Biological Data Analysis: A Complex Networks Perspective

Katerina S. Pechlivanidou

katpechliv@gmail.com

Department of Electrical and Computer Engineering
University of Thessaly

September 2013

Supervisor :
Katsaros Dimitrios
Lecturer
Department of Electrical & Computer Engineering
University of Thessaly

co-Supervisor:
Hatzigeorgiou Artemis
Professor
Department of Electrical & Computer Engineering
University of Thessaly

To my Family

Words of Thanks

My special thanks to my supervisor Mr. Katsaros Dimitrios for the overall support and his guidance throughout this work. I would like to express my appreciation to him for the invaluable advice, for his confidence in my work and for teaching me how to think in a scientific manner.

I also would like to thank all of my dear friends for their unwavering support and for creating an atmosphere of inspiration during my studies. Many thanks for being there for me during every academic and personal challenging experience.

Finally, I owe special thanks to my family for their constant support and their trust in me. They gave me the opportunity to study and complete this thesis. Their encouragement and motivation force me to fulfill my dreams.

Katerina S. Pechlivanidou

Contents

Figure List

Περίληψη

Η ανάλυση των Σύνθετων Δικτύων έχει σημειώσει σημαντική ανάπτυξη κατά τη διάρκεια της τελευταίας δεκαετίας, καθώς μεγάλος όγκος πληροφοριών που αφορούν ποικιλία σύνθετων δικτύων έγινε ευρέως διαθέσιμος. Προς αυτή την κατεύθυνση, πολλά σύνθετα συστήματα μπορούν να περιγραφούν σαν σύνθετα δίκτυα, όπου οι συνιστώσες τους αναπαριστώνται ως κορυφές και οι συνδέσεις τους ως ακμές. Μεταξύ άλλων, η αναπαράσταση με σύνθετα δίκτυα έχει βρει μεγάλο αριθμό εφαρμογών σε πεδία όπως βιοπληροφορική, οπτικοποίηση γράφων, κοινωνιολογία και στην ανάλυση κατανεμημένων συστημάτων. Στη βιολογία, τα σύνθετα δίκτυα αναπαριστούν μια ποικιλία βιολογικών οντοτήτων, από απλούς οργανισμούς μέχρι αντιδράσεις πρωτεϊνών. Μεγάλη προσπάθεια έχει γίνει για την κατηγοριοποίηση κάθε κόμβου του δικτύου ανάλογα με τη θέση τους στη δομή του βιολογικού δικτύου και για τη διάκριση των κόμβων που έχουν μεγάλη επιρροή στο δίκτυο από τους κόμβους των οποίων η απώλεια δεν θα επηρεάσει τη συνοχή και λειτουργικότητα του συνολικού δικτύου.

Ένας γενικός στόχος όταν μελετάμε τέτοιου είδους δίκτυα είναι να ορίσουμε την ευρωστία του συνολικού δικτύου σχετικά με σφάλματα των τμημάτων του. Πιο συγκεκριμένα, η ευρωστία μπορεί να καθοριστεί παρατηρώντας τις αλλαγές του δικτύου καθώς αφαιρούμε τους κόμβους και τις ακμές του ΄ διαγραφές αυτού του είδους μπορούν να θεωρηθούν ως επιθέσεις σύνθετων δικτύων. Θεωρούμε τρείς τύπους επιθέσεων: οι κορυφές μπορούν να αφαιρεθούν ομοιόμορφα με τυχαίο τρόπο, με φθίνουσα σειρά τιμών των μετρικών κεντρικότητάς τους ταυτόχρονα και με φθίνουσα σειρά τιμών των μετρικών κεντρικότητάς τους ακολουθιακά. Η ενδιάμεση κεντρικότητα, ο βαθμός, η κεντρικότητα εγγύτητας και η κεντρικότητα ιδιοδυανύσματος είναι κάποια παραδείγματα μετρικών που έχουν ήδη εμφανιστεί στην ανάλυση σύνθετων δικτύων και έχουν επισημανθεί σε προηγούμενες μελέτες στην προσπάθεια καθορισμού της ευρωστίας τους.

Στην παρούσα εργασία προσπαθούμε να προσδιορίσουμε την επίδραση που έχει στην υποκείμενη δομή των σύνθετων δικτύων η στοχοποίηση κόμβων για διαγραφή σύμφωνα με την τιμή της τοπικής και μη τοπικής μετρικής τους. Εφαρμόζουμε όλα τα προαναφερθέντα είδη επιθέσεων σε βιολογικά δίκτυα και επεκτείνουμε τις ήδη υπάρχουσες εργασίες προσπαθώντας να πραγματοποιήσουμε στοχευμένες επιθέσεις βασισμένες στις τιμές της μ-PCI τιμής κάθε κόμβου. Επιπλέον, προσπαθούμε να εκτελέσουμε επιθέσεις σε μία ομάδα κόμβων σύμφωνα με την k-shell τιμή τους, αφαιρώντας κάθε φορά αυτούς που ανήκουν στο πιο κεντρικό k-core. Σε αυτή τη μελέτη, ο στόχος είναι να εφαρμόσουμε επιθέσεις τόσο σε πραγματικά όσο και σε συνθετικά βιολογικά σύνθετα δίκτυα και να αξιολογήσουμε εκτενώς την ευρωστία και ευπάθειά τους σε στοχευμένες και τυχαίες αστοχίες.

Τα αποτελέσματα δείχνουν ότι η στοχευμένες επιθέσεις σε βιολογικά σύνθετα δίκτυα με βάση τη φθίνουσα σειρά της μ-PCI τιμής των κόμβων του τόσο ταυτόχρονα όσο και ακολουθιακά μπορούν να πραγματοποιηθούν για να προσεγγίσουμε καλύτερα τη μέση τιμή της ευρωστίας και της ευπάθειάς τους, συγκριτικά με τις άλλες μετρικές κεντρικότητας. Πιο συγκεκριμένα, το 2-PCI επιτυγχάνει την καλύτερη προσέγγιση συγκριτικά με όλες τις άλλες μετρικές. Το k-shell κατορθώνει να εκθέσει καλύτερα την ευπάθεια του δικτύου απ’ ότι να απεικονίσει την ευρωστία του, ενώ για την περίπτωση των συνθετικών δικτύων ισχύει ακριβώς το αντίθετο. Το k-shell δεν μπορεί να θεωρηθεί ως βέλτιστη κεντρικότητα για επίθεση. Ο βαθμός των κόμβων μπορεί να θεωρηθεί ως βέλτιστη κεντρικότητα ώστε να

αναδείξουμε την ευρωστία βιολογικών σύνθετων δικτύων όταν συμβαίνουν σφάλματα κατά τη διάρκεια ταυτόχρονων επιθέσεων και την ενδιάμεση κεντρικότητα και την εκκεντρικότητα όταν εφαρμόζουμε ακολουθιακές επιθέσεις.

Abstract

The analysis of **Complex Networks** has received considerable development during the last decade, since huge amount of information of a variety of complex networks became widely available. In this direction, many complex systems can be described by complex networks, where the components are represented by vertices and their connections by edges. Among others, complex network representation has found a number of applications in areas as bioinformatics, graph visualization, sociology and distributed system analysis. In biology, complex networks represent a variety of biological units, from simple organisms to protein interactions. Huge amount of effort has been devoted on classifying the role of each individual node according to their position in the biological network structure and distinguishing high-impact nodes from nodes whose loss will not affect the consistency and functionality of the system.

An overall goal when studying such networks is defining the robustness of the entire system to the failure of its parts. In particular, robustness definition can be addressed by observing how the structure of the network changes as vertices and nodes are removed; these kinds of edge and node removals can be considered as attacks of the complex network. We consider three types of attacks: vertices are removed uniformly at random, in decreasing order of their centrality measure simultaneously and in decreasing order of their centrality measure sequentially. Betweenness, eccentricity, degree, closeness and eigenvector centrality are some examples of the metrics that have already been introduced in complex network analysis and that have already been considered in previous work when trying to identify the robustness of complex networks.

Here we are trying to identify the effect on the underlying network structure of targeting vertices for removal according to their value of local and non-local measures. We apply all aforementioned attacks on biological networks and we extend the existing studies by trying to employ targeted attacks based on the μ-Power Community Index (μ-PCI) value of each vertex. In addition, we try to perform attacks on a group of nodes according to their k-shell value, by removing each time those which belong to the most central one. In this study, the goal is to perform node and edge attacks both on empirical and on synthetic biological complex networks and evaluate extensively their robustness and vulnerability towards malicious and random failures.

The results show that sequential and simultaneous target attacks by descending order of μ-PCI value of nodes can be performed on biological complex networks to approximate the average robustness and vulnerability of the network better than the other centrality measures. Particularly, 2-PCI has the best approximation compared to all other centralities. K-shell manages to capture vulnerability better than exposing the fragility of the network against malicious attacks and the opposite is true for synthetic networks; either way, k-shell cannot be considered as an optimal centrality measure for attack. Degree can be considered as the superior centrality in order to highlight the robustness of a biological complex network against simultaneous target attacks and eccentricity or closeness when performing the attacks sequentially.

Keywords: Biological Complex Networks, Robustness, Vulnerability, Simultaneous Target Attack, Sequential Target Attack, Random Attack, μ-PCI, k-shell

1 Introduction

Many complex systems can be represented by complex networks, even if they initially seem unrelated to this concept. We refer to complex networks, when their components are described by vertices (nodes) and their connections by edges. There are numerous available examples of networks in many disciplines such as technology, sociology and biology. We face some of the most important examples of complex networks in the latter category, since the availability of large biological datasets has led to the recent popularity of the study of Protein Interaction Networks and to the development of such network types. In particular:

- Protein Interaction Networks (PINs), which is the most important category of networks in complex biological systems analysis. In PINs representation, proteins are represented as nodes and their interactions as edges.
- Neuronal networks (NNs)
- Gene regulatory networks (DNA-protein interaction networks)
- Signaling Networks
- Species Interaction Networks
- Metabolic Networks
- Food webs

are only some of the most representative biological complex networks examples. In biology, complex network studying has shifted its focus on large-scale network analysis, since almost every biological system is described by very large networks. Large-scale biological networks are referenced as "omes", such as genome, interactome, proteome, diseasome, biome and so on.[18]

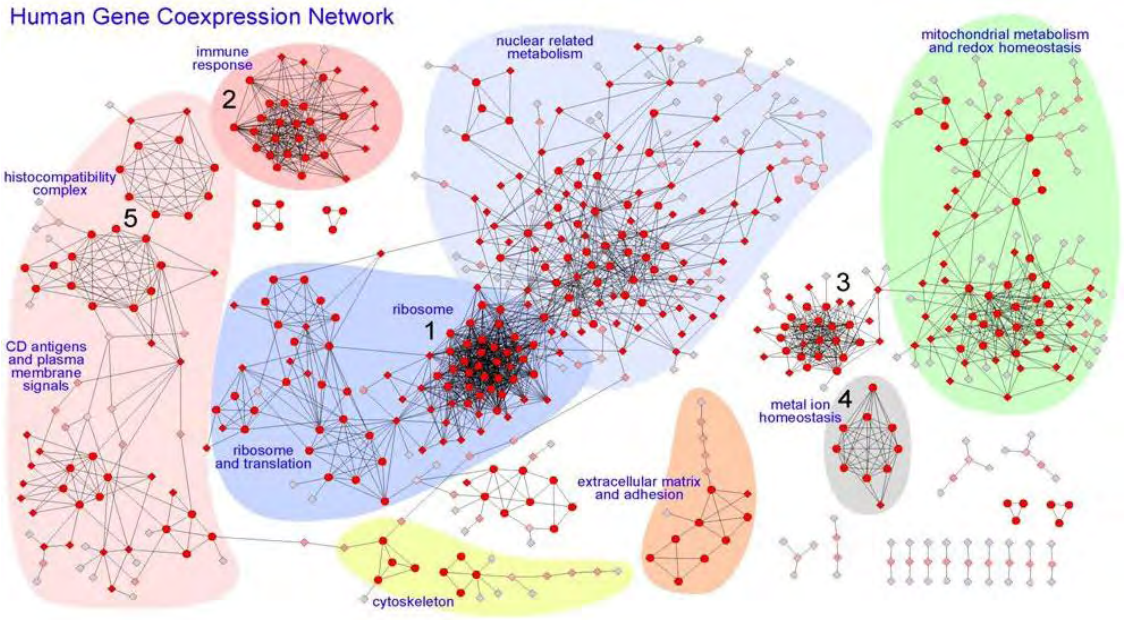


Figure 1| The Human Gene Co-expression Network Image can be found here: http://bioinfow.dep.usal.es/coexpression/network.jpg

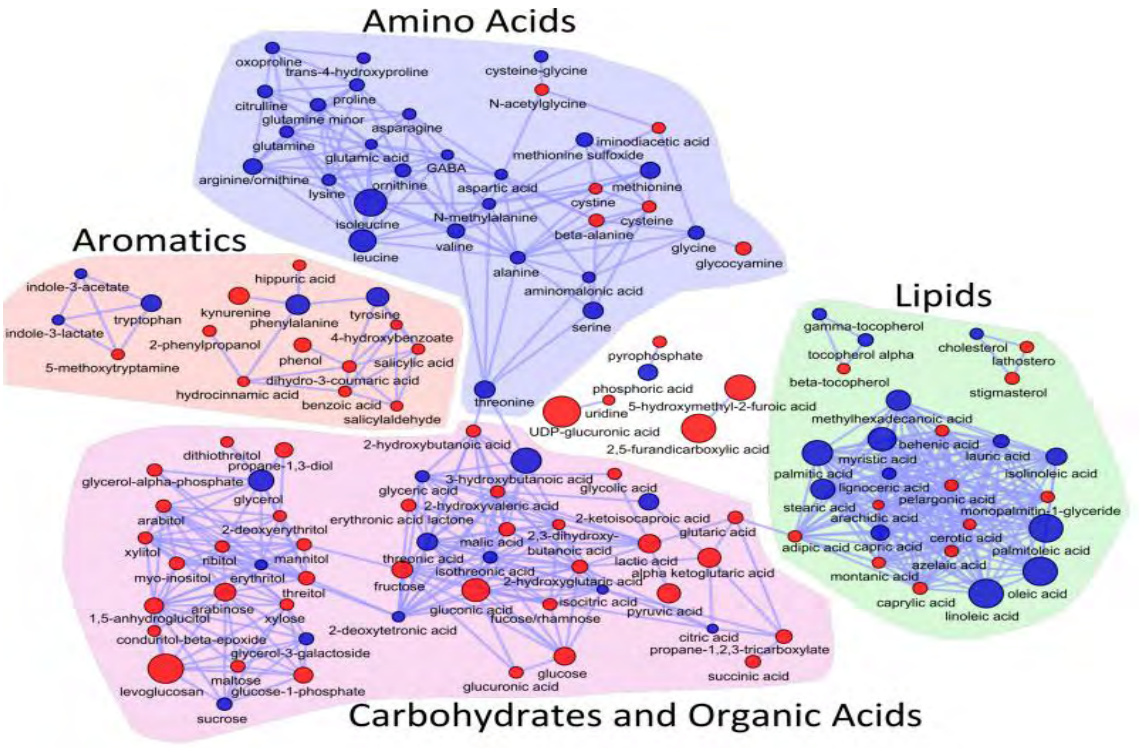


Figure 2| A Metabolic Network
Image can be found here:
http://fiehnlab.ucdavis.edu/staff/grapov/grapov-metabolic-network-jpg.png

The construction of a robust network of complex systems is a great challenge to the scientific community. An important aspect of studying the behavior of complex networks is to define the effect of random and malicious failures on the individual components of the network and on the whole system when vertices along with the edges attached to them are removed according to a centrality value. It is clear, that the importance of each node differs in each case of networks. Consider an example in which an attacker installs a virus on a machine. In this case, the attacker has to choose the most important node, here the hub node which is the node with the most connections to other nodes, in order to spread the virus installed efficiently. Now, consider a network of proteinomics, in which proteins are represented as nodes and their interactions as edges. In this case, deleting a hub-node is more likely to be fatal to an organism than non-hub, a phenomenon known as the centrality-lethality rule. [1]

It became imperative, regardless of the network type, to define some characteristics of complex networks that can indicate a robust underlying structure. At this point, we have to clarify what robustness is without giving yet a strict but a more free and intuitive definition. Among other definitions, robustness of a network can be described as the ability of a network to maintain its total throughput under node and link removal [2]; this means that a robust network maintains to keep its components strongly connected and thus it is less sensitive to node attacks.

In order to measure the robustness of complex networks and the importance of nodes within them, in the last few years, a significant amount of metrics and methods have been

evolved and introduced in network analysis. Significant effort was made for this purpose. In previous related works, studies for the effect on the performance of the whole network was devoted and particularly cases in which node deletion in random or descending order of their degree, betweenness, eigenvector, eccentricity centrality value or their k-coreness is being performed. The purpose of this work is to extend previous work and also to investigate the robustness of biological networks when μ-PCI attacks are performed.

Moreover, in each case other network characteristics, like vulnerability, maximum component size, number of triangles in the network, number of k-cores etc. have been studied and results are provided. Networks that are used in this study are retrieved both from real-world biological datasets and from synthetic networks like Barabási -Albert and Erdős–Rényi graphs. The last two graph types are selected for our purpose, since they are the most similar to real-world network representations, with the first one to be a much better representation; the second one has a lack of many characteristics of empirical biological networks.

The rest parts of this paper are organized as follows. In Sect. 2 we review the related work, our motivations and contributions; in Sect. 3 we describe Biological Network models, the centralities we used in our work, the network measures and the three attack types. We also define robustness and vulnerability in the same section. In Sect. 4 we describe the Network models we used in our work, our simulation model and the results retrieved. Finally, Sect. 5 concludes our work.

2 Related Work

Defining the robustness of a network against random and malicious failures on the individual components of the network and on the whole system has attracted significant attention in the literature concerning various types of networked systems. The studies along this line have been able to examine the robustness of, e.g. technological networks like the Internet topology at the autonomous system (AS) level [3], social networks like the Facebook network [4], Biological networks like Protein Interaction Networks (PINs) [5], and so on.

In the context of network robustness, it is beneficial to investigate in depth the attack of nodes according to their value of as many centrality measures as possible in order to define the most aggressive attack or the robust network infrastructure. Huge amount of effort has been devoted to study the robustness when performing deletion of nodes according to their degree and betweenness value. Particularly, Ali Sydney et al. [2] considered the latter attacks and elasticity as a robustness measure and showed that link redundancy is a sufficient but not an essential criterion for robustness.

A more thorough research on the on the definition of robustness of a network has been made by Iyer, Killingback, Sundaram and Wang in [5]. They suggested as robustness measurement the largest existing component of the network compared to the fraction of nodes that have been removed. The deletion of the vertices follows the descending order of betweenness, eccentricity, closeness, degree and eigenvector values that we describe in the next chapter in detail. They staged the attacks one step further, considering not only random and malicious failures, but they focus both on simultaneous and on sequential targeted attacks when targeting the nodes.

Another metric used to highlight the underlying structure of networks, mostly very large sized, and its hierarchies which cannot be captured using other metrics due to their size, is the k-shell or k-core of a graph. Since real-world networks tent to have enormous size, the problem of the k-core decomposition has emerged and was addressed in recent works [8] [9] [10]. There are numerous examples of networks that have been "decomposed" based on this method; study of the Internet topology at the autonomous system (AS) level, discovery of the role of proteins in complex proteinomic networks are only some of the examples that can be mentioned. The k-coreness is used to identify a group of nodes with degree at least k. It became clear, that higher values of k-coreness correspond to more central nodes of the network. In our study, we try to process a targeted attack on nodes with the largest coreness in a biological complex network and we present the obtained results in Sect.4 and 5.

Similar work has been provided in [7] by Nicosia, Criado, Romance, Russo and Latora but in their study the problem of identifying the central elements in a network is described inversely. Therefore, they define a subset of nodes, called controlling set, which can prescribe a set of centrality values to all nodes of the network. Although they do not define the robustness of a network, their work proves that we can find a set of nodes whose role is more important than other nodes in the network and thus they can be considered as perfect candidates in a targeted attack.

Although it cannot directly be related to our work, in [6] Katsaros, Tassioulas, Dimokas and Manolopoulos developed a new metric called μ-Power Community Index (μ-PCI) which is more informative that the node degree and it is not affected by any isolated nodes. This latter measure provides a more localized/centralized centrality measurement and it was proposed as a criterion to select sensors with a special role in a grid of sensors based on their ability to influence the communication between multi-hop connected nodes.

2.1 Motivations and contributions

The proposed centralities so far are used as a criterion to rank nodes according to their importance in order to measure robustness of complex networks focus on the power each node individually in a network has. These kinds of metrics are strictly related to the characteristics of each node. All of these centralities examine each node as an individual entity in the network. But how can its importance be affected if it belongs to a powerful neighborhood? What happens if this node is deleted? Will the network collapse? What happens if we delete a whole set of nodes and particularly the most 'central' k-shell? Can a biological network maintain its overall throughput under such a node deletion or will it degrade rapidly? These questions we will try to answer in this work.

Motivated by the research in [6] for defining the most powerful sensor, here node, we propose the μ-PCI as a new criterion to target nodes for deletion in attacks on Biological networks. Moreover, considering that biological networks have a modular organization, we also extend previous studies and investigate the evolution of the giant component's size during a node removal according to their k-core values. Then, we use the results to calculate numerically the robustness and vulnerability of both empirical and synthetic biological complex networks.

In summary, our work's contributions are the followings:

- Definition of the robustness of biological complex networks. We study of the effect of random and malicious failures on the individual components of the network and on the whole system when vertices are removed according to their:
    - μ-PCI value (sequentially & simultaneously)
    - k-coreness (simultaneously)
    - degree, eccentricity, eigenvector, closeness and betweenness centrality value (sequentially & simultaneously)

- Definition of the robustness of biological complex networks against random failures.

- Evaluation of the outcomes. Robustness of the biological network is used as the main criterion to characterize its infrastructure, its resistance to disconnection of its components and its throughput under node and link removal.

3 Complex Networks

3.1 Complex Networks

A complex network is a graph (network) with non-trivial topological features - features that do not occur in simple networks such as lattices or random graphs but often occur in real graphs [19]. In all cases of Complex Networks, vertices are considered to be the elements of the represented complex system and the edges between them mean that they are associated in some order level; see Figure 3 and 4 for example.

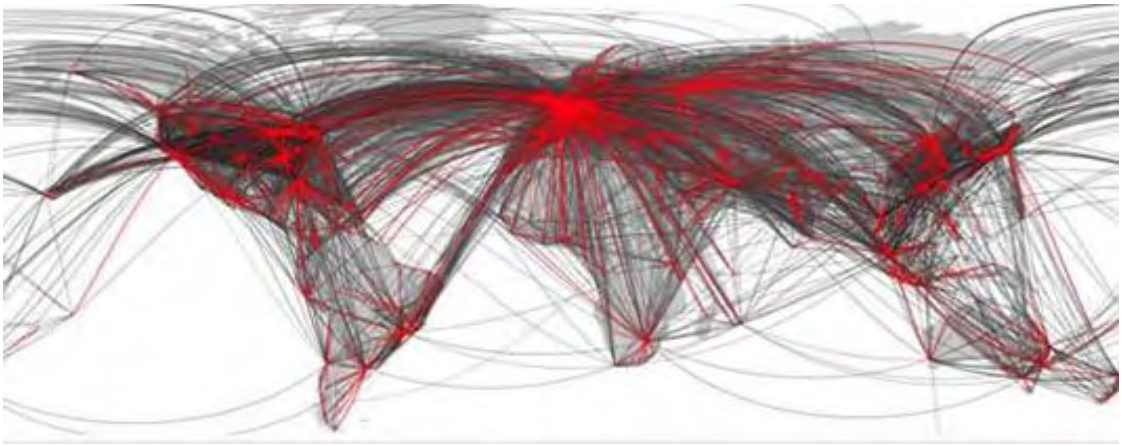


Figure 3| The worldwide air transportation network. Each grey link reassembles traffic of passengers between more than represent the network's skeleton, a tree-like structure only 1,300 links that represents the core structure of the network. Link in the skeleton are the most important connections of the network. Image can be found here: http://optimizationandanalytics.files.wordpress.com/2013/01/complex-network-structure.png

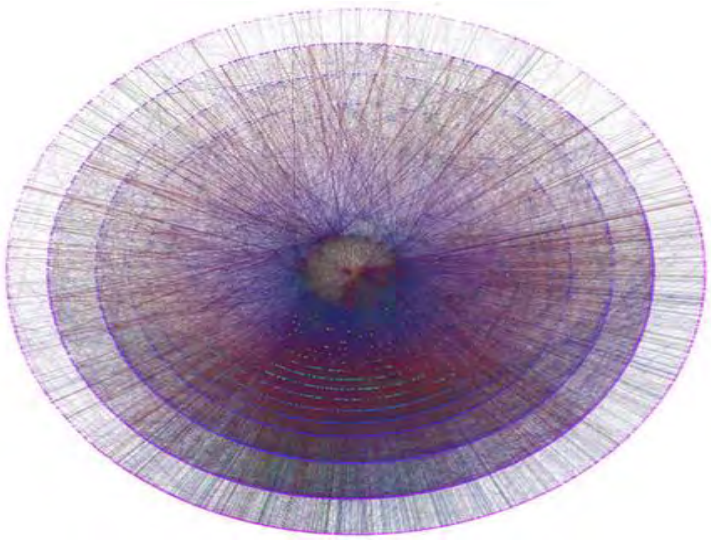


Figure 4| The Internet AS The m-core decomposition of the Internet AS. Light purple means 1-coreness and red means 21-coreness. Image can be found here: http://www.nature.com/srep/2013/130827/srep02517/fig_tab/srep02517_F5.html

3.2 Biological Complex Network Models

3.2 .1 The Biological Complex Network Model

A biological network is any network that applies to biological systems. A network is any system with sub-units that are linked into a whole, such as species units linked into a whole food web. Complex biological systems may be represented and analyzed as computable networks [18]. Protein-Protein Interactions, the most common biological network model, are mainly represented by Biological Complex Networks, e.g. PINs (Protein Interaction Networks), where the proteins are represented as nodes and the connections between the interacting proteins are shown as edges [11].

In this work, we will use undirected graphs to form the desired biological networks since they are commonly represented as undirected graphs, i.e. graphs where the edges are not directed and therefore edge (u,v) is identical to (v,u).

Definition 1: A graph or an undirected graph is a tuple G = (V,E) with a nonempty set V whose elements are called vertices, nodes or points a (possibly empty) set E of unordered pairs of elements of V called links or edges.[20]

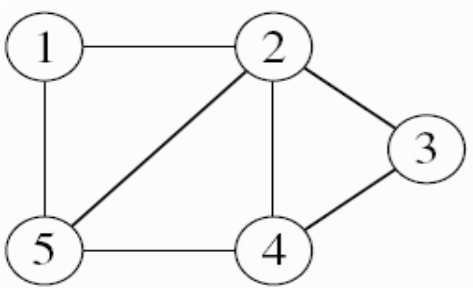


Figure 5| Undirected Graph

Image can be found here: http://homepages.ius.edu/rwisman/C455/html/notes/AppendixB4/B4-2.gif

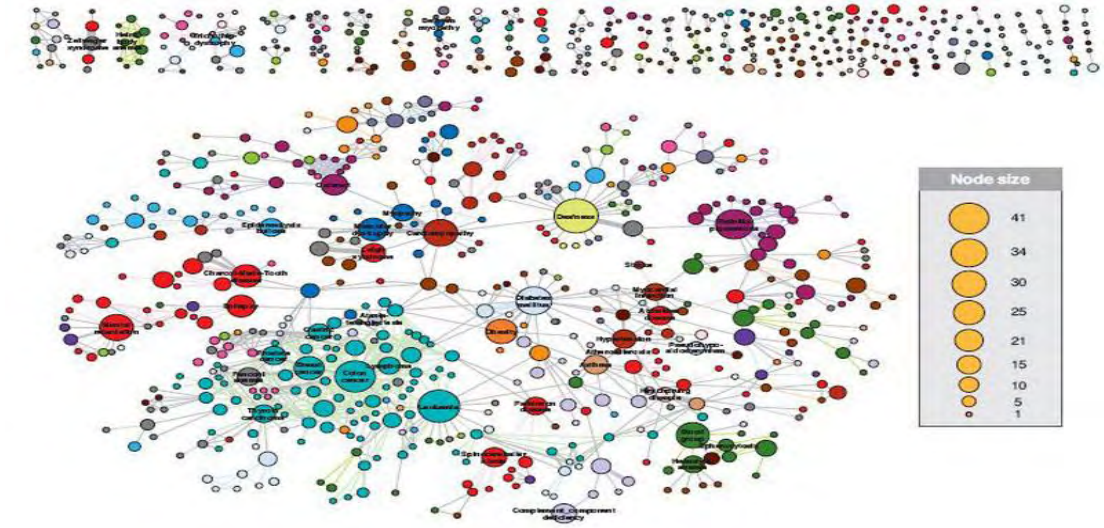


Figure 6| Human disease network (HDN) In the HDN, each node corresponds to a distinct disorder, colored based on the disorder class to which it belongs. A link between disorders in the same disorder class is colored with the corresponding dimmer color, and links connecting different disorder classes are gray. The size of each node is proportional to the number of genes participating in the corresponding disorder (see key), and the link thickness is proportional to the number of genes shared by the disorders it connects.

Biological networks share a number of common global features which are listed below. Biological networks have [17]:

- **a scale-free degree distribution;** this means that they contain a number of hub nodes that are the most important in the network [15]
- **a small average shortest path length** between any two nodes; this is also knows as small-world model
- **a disassortative nature**
- **a modular organization**
- **a structural and dynamical robustness**

In our research we tried to examine as many biological network types as possible. We provide both empirical biological networks and synthetic ones. In order to highlight topological properties of synthetic biological networks, the scale-free network model proposed by Barabási Albert and the Erdős–Rényi model for random graphs where used. The first model is suitable to represent biological complex networks; the second is simpler than the real networks but since real-world networks have a small average diameter like the ER model we used it to see the effects of attacks on its structure.

3.2.2 The Barabási– Albert (BA) scale-free Network Model

Definition 2: Barabási– Albert scale free model

The Barabási– Albert (BA) model is an algorithm for generating random scale-free networks using a preferential attachment mechanism. Scale-free networks are widely observed in natural and human-made systems and therefore ideal to represent synthetic biological networks. [21] [12]

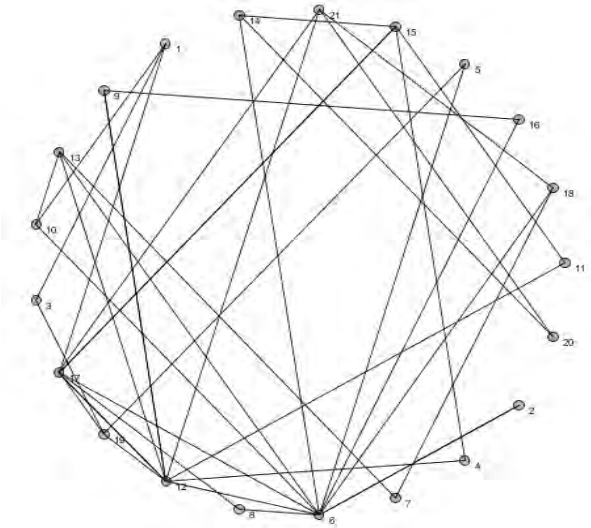


Figure 7| The Barabási– Albert Graph The graph was created with CentiBin. Number of iterations was set to 20 and random seed to 10.

The Barabási– Albert Algorithm

The network begins with an initial connected network of m_0 nodes. New nodes are added to the network one at a time. Each new node is connected to $m \leq m_0$ existing nodes with a probability that is proportional to the number of links that the existing nodes already have. Formally, the probability p_i that the new node is connected to node i is:

$$p_i = \frac{k_i}{\sum_j k_j}$$

where k_i is the degree of node i and the sum is made over all pre-existing nodes j (i.e. the denominator results in the current number of edges in the network). [21]

The most important properties of Barabási– Albert scale free model [21]

- Heavily linked nodes ("hubs") tend to quickly accumulate even more links, while nodes with only a few links are unlikely to be chosen as the destination for a new link. The new nodes have a "preference" to attach themselves to the already heavily linked nodes.

- The degree distribution resulting from the BA model is scale free, in particular, it is a power law of the form $P(k) \sim k^{-3}$

- The average path length of the BA model increases approximately logarithmically with the size of the network.

- While there is no analytical result for the clustering coefficient of the BA model, the empirically determined clustering coefficients are generally significantly higher for the BA model than for random networks. The clustering coefficient also scales with network size following approximately a power law $C \sim N^{-0.75}$

- The Barabási– Albert (BA) model is suitable for Metabolic networks and food-webs and belongs to the Scale-free networks category with networks with a power-law distribution $P(k) \sim k^{-\gamma}$

3.2.3 The Erdős–Rényi Random Network model

Definition 3: Erdős–Rényi Random Graph model [22]

The Erdős–Rényi model is either of two closely related models for generating random graphs, including one that sets an edge between each pair of nodes with equal probability, independently of the other edges.

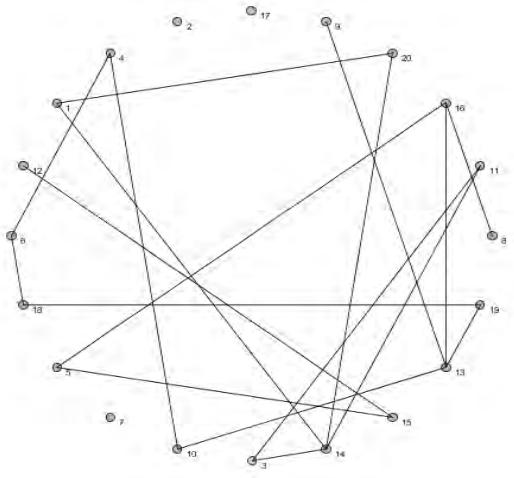


Figure 8| The Erdős–Rényi Graph The Graph was created with CentiBin with edge probability set to 0.09

The most important properties of Erdős–Rényi Random Graph model [22]

- The expected number of edges in G(n, p) is $\binom{n}{2} p$
- If np < 1, then a graph in G(n, p) will almost surely have no connected components of size larger than O(log(n)).
- If np = 1, then a graph in G(n, p) will almost surely have a largest component whose size is of order $n^{2/3}$
- If $np \rightarrow c > 1$, where *c* is a constant, then a graph in G(n, p) will almost surely have a unique giant component containing a positive fraction of the vertices. No other component will contain more than O(log(n)) vertices.
- If $p < \frac{(1-\epsilon)\ln n}{n}$, then a graph in G(n, p) will almost surely contain isolated vertices, and thus be disconnected.
- If $p > \frac{(1+\epsilon)\ln n}{n}$, then a graph in G(n, p) will almost surely be connected.

The Erdős–Rényi (ER) model is suitable for uncorrelated random graphs and belongs to the single-scale network category with a sharp distribution of vertex degrees exhibiting exponential or Gaussian tails. [14] The construction of the network can be described as following: At first we have N disconnected nodes and then we add edges according to a fixed edge probability.

3.3 Robustness & Vulnerability

Complex Networks provide significant insides into the ability of a complex system to maintain its throughput under node attack. The idea for defining its robustness and therefore its vulnerability too, is to observe the evolution of the size of the giant component of the network during the attack. The process of deleting a fraction of nodes together with the edges connected to them from a network is known as percolation. Percolation is a key approach to measure the robustness of a network. An important aspect of understanding this concept is to understand beforehand that the larger the largest component is compared to the size of the network the harder it is for its components to fail against node and link attack. The simple example below can describe at a very basic level the robustness concept in networks. Of course, simple networks like the one below are unlikely to occur in any real-world complex systems and we use it only to point out the meaning of robustness.

To make an intuitive approach of network robustness we assume these two networks:

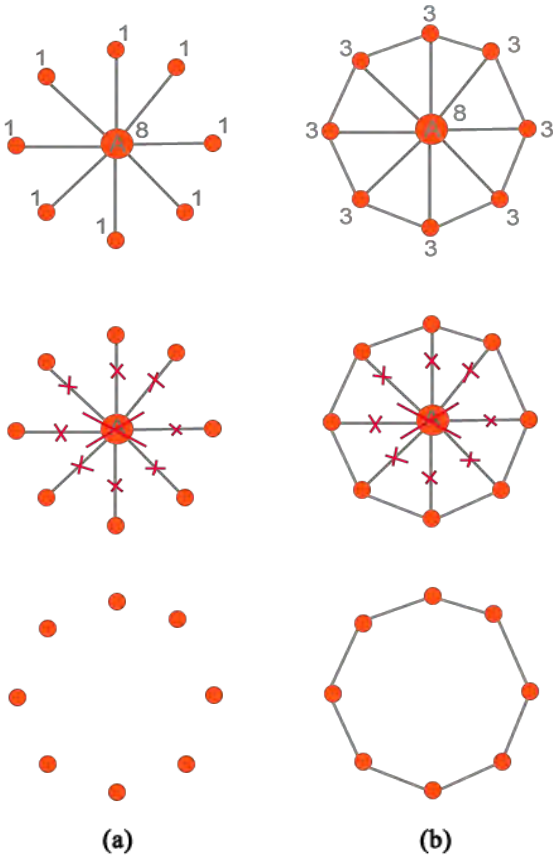


Figure 9| (a) Star Graph (b) Ring Graph with node A in center. A has edges to all peripheral nodes. Every node in 9a has degree=1 except from node A which has degree=8. Every node in 9b has degree=3 except from node A which has degree=8. In both cases degree attack is performed and thus node A, which has the highest degree value, is removed.

In the first case we see that if a degree strategic attack takes place, then node A is chosen to be removed first because of its high degree value. In the second case we observe that using the same criterion node A is removed again. Although we have the same number of nodes, each attack leads to a different result; case 9b appears to be more robust because after one node attack it does not fall apart directly. Contrariwise, network in 9a cannot manage to hold any of its components together. Therefore, one can assume that network 9b is more robust than network 9a.

To calculate the ability described above numerically we define robustness as follows [5]: Let N be an initial network of size N. Let N_p be the network that results when a fraction p of the vertices is being removed together with the edges connected to them. Consider vertex removal either uniformly randomly either targeted in descending order of any measurement that shows their importance. We will clarify and analyze thoroughly node deletion in the following chapter. The largest component of N_p will be denoted by N_p^c. In order to quantify the robustness, we have first to calculate the fraction $\sigma(\rho) = | N_p^c | / N$.

We define the quantity of the robustness R as follows:

$$R = \frac{1}{N} \sum_{i=1}^{N} \sigma(i/N) \qquad , R \in [0, 1/2]$$

The fraction $\frac{1}{N}$ allows the robustness to be compared with the R value of other networks of various sizes and structures. The extremities of R correspond to the following networks: we observe R value equal to $\frac{1}{N}$ when a star graph is provided and equal to $\frac{1}{2}(1 - \frac{1}{N})$ when a fully connected graph is provided.

We define the quantity of vulnerability V of the network as follows:

$$V = \frac{1}{2} - R \qquad , V \in [0, 1/2]$$

Think of vulnerability as a measure of the weakness of the network to resist to a thread, i.e. node attack.

3.4 Centrality Measures

Vertex and edge removal presuppose identifying a centrality measure that quantifies the importance of each vertex (node) of the network. This allows arranging the nodes according to their importance. Below are the definitions and the descriptions of each centrality measure used in this work.

Degree Centrality

Degree of a node is assumed to be the number of edges it is connected to. It is the simplest centrality measure. Given an undirected graph G = (V, E), with size N = |V| the degree centrality of a node (vertex) v ∈ V of the network is equal to deg(v).

$$d_v = \deg(v)$$

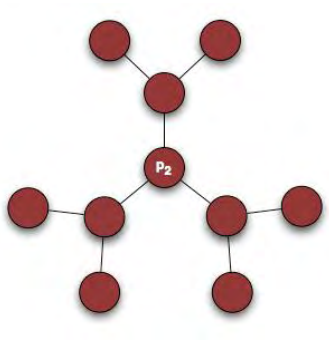


Figure 10| A Complex Network. Node P2 has degree centrality equal to 3 and therefore dv =3
Image can be found here: http://assets.20bits.com/misc/low-degree.png

Betweenness Centrality

Betweenness Centrality is another useful centrality measure that takes proper account of the importance and load of the node. It is equal to the number of shortest paths from all vertices to all others that pass through that node. The betweenness centrality of a node v is given by the expression:

$$B_v = \sum_{s \neq v \neq t} \frac{\sigma_{st}(v)}{\sigma_{st}}$$

where σ_{st} is the total number of shortest paths from node s to node t and σ_{st} is the number of those paths that pass through v. Since we deal with undirected networks and in order to

normalize the betweenness centrality value, i.e. $g \in [0,1]$, it has to be divided by $\frac{(N-1)(N-2)}{2}$, where N is the number of nodes in the largest component. B_v achieves its highest value when v is cross by every single shortest path [23].

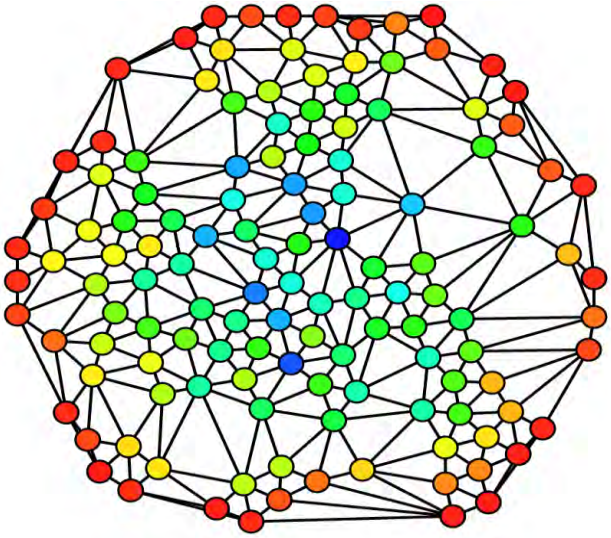


Figure 11| A Network whose vertices are ranked according to their betweenness centrality. Hue from red=0 to blue=max shows the node betweenness. Image can be found here: http://upload.wikimedia.org/wikipedia/commons/6/60/Graph_betweenness.svg

Eigenvector Centrality

Eigenvector centrality is a measure of the influence of a node in a network . It assigns relative scores to all nodes in the network based on the concept that connections to high-scoring nodes contribute more to the score of the node in question than equal connections to low-scoring nodes. [24]

For a given graph G(V,E) with |V| number of vertices let A = $(a_{v,t})$ be the adjacency matrix, i.e. $a_{v,t}$ = 1 if vertex v is linked to vertex t, and $a_{v,t}$ = 0 otherwise. The centrality score of vertex v can be defined as:

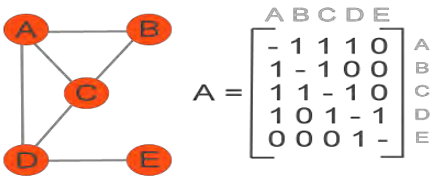
$$x_v = \frac{1}{\lambda}\sum_{t\in M(v)} x_t = \frac{1}{\lambda}\sum_{t\in G} a_{v,t}\, x_t$$

where M(v) is a set of the neighbors of v and λ is a constant. With a small rearrangement this can be rewritten in vector notation as the eigenvector equation

$$A\cdot x = \lambda\cdot x$$

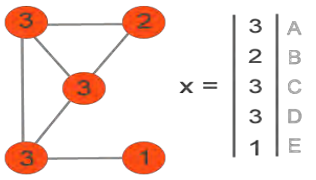
In general, there will be many different eigenvalues λ for which an eigenvector solution exists. However, the additional requirement that all the entries in the eigenvector are positive implies (by the Perron-Frobenius theorem) that only the greatest eigenvalue results in the desired centrality measure. The v^{th} component of the related eigenvector then gives the centrality score of the vertex v in the network. Power iteration is one of many algorithms that may be used to find this dominant eigenvector. Furthermore, this can be generalized so that the entries in A can be real numbers representing connection strengths, as in a stochastic matrix. Consider the following example:

Matrix A is the 5x5 adjacency matrix for this undirected graph G=(V,E)



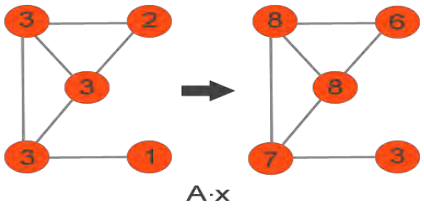
(a)

Vector x contains the degree centrality for each vertex



(b)

If we then multiply A and x the result for aech vertex is shown on the graph below



(c)

Figure 12| A Graph that represents the calculation of the eigenvector centrality. 12a, 12b and 12c show the steps.

If we take a closer look we can see that the degree centrality has been spread out. Each vertex now has a value that corresponds to the sum of the degrees of its neighbors. Our goal now is to find a matrix that has the following characteristic:

$$B \cdot x = \lambda \cdot x$$

$$B \cdot \begin{bmatrix} x1 \\ x2 \\ x3 \\ x4 \\ x5 \end{bmatrix} = \begin{bmatrix} \lambda * x1 \\ \lambda * x2 \\ \lambda * x3 \\ \lambda * x4 \\ \lambda * x5 \end{bmatrix}$$

In this case the vector is called Eigenvector of A and the entries Eigenvector centralities of the vertices. The vector can be multiplied by the adjacency matrix and return itself multiplied by a scalar. [29] The eigenvector and eigenvalues can be calculated by solving the following equation:

$$(A - \lambda I) \cdot x = 0$$

where I is the identity matrix.

Eccentricity Centrality

The distance between two vertices in a graph is the number of edges in a shortest path connecting them. This shortest path is also known as geodesic path and the number as geodesic distance. The eccentricity e_v of a vertex v is the greatest geodesic distance between v and any other vertex. It can be thought of as how far a node is from the node most distant from it in the graph [25] [26] or one can assume that e_v reflects how far is each node from every other node at most in the graph. Therefore, if we consider $dist_{max} v \in V$ to be the maximum distance node v has to any other node in the network then e_v can be considered as

$$e_v = \frac{1}{dist_{max} v \in V} \quad [13]$$

Closeness Centrality

Closeness centrality is based on the mean value of the distance from the node under consideration to all other nodes of the network. Therefore we can find the mean geodesic path as suggested in [5]:

$g_i = \frac{1}{N} \Sigma_{j \in V} \gamma_{ij}$, if path γ_{ij} for i=j is also included in the sum

or

$g_i = \frac{1}{N-1} \Sigma_{j \in V} \gamma_{ij}$, if path γ_{ij} for i=j is not included in the sum

Then we can define closeness centrality as

$$c_i = \frac{1}{g_i}.$$

From the c_i type, it is clear that nodes that are connected to many other nodes through geodesic paths have a larger value of closeness. At this point, it is important to mention that closeness centrality shows how long it will take for information to spread out from a single node v to all other nodes sequentially [24]. Here follows an example:

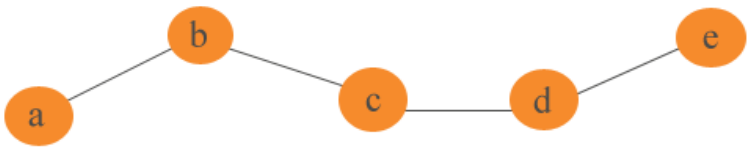


Figure 13| An undirected Graph

We now calculate the closeness for each node in order to find which the most "central" one is.

$$c_a = \frac{1}{g_a} = \frac{4}{1+2+3+4} = \frac{4}{10} = 0.4$$

$$c_b = \frac{1}{g_b} = \frac{4}{1+1+2+3} = \frac{4}{7} \cong 0.57$$

$$c_c = \frac{1}{g_c} = \frac{4}{2+1+1+2} = \frac{4}{6} \cong 0.67$$

$$c_d = \frac{1}{g_d} = \frac{4}{3+2+1+1} = \frac{4}{7} \cong 0.57$$

$$c_e = \frac{1}{g_e} = \frac{4}{4+3+2+1} = \frac{4}{10} = 0.4$$

As the results show, c is definitely the most central node and thus it can reach each other node in a short distance.

k-core / k-shell

A subgraph G(C) induced by the set C⊆V is a k-core iff $\forall u \in C: d_{G(\overline{C})}(u) \geq k$ and G(C) is maximal, i.e., for $\overline{C} \supset C$, there exists $v \in \overline{C}$ such that $d_{G(\overline{C})}(v) < k$. A node v of G is said to have coreness k iff it belongs to the k-core but not the (k+1)-core. All nodes with coreness k constitute a special group of nodes named k-shell. The k-core is obtained by recursively removing all nodes of degree smaller than k, until the degree of all remaining vertices is larger or equal to k; a process also known as k-shell decomposition.[9]

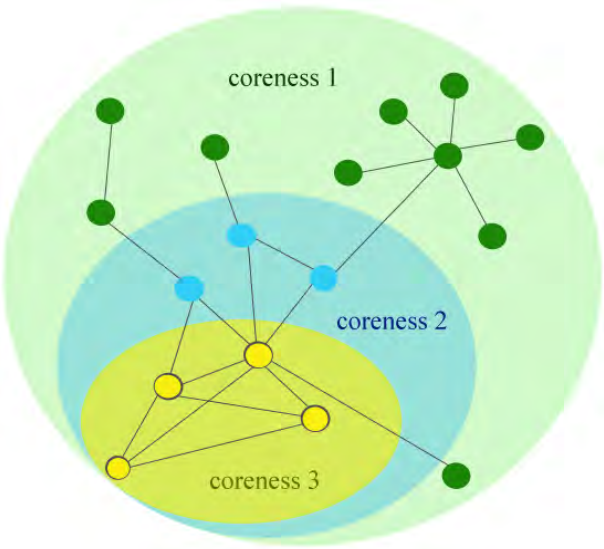


Figure 14| k-cores and k-shells of an undirected graph. It is obvious from the graph that all green nodes constitute 1-shell, because they have degree at least one, they belong to 1-core (the green one) but not the 2-core (the blue one). Therefore, the green core represents the 1-core of the given graph.

k-shell decomposition was introduced as a technique to characterize very large graphs beyond its degree distribution. It helps to highlight the underlying structure of the network that is hindered by its large size. Since biological networks consist of a large number of vertices, we used this tool to make targeted attacks on nodes. The k-shell decomposition of a network suffers two serious shortcomings due to its size: first k-shell decomposition is a very slow process when the size is enormous and second memory constrains occur in this case. We have developed a distributed and parallel algorithm for this purpose to achieve the decomposition. The algorithm is developed in the Hadoop's MapReduce environment, which is a cloud based master-slave platform.

μ-Power Community Index

The μ-Power Community Index (μ-PCI) of a sensor (here a node is assumed to be a sensor) v is equal to k, such there are up to μ x k sensors in the μ-hop neighborhood of v with degree greater than or equal to k and the rest of the sensors within that neighborhood have a degree less than or equal to k [6]. As described in [6], nodes which have more connections (large d_v value) are more likely to be powerful because they can clearly affect more nodes. It is clear that their power depends on the d_v of their one-hop neighbors. If a node has a large μ-PCI value, it means that this node can reach other nodes on short paths.

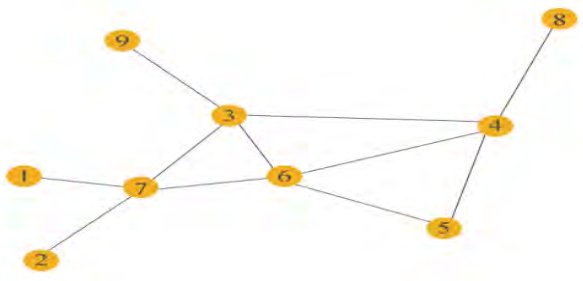
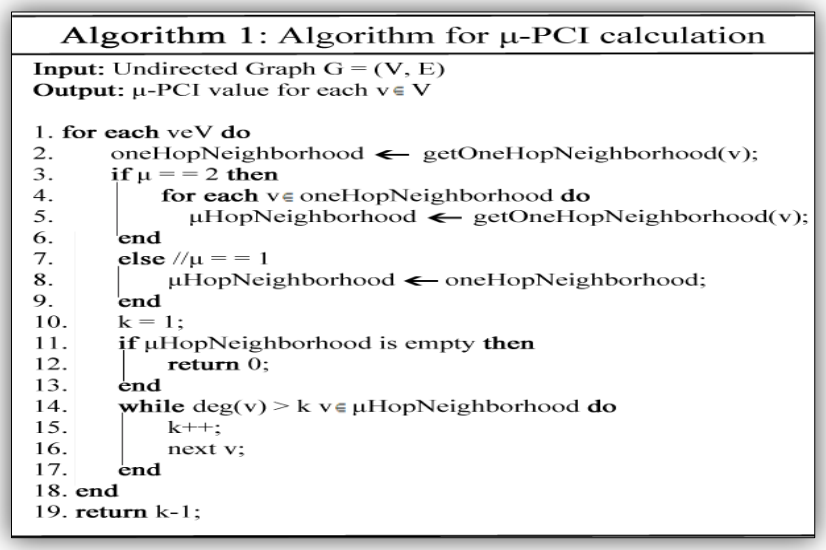


Figure 15| An undirected graph For μ=1, PCI(6) = PCI(3) = 3 and PCI(7) = PCI(4) = 2, PCI(1) = PCI(2) = PCI(8) = PCI(9) = 1, PCI(5) = 2. For μ=2, 2-PCI(6) = 2-PCI(3) = 2-PCI(4) = 1, 2- PCI(1) = 2-PCI(2) = 2-PCI(5) = 2-PCI(7) = 2-PCI(8) = 2 and 2-PCI(9) = 3. Notice that when a node does not have any μHop neighbor then the μPCI value is set to zero.

The μ-PCI algorithm can be calculated for any value of μ; if it is calculated for μ=1, the μ-PCI is then considered as the plain Power Community Index (PCI). In our study we use only the PCI and 2-PCI values. Larger values of μ are not in our interest; μ-PCI values indicate that the node can reach others in a relatively short path and when values of μ are larger than 2 then the concept of direct influence on other nodes from the node under consideration is basically lost. Having in mind the latter note, it is obvious that when a node does not have a μ-hop neighborhood the μ-PCI value is set to zero. It means practically that in this case the node is not able to exert influence to any other node of the network. In case where the goal is a targeted attack, this node is not considered as a perfect candidate for removal.

We now present the algorithm we used to compute PCI and 2-PCI values in this work.



Algorithm 1: Algorithm for μ-PCI calculation

Input: Undirected Graph G = (V, E)
Output: μ-PCI value for each v∈ V

```
1. for each veV do
2.     oneHopNeighborhood ← getOneHopNeighborhood(v);
3.     if μ = = 2 then
4.         for each v∈ oneHopNeighborhood do
5.             μHopNeighborhood ← getOneHopNeighborhood(v);
6.     end
7.     else //μ = = 1
8.         μHopNeighborhood ← oneHopNeighborhood;
9.     end
10.    k = 1;
11.    if μHopNeighborhood is empty then
12.        return 0;
13.    end
14.    while deg(v) > k v∈ μHopNeighborhood do
15.        k++;
16.        next v;
17.    end
18. end
19. return k-1;
```

Figure 16| Algorithm for μ-PCI calculation. Notice that we only considered two possible values for μ in our algorithm; μ=1 and μ=2.

3.5 Network Measures

Besides the centrality metrics we included other more global metrics to obtain the statistics and draw our conclusions. The above centrality measures are used to retrieve the importance of each node in the network and to arrange them in order of their importance. Then this arrangement is used to choose which node to remove in a targeted attack. More global measures are used to characterize the whole complex system. Below we describe these global metrics to clarify their concept.

Clustering Coefficient

The Clustering Coefficient of a network measures the average probability that two neighbors of a vertex are themselves adjacent. The local clustering coefficient C_i of a vertex i ∈ N is defined as [5]:

$$C_i = \frac{(\text{number of pairs of neighbors of i that are adjacent})}{(\text{number of pairs of neighbors of i})}$$

The global clustering coefficient *C* for the whole network is defined as the mean value of all C_i, i.e.

$$C = \frac{1}{N}\sum_{i=1}^{N} C_i \ , \quad C \in [0,1]$$

Diameter of Network

The diameter d of a network is the maximum eccentricity of any vertex in the graph; d is the greatest distance between any pair of vertices or, alternatively,

$$d = max_{v \in V}\ e_v \ .$$

To find the diameter of a graph, first find the shortest path between each pair of vertices and then consider the greatest length of any of these paths as the diameter of the graph [25].

3.6 Attack Types

The robustness of the network can be considered as its resistance to disconnection of its components and its throughput under node and link removal. Before the removal takes place, centrality measures have to be obtained for each node in the network. After degree, betweenness, eccentricity, eigenvector, closeness, k-shell or μ-PCI centrality is calculated for each node, all nodes are arranged in decreasing rank of the specified centrality measure. This is a conduct process so that the individual importance of each node in the network can be highlighted. As a next step the network gets "pruned", while removing one by one the nodes in the latter described order. During this reduction we study the effect this process has on the largest component of the network, i.e. we calculate the robustness. Several types of strategic node attack exist.

The first attack type to note is the simplest and unlikely to occur when targeted attack is aimed. In this case, nodes are deleted together with the edges connected to them in a random order, true to its name, regardless to their centrality values.

In situations where the goal is to force targeted attacks, simultaneous attacks are the first to be taken under consideration. Simultaneous targeted attacks describes a situation where the centrality measure is calculated firstly for all nodes and then nodes along with their edges are removed in the order of the centrality measure, from the highest to the lowest.

A second approach for coordinated attacks is more complex than the latter and known as the sequential target attack. Sequential target attack is a malicious attack and a situation where the centrality measure for all nodes of the initial network is calculated and then the "attacker" picks the node with the highest value. In order to delete the next node, centrality measures have to be calculated from scratch because now the network under consideration has changed. As a result the role and significance of each node in the network has changed too and this it is wise to recalculate the centralities to highlight the most powerful node.

4 Evaluation

4.1 Simulation model

We developed a graph simulation model based on Gephi, an open source interactive visualization and exploration platform suitable for complex networks analysis. We also developed a user interface shown in Figure 17. For this purpose we used NetBeans IDE (Integrated Development Environment) and java version java1.7.0_03. Our system has the following features: 6GB RAM and IntelCore i7 950 @ 3,7 GHz. We also had to increase heap size of this project up to 1MB in order to avoid Java's Garbage Collection during graph initialization, especially when the graph was dense.

In order to retrieve the k-cores and k-shells of the graph we developed a distributed algorithm for Hadoop's MapReduce. We used the fully distributed version of Hadoop; one master node and a pair of slaves were provided for this purpose. Hadoop was installed on a blade server and on each Daemon CentOS (Community ENTerprise Operating System) was installed. Also each Daemon has a 30GB disk, a 12GB RAM and 8 cores.

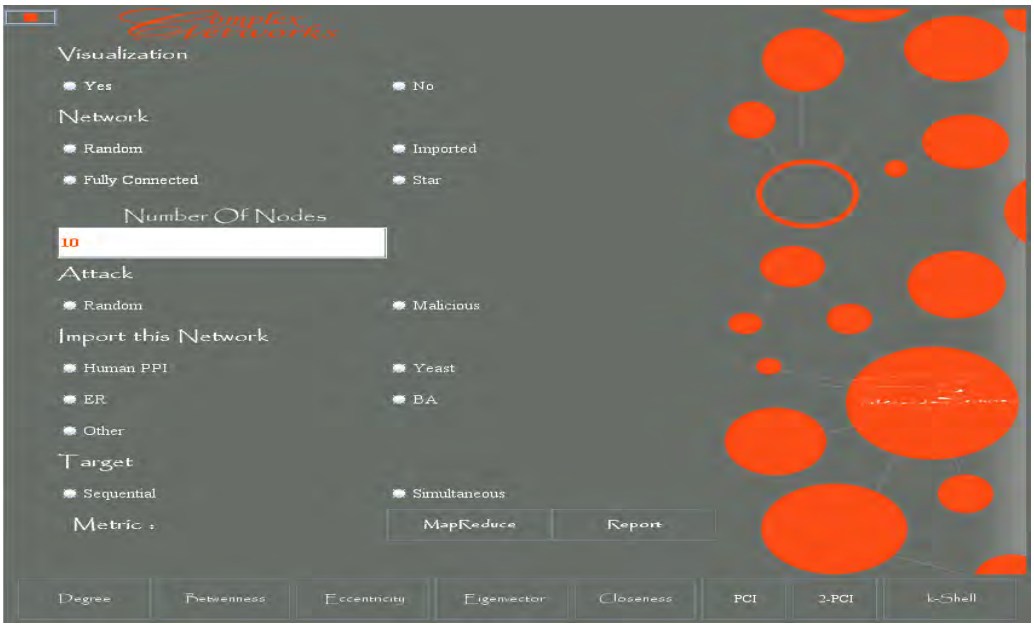


Figure 17| Our application Interface

4.2 Our Network Models

The networks we used for our experiment were both empirical biological networks and synthetic ones. A Protein-to-Protein interaction network in budding Yeast (Yeast) and a Human Disease Network (HDN) were included in the category of empirical biological networks (the original datasets can be found in [27] and [28]). Barabási –Albert (BA) and Erdős–Rényi (ER) networks were used to examine the robustness of synthetic networks that share similar features with the empirical ones. The latter networks were generated with CentiBin software. All networks have different sizes and attributes.

4.2.1| Protein-to-Protein Interaction Network in budding Yeast

The Protein-to-Protein Interaction Network in budding Yeast is an empirical network that consists of 2361 nodes and 7182 edges. For better visualization we provide the network structure in Figure 17 that follows. In Figure 20 correlations between centrality measures are shown for the Yeast Protein Network. In Figures 18 and 19 the distribution of each centrality is provided.

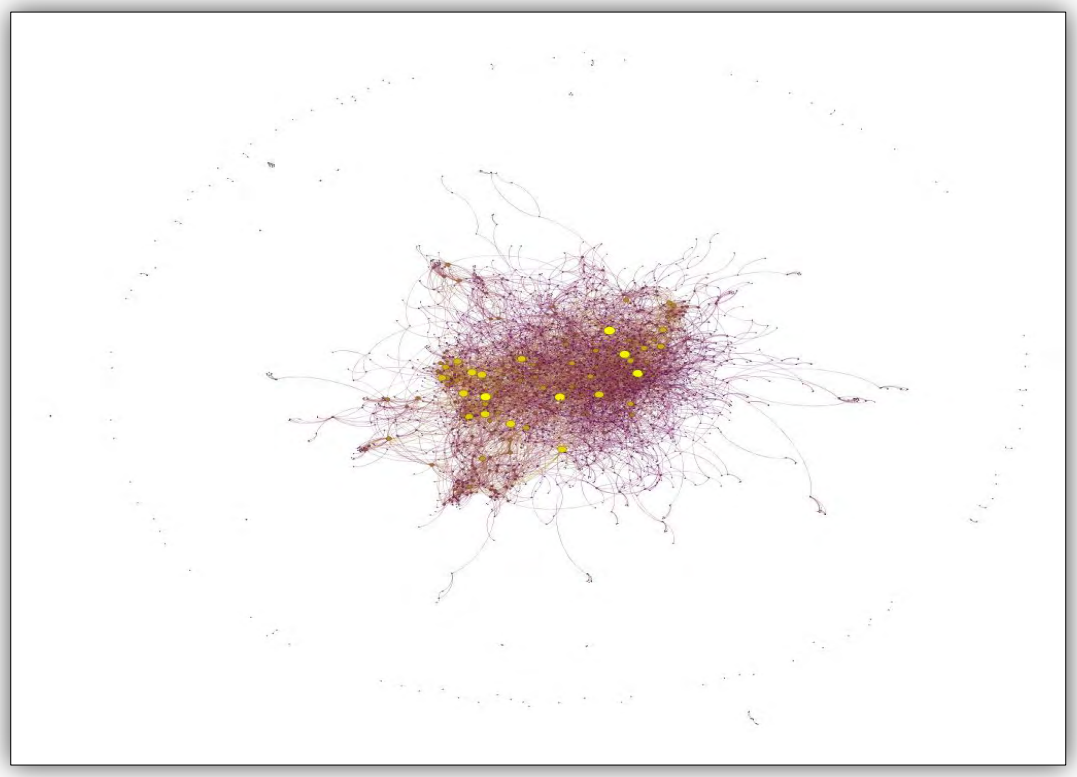


Figure 18| The Yeast Protein Interaction Network – Proteome with 2361 nodes and 7182 edges The network is created with Gephi visualization software. For this purpose and for our experiment dataset that is provided in the Gephi wiki is used. Dataset can be found here **http://wiki.gephi.org/index.php/Datasets**

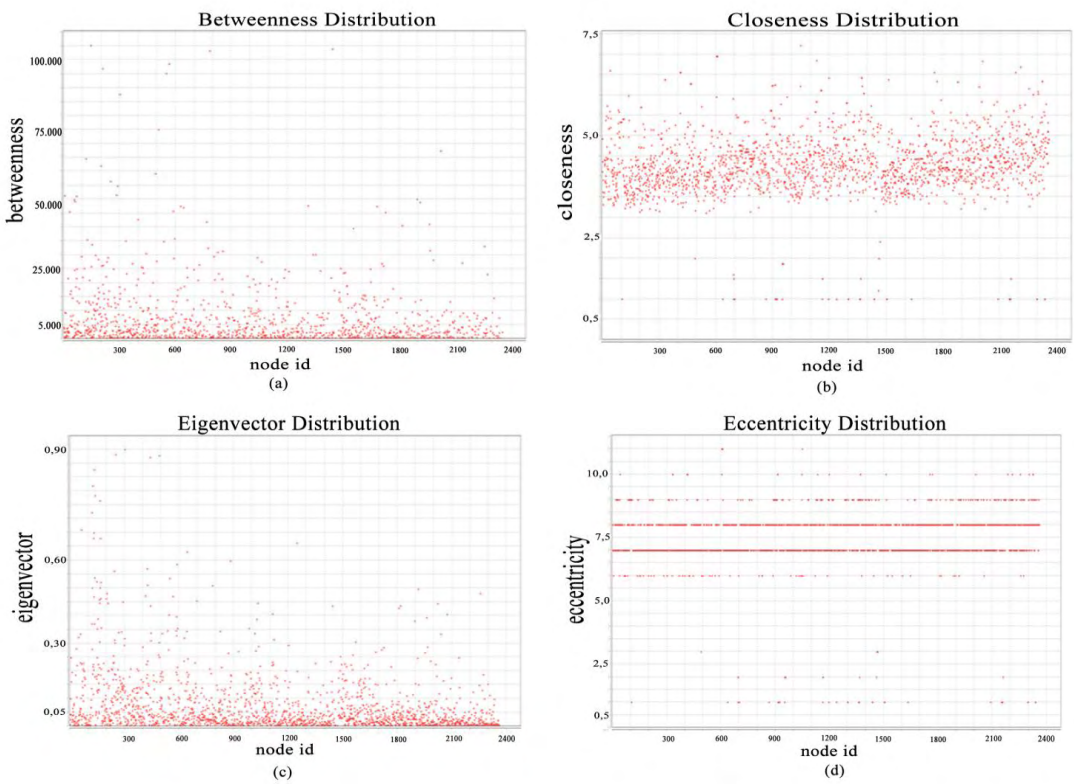


Figure 19| Centrality measures Distributions of Protein-to-Protein Interaction Network in budding Yeast with 2361 nodes and 6646 edges. (a) betweenness; (b) closeness; (c) eigenvector; (d) eccentricity respectively.

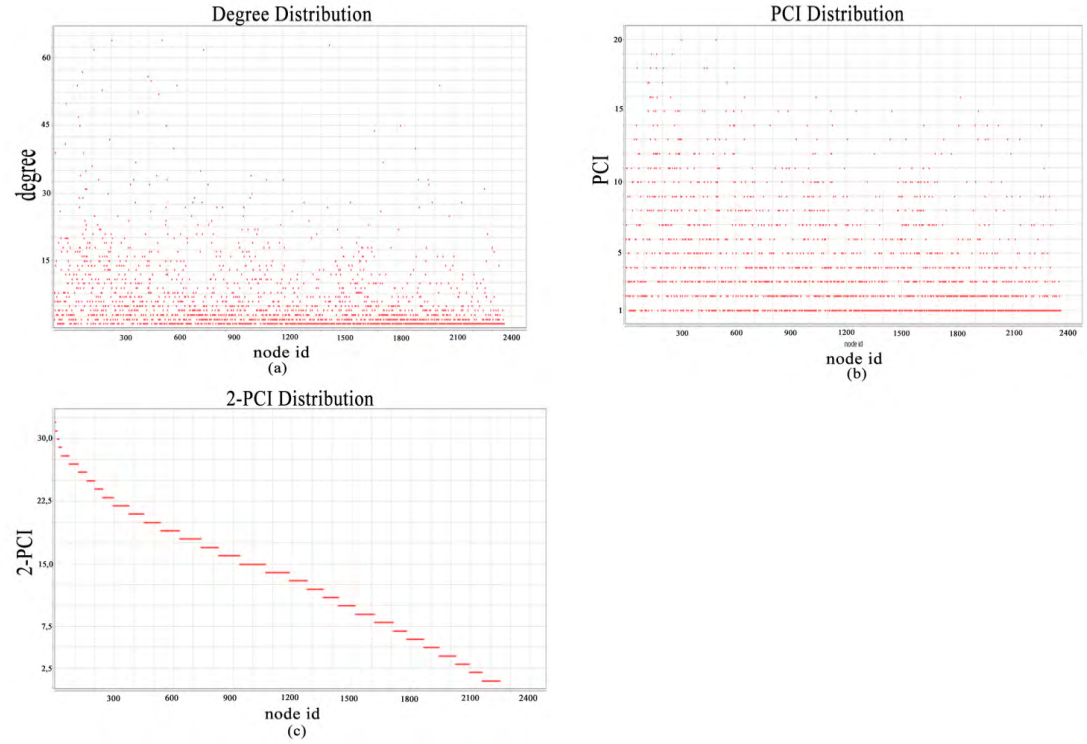


Figure 20| Centrality measures Distributions of Protein-to-Protein Interaction Network in budding Yeast with 2361 nodes and 6646 edges. (a) degree; (b) PCI; (c) 2-PCI respectively.

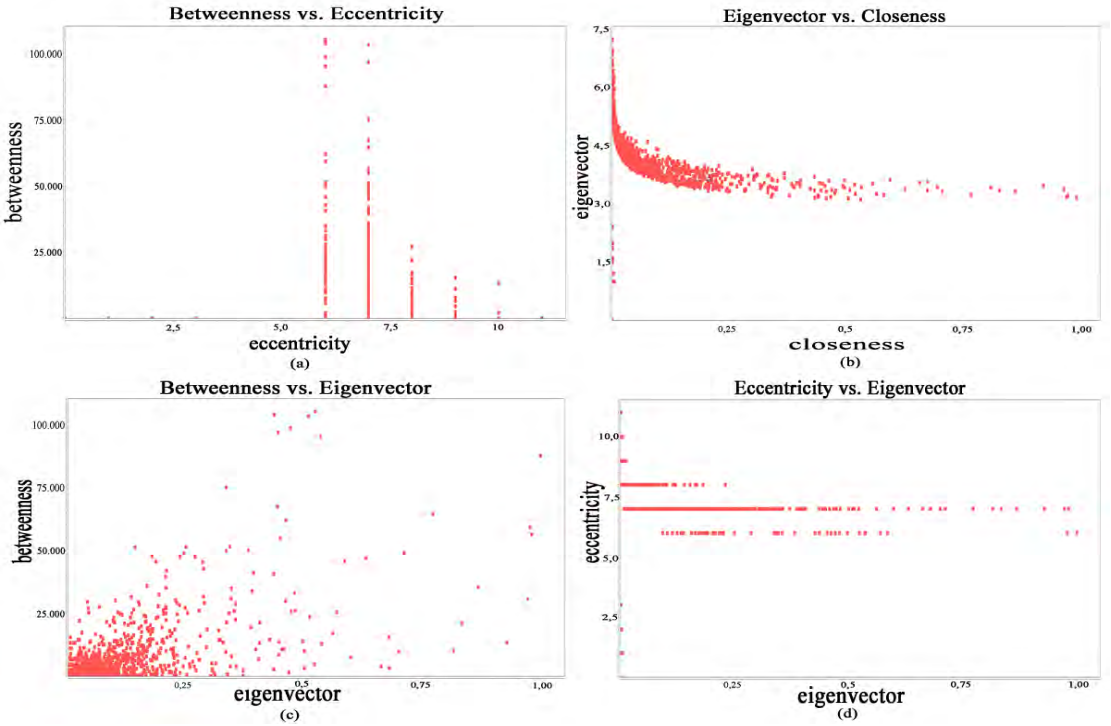


Figure 21| Correlations between centrality measures of Protein Interaction Network in budding Yeast with 2361 nodes and 6646 edges. (a) betweenness versus eccentricity; (b) betweenness versus eigenvector; (c) eccentricity versus eigenvector; (d) eigenvector versus closeness

4.2.2| The Human Disease Network

The Human Disease Network is an empirical network that consists of 7533 nodes and 22052 edges. For better visualization we provide the network structure in Figure 21 that follows. Figure 24 shows the correlations between centrality measures for the Human Disease Network. The distribution of each centrality is provided in Figures 22 and 23.

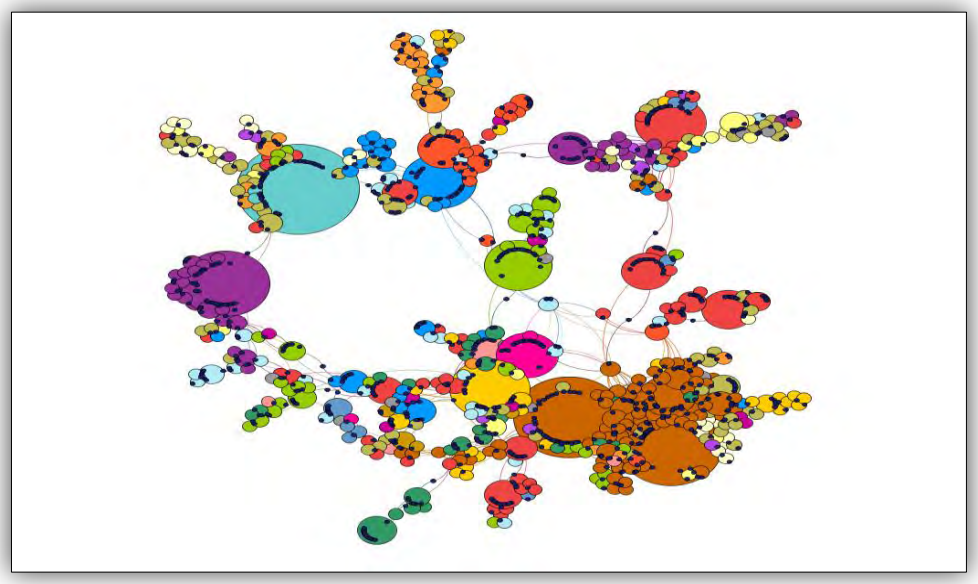


Figure 22| The Human Disease Network – Diseasome The network is created with Gephi visualization software. For this purpose and for our experiment dataset that is provided in the Gephi wiki is used. Dataset can be found here **http://wiki.gephi.org/index.php/Datasets**

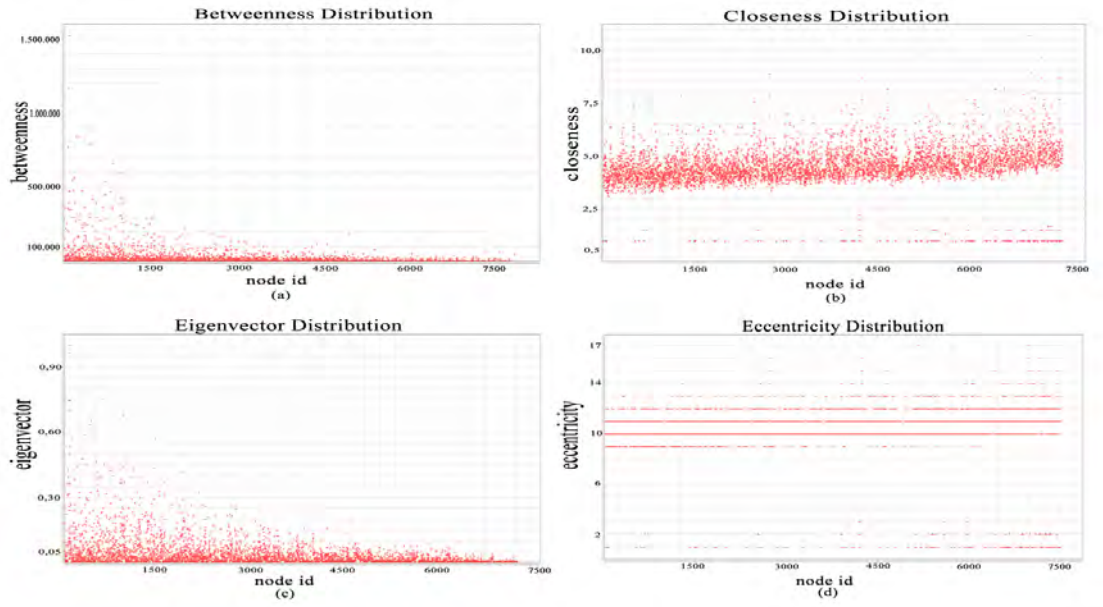


Figure 23| Centrality measures Distributions of the Human Disease Network with 7533 nodes and 22052 edges. (a) betweenness; (b) closeness; (c) eigenvector; (d) eccentricity respectively.

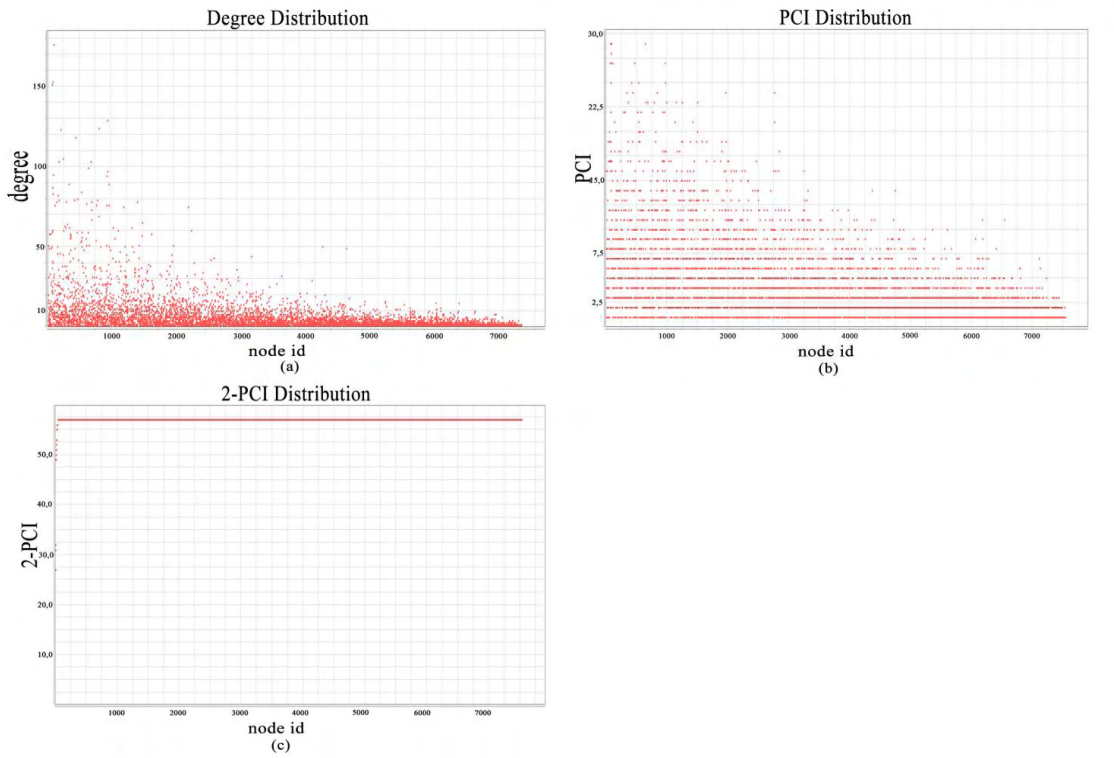


Figure 24| Centrality measures Distributions of the Human Disease Network with 7533 nodes and 22052 edges. (a) degree; (b) PCI; (c) 2-PCI respectively.

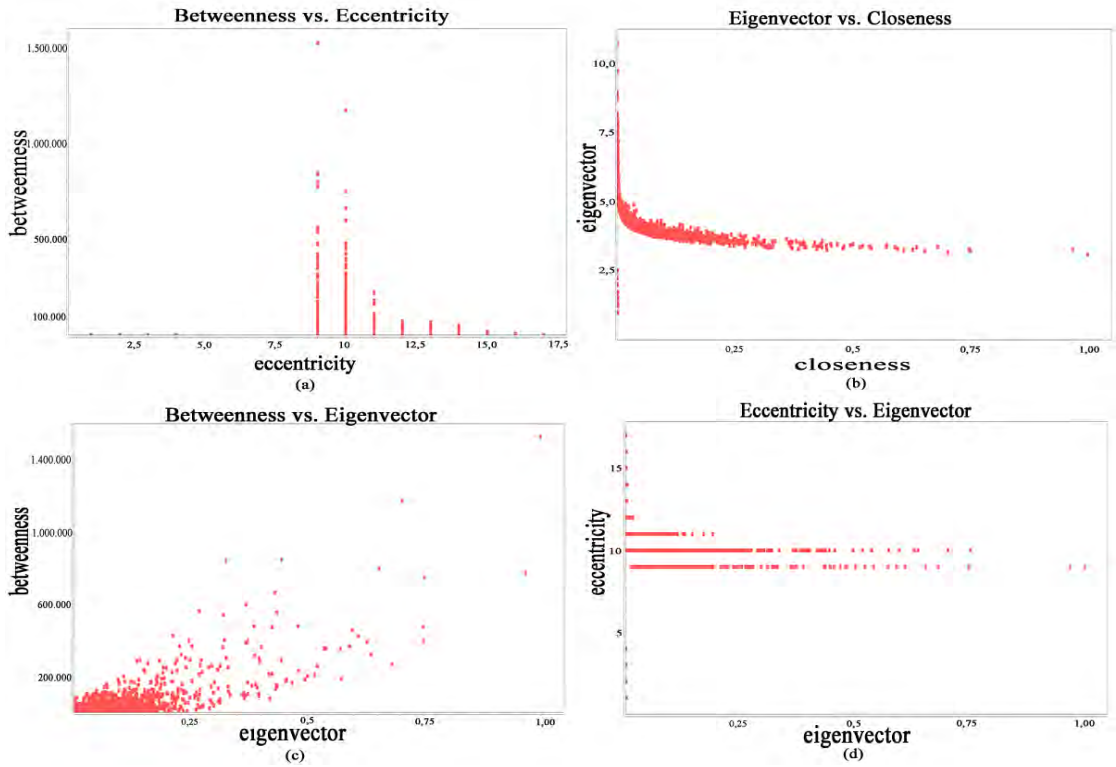


Figure 25| Correlations between centrality measures of the Human Disease Network with 7533 nodes and 22052 edges. (a) betweenness versus eccentricity; (b) betweenness versus eigenvector; (c) eccentricity versus eigenvector; (d) eigenvector versus closeness

4.2.3| The Barabási–Albert Network Model

The Barab*á*si–Albert Network is a synthetic network that consists of 5001 nodes and 14947 edges. We used a random seed equal to 1212 and we processed 5000 iterations to retrieve this network.

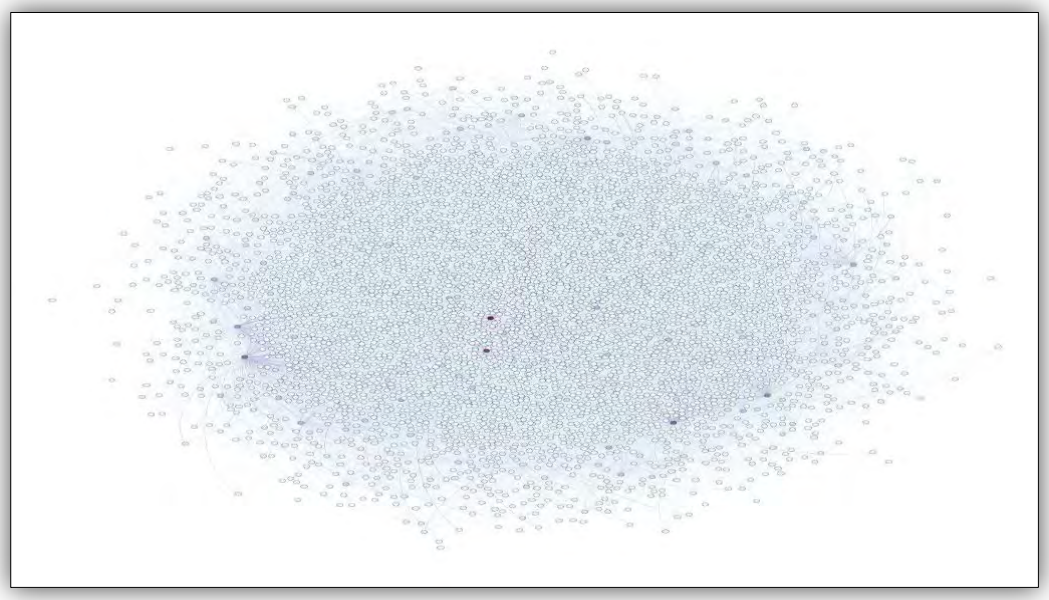


Figure 26| The Barabási–Albert Network with 5001 nodes and 14947 edges. The dataset for the graph was created with CentiBin and the visualization was created with Gephi. The purple nodes are hub nodes.

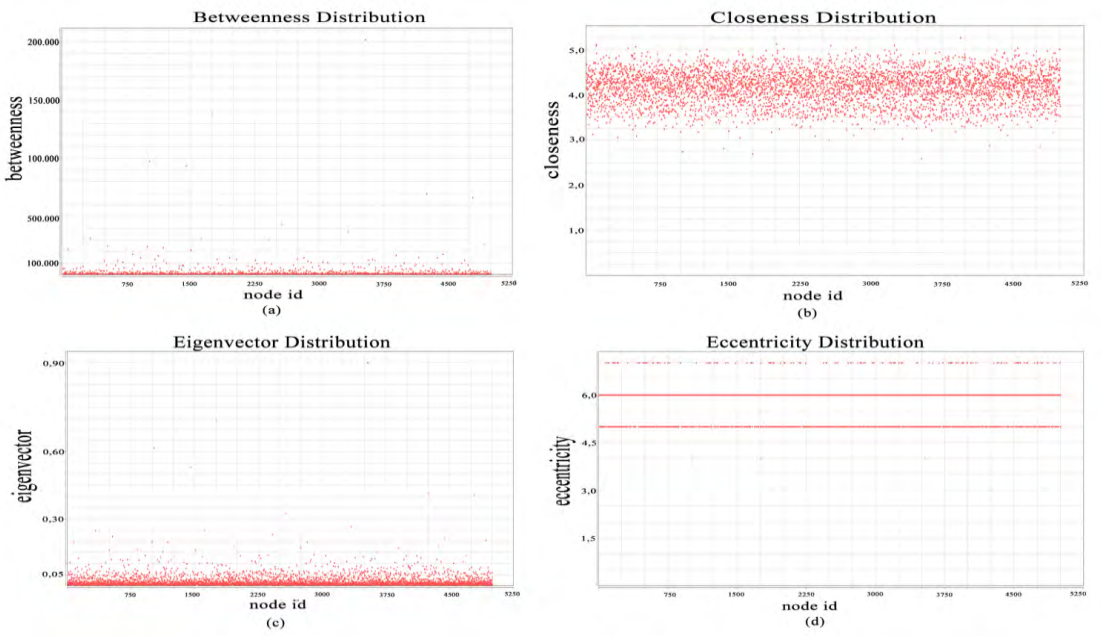


Figure 27| Centrality measures Distributions of the Barabási–Albert Network with 5001 nodes and 14947 edges. (a) betweenness; (b) closeness; (c) eigenvector; (d) eccentricity respectively.

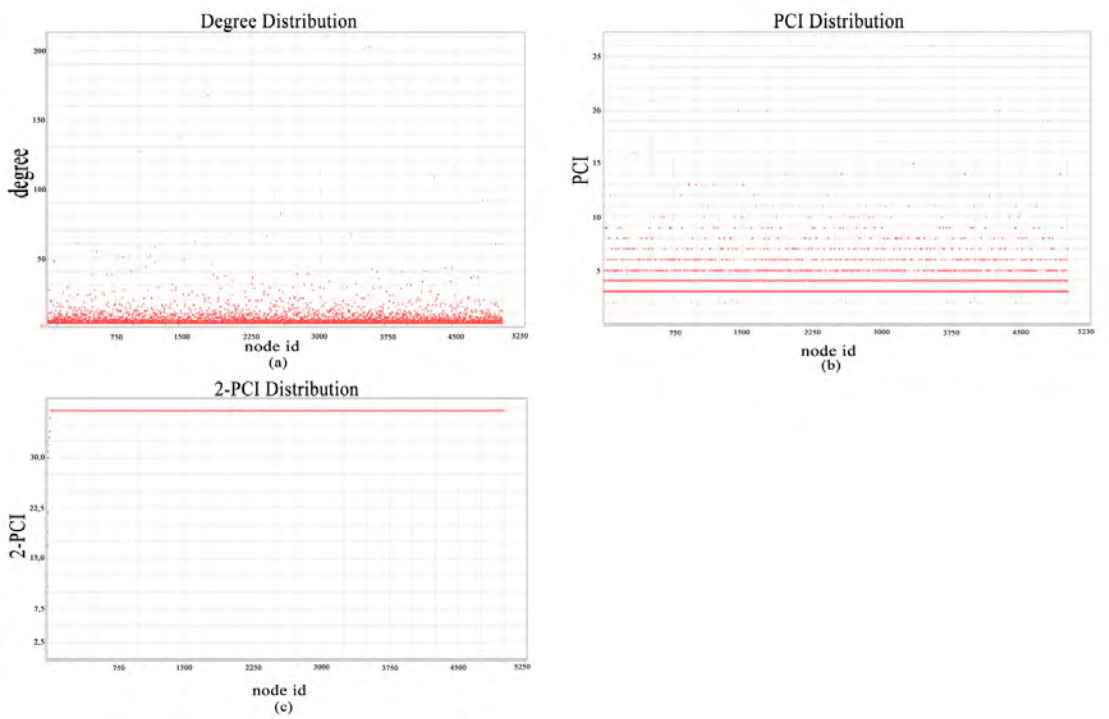


Figure 28| Centrality measures Distributions of the Barabási–Albert Network with 5001 nodes and 14947 edges. (a) degree; (b) PCI; (c) 2-PCI respectively.

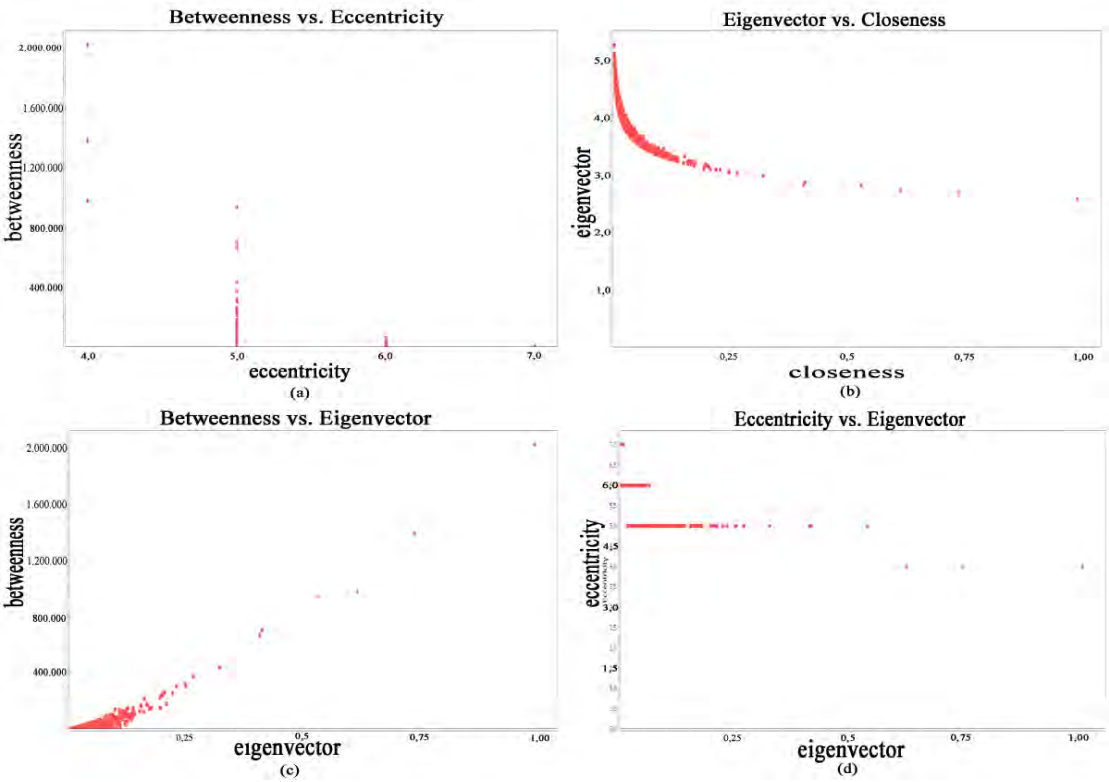


Figure 29| Correlations between centrality measures of the Barabási–Albert Network with 5001 nodes and 14947 edges. (a) betweenness versus eccentricity; (b) betweenness versus eigenvector; (c) eccentricity versus eigenvector; (d) eigenvector versus closeness.

4.2.4| The Erdős-Rényi Network Model

The Erdős-Rényi Network is a synthetic network that consists of 5000 nodes and 12536 edges. We used an edge probability equal to 0.001to retrieve this network.

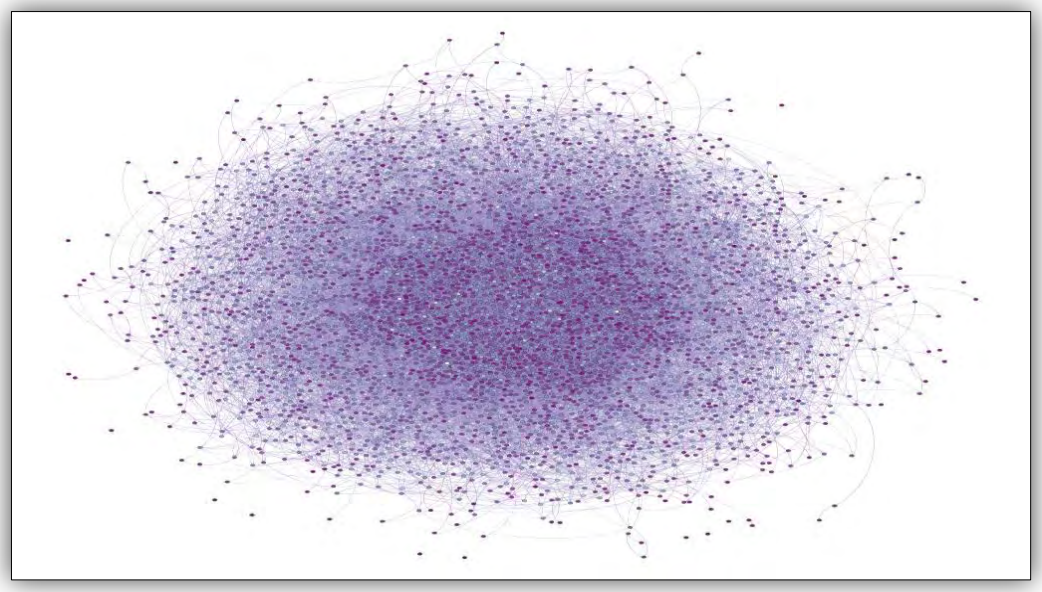


Figure 30| The Erdős-Rényi Network with 5000 nodes and 12536 edges. The dataset for the graph was created with CentiBin and the visualization was created with Gephi.

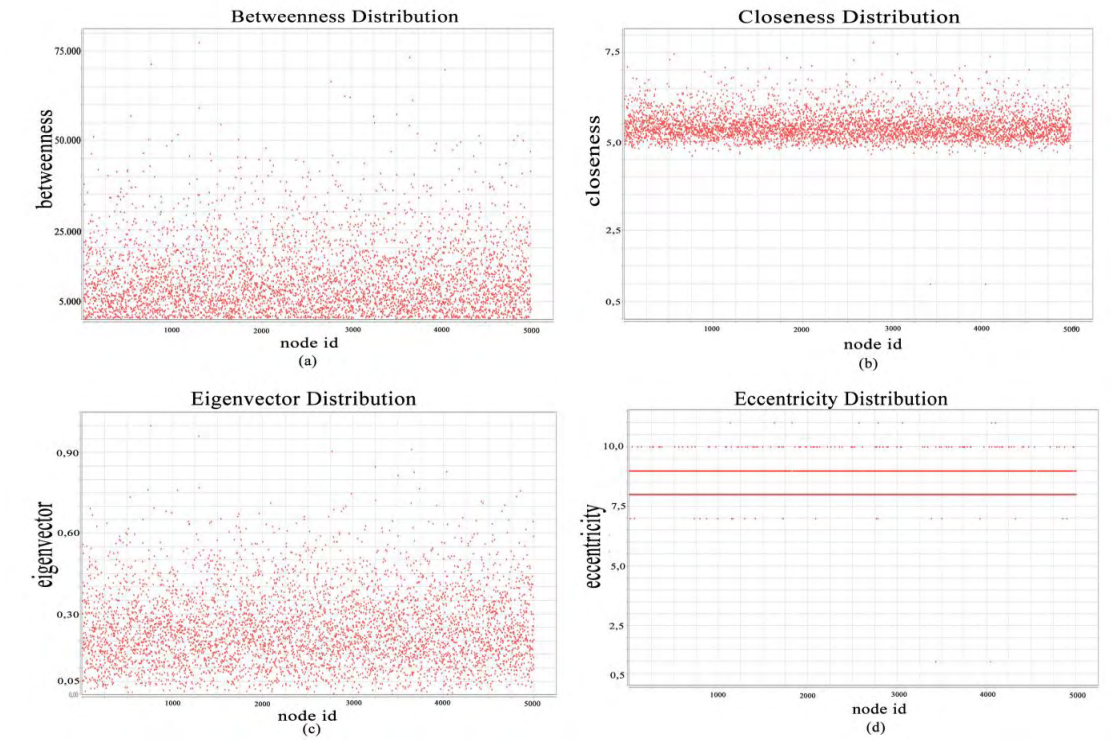


Figure 31| Centrality measures Distributions of the Erdős-Rényi Network with 5000 nodes and 12536 edges. (a) betweenness; (b) closeness; (c) eigenvector; (d) eccentricity respectively.

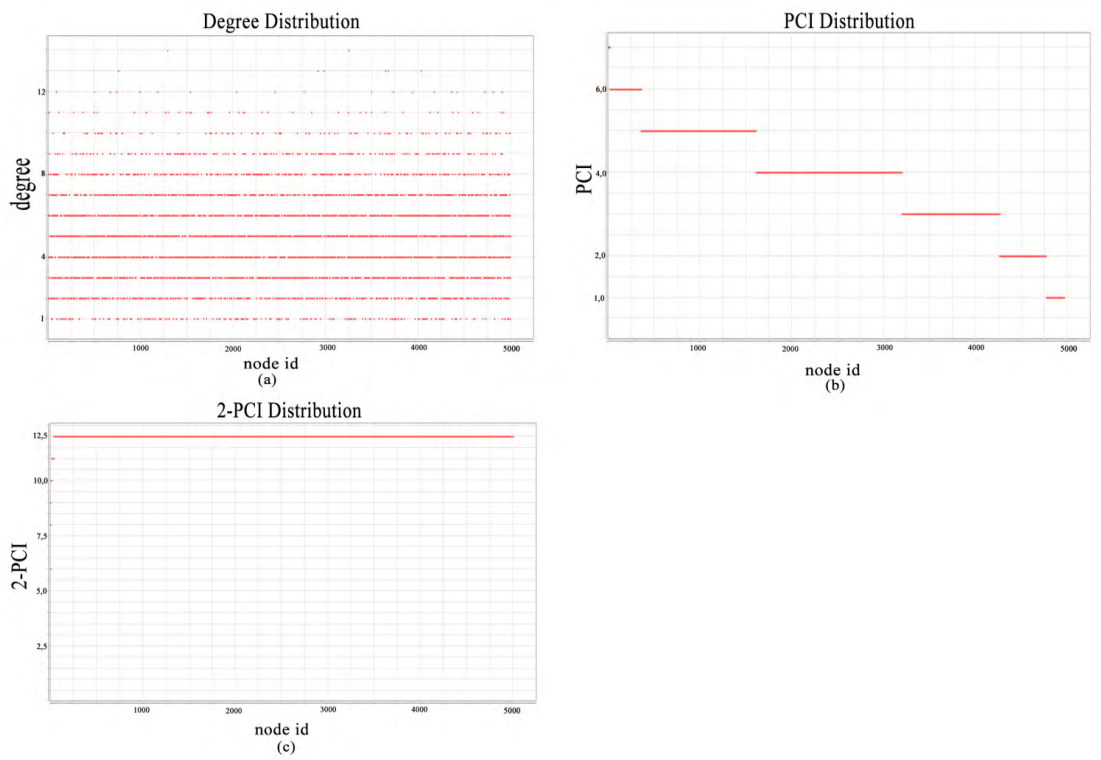


Figure 32| Centrality measures Distributions of the Erdős-Rényi Network with 5001 nodes and 12536 edges. (a) degree; (b) PCI; (c) 2-PCI respectively.

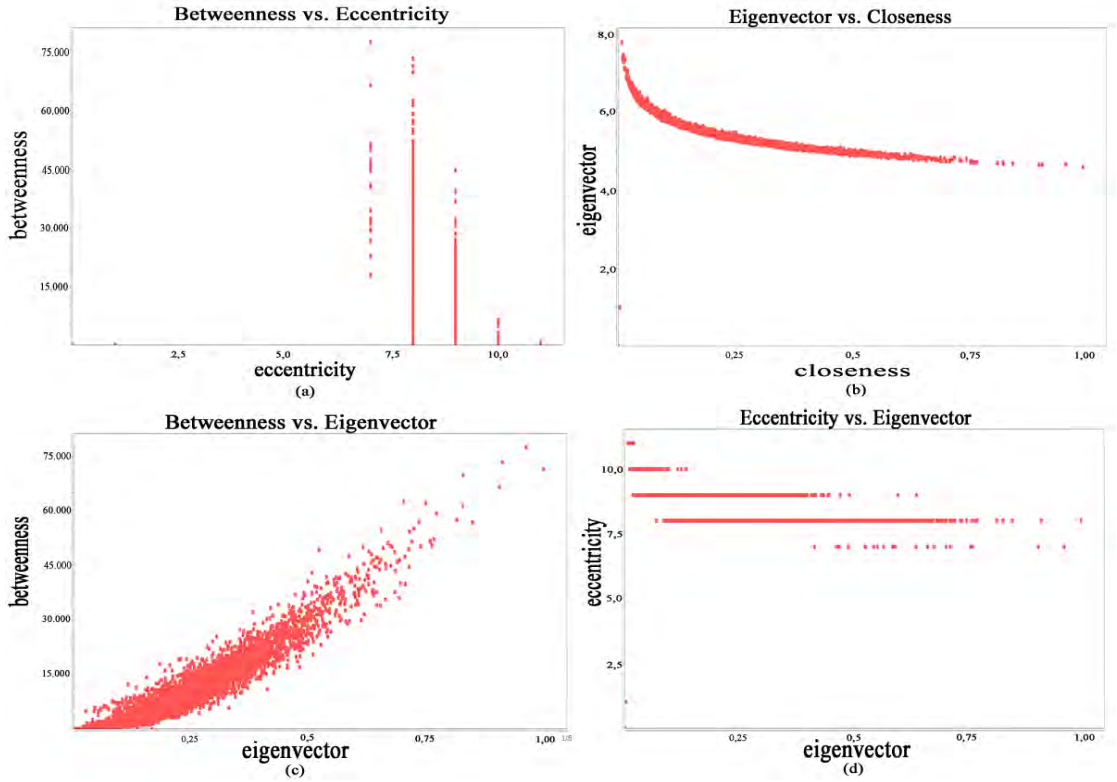


Figure 33| Correlations between centrality measures of Erdős-Rényi Network with 5000 nodes and 12536 edges. (a) betweenness versus eccentricity; (b) betweenness versus eigenvector; (c) eccentricity versus eigenvector; (d) eigenvector versus closeness.

Before we start to describe our results we first have to provide some important insights of the network models that have been included in our work. Table 1 and 2 that follow store the characteristics of each of them.

Network Models and Features

Network Type	Number of Triangles	Diameter	Average Path length	Number of shortest paths	Number of k-cores
Yeast	3530	11	4.376	4.944.096	10
HDN	7043	17	4.629	52.976.918	10
BA	341	7	4.169	25.005.000	2
ER	20	11	5.455	24.656.192	3

Table 1| Table containing all networks along with their characteristics

Network Models and Features

Network Type	Short Description	Undirected	#V	#E	Max Largest Component	Max Avg Clustering Coefficient
Yeast	Protein-to-Protein Interaction Network in budding Yeast	✔	2.361	6.646	2.224	0.200
HDN	A network of disorders and disease genes linked by known disorder-gene associations, indicating the common genetic origin of many diseases [http://wiki.gephi.org/index.php/Datasets]	✔	7.533	22.052	7.279	0.106
BA	Barabási–Albert network, edge probability=0.0001.The network was generated with CentiBin.	✔	5.001	14.947	5.001	0.007
ER	Erdős-Rényi network, random seed=1212 & number of iterations=200.The network was generated with CentiBin.	✔	5.000	12.536	4.966	0.001

Table 2| Table containing all the networks along with their characteristics

4.3 Attack Performance Evaluation

4.3.1| Simultaneous Target Attack

Our first experiment includes simultaneous target attacks. In this section we study the effect of node deletion on the largest component of the network. The node and edge removal follows a simultaneous strategy; this means that degree, betweenness, eccentricity, closeness, eigenvector, PCI, 2-PCI and k-shell values are calculated for all nodes and nodes are then deleted in descending order of their centralities in each case. This process was applied to a variety of network models, such as scale-free networks (BA), a yeast protein-to-protein interaction network (Yeast), a network of disorders and disease genes (HDN) and a random network (ER). It is important to notice that in case of the k-shell, only simultaneous attack is performed.

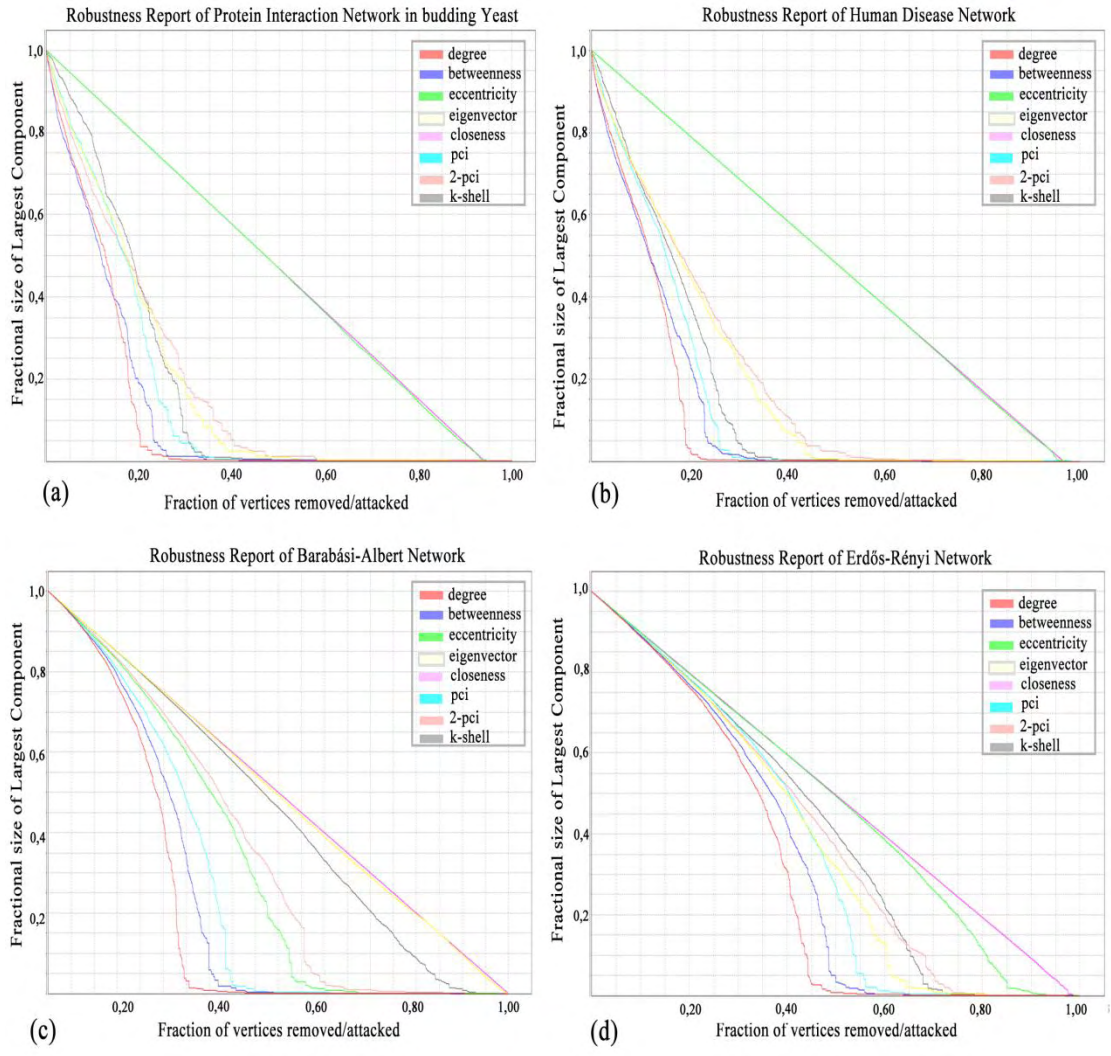


Figure 34| Robustness against simultaneous target attack of (a) the Protein Interaction Network in budding Yeast (b) the Human Disease Network, (c) Barabási–Albert Network and (d) the Erdős-Rényi Network

Figure 34 shows robustness results for each network used in our study against attacks by each centrality measure aforementioned. It is clear from the graphs that performing simultaneous target attacks according to the degree centrality turns the network into a less robust but more vulnerable one, in each studied case. Hence, simultaneous degree attack is the most effective way to decompose the network and destroy the consistency of its structure. Simultaneous target attack according to betweenness centrality seems to be the second most effective attack in order to degrade the network structure.

Eigenvector centrality has a medium impact on the robustness of real-world biological networks compared to the other centralities of this study and a medium to good impact on synthetic ones. The same can be assumed when performing the simultaneous strategic attack but now according to the k-shell values. Although different networks might have a different number of k-cores, the largest component size is medially to slightly low affected by node removal in each case.

An attacker is unlikely to achieve network destruction when trying to delete nodes in descending order of their closeness centrality simultaneously. In all situations, the network is decomposed in a very slow way and therefore it is less vulnerable to attacks than other metric values. Eccentricity attack has also a relatively low performance when the goal is to destroy the biological network. One can see in Figure 34 that in both in the PIN and the HDN the network maintains its total throughput in a very good level and in the less better in the ER and almost in an average level in the BA.

A very interesting outcome is that when attacking nodes of both biological and synthetic networks according to their PCI (μ=1) value, their individual components disconnect relatively rapid. Although all four networks nodes have different PCI values, PCI simultaneous target attack has very similar results on each one. This means that deleting nodes which have one-hop neighbors that are effective information spreaders, i.e. hub nodes, makes the network vulnerable to attacks. In real-world biological networks performing the same attack as the latter, but this time aiming nodes who have two-hop hub neighbors, has a less effective result. In this kind of attacks, the network appears to have a better resistance to structure degradation than the PCI attack. We believe that when performing simultaneous target attacks, 2-PCI has on the average a mediocre performance but is steadily medium in each network considered. Generally, apparent from the results 2-PCI manages to approximate the mean value[1] of robustness and vulnerability in the best way compared to all other six centralities that have been tested.

At this point let us examine the case of the HDN in a little more detailed way since its topological features affect the μ-PCI values in an interesting aspect. Previous studies show that hub proteins (nodes) are likely to be encountered in essential human genes and their role is dominant in the structure of the interactome [15]. On the other hand, non essential proteins have mostly a localized role in the network. Considering that μ-PCI is a localized measure, more informative than the degree and also not influenced by the isolated nodes we can observe that PCI reflects more the fragility against sequential removal of nodes with a more

[1] The average value of robustness is considered as the mean value of the robustness that is calculated when degree, eigenvector, eccentricity, closeness and k-shell simultaneous target attacks are performed.

powerful effect to their neighborhood than eccentricity or closeness centrality. Therefore, nodes that are attached to many hub nodes are removed and make hub nodes less powerful. In this situation essential genes become gradually more non essence nodes. Hub nodes play a central and functional role in the human interactome and hence making them more powerless inevitably destroys the underlying network structure. Taking under account the impact that PCI on the HDN has it is logical to assume that 2-PCI has less effect on the network degradation than PCI; the results are confirming this aspect.

We believe that the best way to attack simultaneously a biological network is to aim for nodes with the highest degree; not only is this the most efficient way but it is also the fastest computational way. This result is a consequence of the fact that high degree nodes are essential and play a central role in empirical networks and are hub nodes in synthetic networks and therefore when removing them and the links attached to them the network decomposes relatively fast. Specifically as mentioned in [1], Protein-to-Protein Interaction Networks have only a small number of hub nodes and the other nodes are not highly connected. Removing a hub protein (node) from the PIN network probably causes a fatal failure to the network's structure rather than removing a non-hub node, i.e. a node with poor connections to other nodes, which shall not have a great impact on the proteome. This rule we confirmed in our studies is known in Genome analysis as centrality-lethality rule.

PCI centrality can be used to highlight the fragility of the network against simultaneous target attacks, since it results a rapid decrease of the giant component. Removing nodes with high PCI values diminishes the cohesion of the network structure relatively fast. Less effective than the latter centrality attack, is to perform a simultaneous attack by a class higher μ (μ=2). It does provide less significant insights into biological system vulnerability than the PCI centrality. We believe that both power community indexes can be considered as a better option than other centrality measures, like closeness, to expose the vulnerability of the network and generally a good choice if the goal is to examine the average robustness compared to other centralities. Also, 2-PCI can be considered as the best option to approximate the average value of robustness[2] compared to all other centrality measures used in this work.

Regardless of the number of k-shells and k-cores and despite that we remove the most 'central' nodes in an onion-like shaped structure after retrieving the k-cores (see Figure 14 for better visualization and understanding of the structure decomposition), simultaneous targeted node deletion has a poor ability to capture both the robustness and the vulnerability of the network under consideration. It is clear though that robustness is better exposed for synthetic and vulnerability for empirical networks. Another observation is that the hierarchy of empirical networks comprises ten k-cores in contrast to the BA and ER in which two and three k-cores appear respectively.

In contrast, a rather inferior choice would be to perform a simultaneous target attack according to the descending order of closeness centrality. In this case nodes that are as close to all nodes as possible and not necessarily directly connected to them as in degree centrality are chosen for deletion. The result in this case is the opposite of what an attacker is actually

[2] The average value of robustness is considered as the mean value of the robustness that is calculated when degree, eigenvector, eccentricity, closeness and k-shell simultaneous target attacks are

aiming for. But when the goal is to provide the robustness of the underling network structure deleting nodes by decreasing order of their closeness centrality will be by far the best choice.

Robustness Report Simultaneous Target Attack								
Network	**Degree**	**Betweenness**	**Eccentricity**	**Eigenvector**	**Closeness**	**PCI**	**2-PCI**	**k-shell**
Yeast	0.109	0.114	0.442	0.169	0.444	0.149	0.174	0.174
HDN	0.107	0.119	0.466	0.186	0.467	0.140	0.194	0.156
BA	0.210	0.239	0.500	0.330	0.500	0.267	0.353	0.455
ER	0.301	0.327	0.471	0.372	0.493	0.358	0.393	0.404
max-min	0.194	0.213	0.058	0.203	0.056	0.218	0.219	0.299

Table 3| Robustness against simultaneous target attack by degree, betweenness, eccentricity, eigenvector, PCI, 2-PCI, closeness and k-shell of all networks. All values are rounded to the third digit.

Numerically speaking, we can confirm from results displayed in Table 3 that closeness has the same effect in almost each network and therefore the difference of the maximum and minimum value of robustness is very small (0.056). The same applies for the eccentricity centrality case (0.058). The lack of robustness against simultaneous target attack by degree is confirmed in the Degree column of Table 3.

Vulnerability Report Simultaneous Target Attack								
Network	**Degree**	**Betweenness**	**Eccentricity**	**Eigenvector**	**Closeness**	**PCI**	**2-PCI**	**k-shell**
Yeast	0.391	0.386	0.058	0.331	0.056	0.351	0.326	0.326
HDN	0.393	0.381	0.034	0.314	0.033	0.360	0.306	0.344
BA	0.290	0.261	0.000	0.170	0.000	0.233	0.147	0.045
ER	0.199	0.173	0.029	0.128	0.007	0.142	0.107	0.096
max-min	0.194	0.213	0.058	0.203	0.056	0.218	0.219	0.299

Table 4| Vulnerability against simultaneous target attack by degree, betweenness, eccentricity, eigenvector, PCI, 2- PCI, closeness and k-shell of all networks. All values are rounded to the third digit.

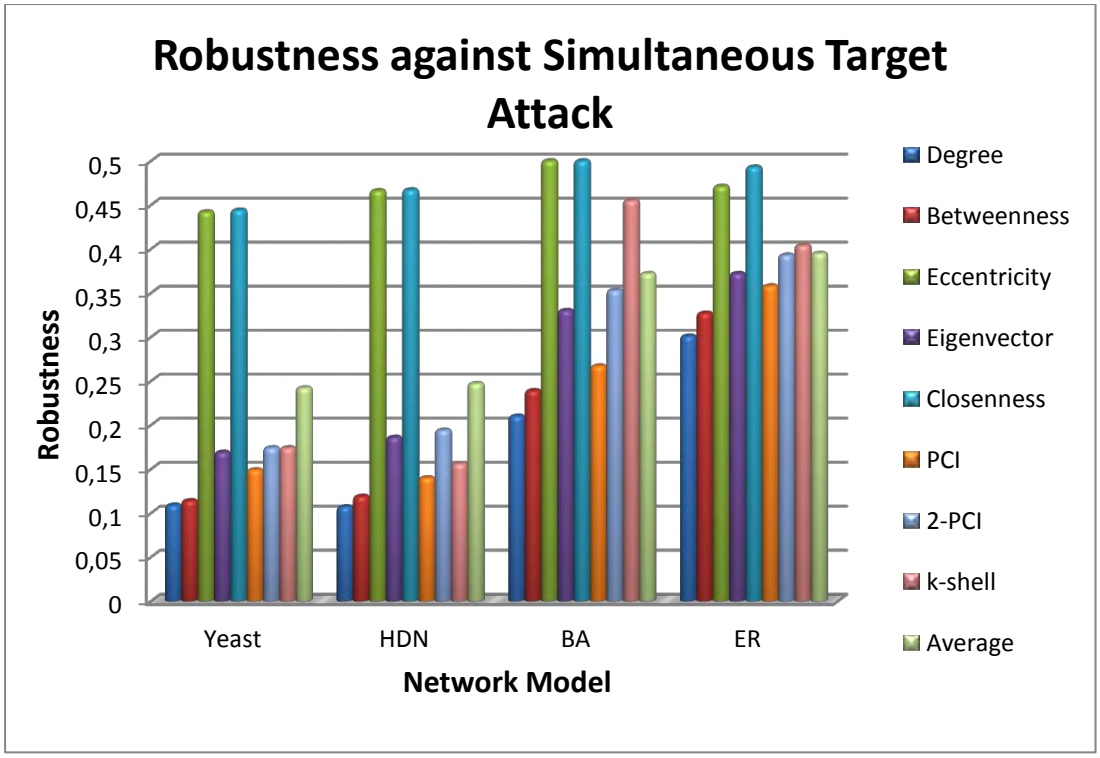


Figure 35| All network models and their robustness against Simultaneous Target Attacks. Average Robustness is also shown.

4.3.2| Sequential Target Attack

Our second experiment includes sequential target attacks. We study the effect on the largest component of the network node deletion has. The node removal follows a sequential strategy; this means that degree, betweenness, eccentricity, closeness, eigenvector, PCI and 2-PCI values are calculated for all nodes and they are then deleted in descending order of their centrality values. A next step is to consider the network N_p that results from deletion of a fraction of ρ nodes as a new network and force a new round of calculations. This process is repeated until no nodes are left and highlights each time the most 'central' node. We applied sequential targeted removals to a variety of network models, such as a scale-free network (BA), a yeast protein-to-protein interaction network (Yeast), a network of disorders and disease genes (HDN) and a random network (ER).

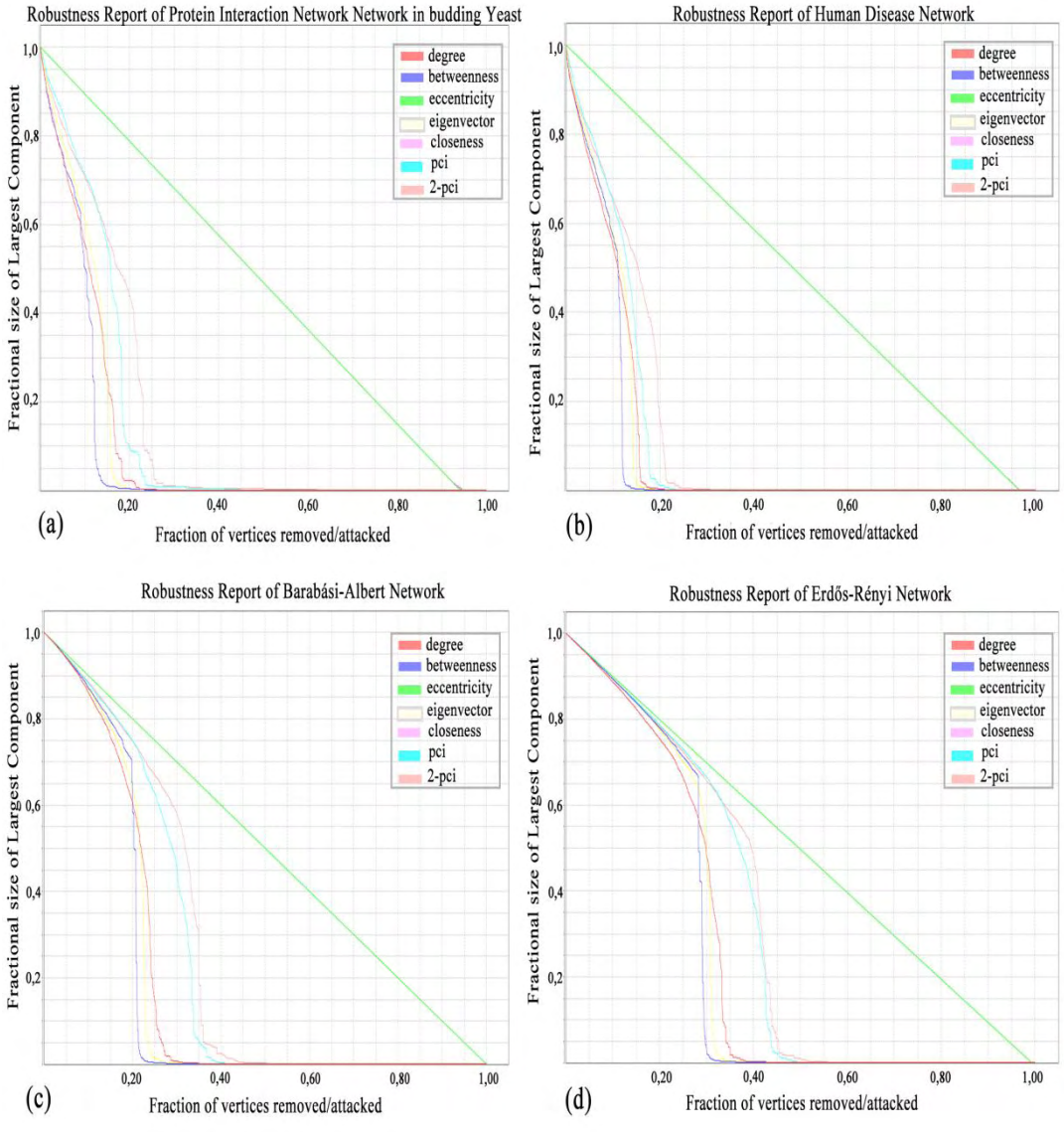


Figure 36| Robustness against simultaneous target attack of (a) the Protein Interaction Network in budding Yeast (b) the Human Disease Network, (c) Barabási–Albert Network and (d) the Erdős-Rényi Network

Figure 36 shows that the same networks behaves differently under sequential target attack when compared to simultaneous. Sequential target attack has on average greater impact than simultaneous targeted attack by almost each type of centrality measure. The preceding results can be qualitatively understood considering that in each round of the attacks the most important node according to the centrality values is removed along with its connections. Hence, this procedure reduces the resistance of the network to failures and degrades it rapidly.

Specifically, it is obvious from the diagram that sequential attack by both eccentricity and closeness can highlight the ability of a network to maintain its total throughput under node and link removal better than other centralities. In both cases the biological complex systems under consideration disconnect their components quiet hardly; we observe that only after removing 45-50% of the nodes the size of the giant component is cut in halve.

Conversely, the outcomes for the robustness when targeting nodes according to other centralities are significantly different. Betweenness centrality appears as the superior choice in order to attack a biological network in a sequential process. This aspect is respected to result as a node with a high betweenness level is the node which achieves the maximum number of shortest paths from all nodes to all others that pass through that node. If this node is deleted then we delete concurrently the connections between other nodes. As mentioned previously, protein networks tend to have hub nodes that play a crucial role in its vital functionality. When removing hub nodes and particularly the most important that is placed in between other nodes the network degrades rapidly.

We also observe that in contrast to simultaneous target attacks, degree has lost here its ability to expose vulnerability in the same superior level as in the previous category of failures. PCI and 2-PCI values as a measure of robustness manage to capture again the best approximation of the mean value[3] of both robustness and vulnerability compared to all other values of robustness calculated in the other cases of centralities; 2-PCI maintains a better approach of the average robustness value than PCI. PCI again exhibits in a better order the vulnerability of the network against malicious failures than 2-PCI. Of course, this can be understood considering that in each round of deletion the most important spreader in the network is removed; after all, μ-PCI tries to discover the node that can achieve the maximum influence to other nodes.

Finally, sequential targeted node removal by eigenvector centrality reaches again as in simultaneous attacks a mediocre performance in determining either the robustness or the vulnerability.

We believe that when the goal is to capture the robustness of a network eccentricity and closeness are the best choice. On the other hand, node deletion by decreasing betweenness centrality order is more likely to expose vulnerability against sequential attacks. If the case is to define the average robustness and vulnerability of malicious failures μ-PCI, for $\mu=1$ and even better for $\mu=2$, meets the best conditions to capture the throughput of the network under node removal.

[3] The average value of robustness is considered as the mean value of the robustness that is calculated when degree, eigenvector, eccentricity and closeness sequential target attacks are performed.

Robustness Report							
Sequential Target Attack							
Network	**Degree**	**Betweenness**	**Eccentricity**	**Eigenvector**	**Closeness**	**PCI**	**2-PCI**
Yeast	0.097	0.081	0.444	0.099	0.444	0.129	0.146
HDN	0.093	0.083	0.467	0.093	0.467	0.110	0.127
BA	0.193	0.179	0.500	0.187	0.500	0.252	0.270
ER	0.253	0.242	0.493	0.251	0.493	0.314	0.321
max-min	0.160	0.161	0.056	0.158	0.056	0.204	0.194

Table 5| Robustness against simultaneous target attack by degree, betweenness, eccentricity, eigenvector, PCI, 2-PCI, closeness and k-shell of all networks. All values are rounded to the third digit.

Numbers in Table 5 and 6 confirm our previous opinions about the classification of the strategies that should be followed in order to provide the robustness and vulnerability of the network under sequential node deletion. Particularly, every centrality captures poor robustness, except from eccentricity and closeness. Observing the last row we can see that although degree, betweenness, PCI and 2-PCI have a negative impact on the robustness they have a different effect on each network; therefore we can see that the difference from the maximum and minimum value of robustness is in a range from 0.160 up to 0.204. On the other hand, eccentricity and closeness manage to affect the size of the largest component in a very similar way in each network model during node and edge removal.

Vulnerability Report							
Sequential Target Attack							
Network	**Degree**	**Betweenness**	**Eccentricity**	**Eigenvector**	**Closeness**	**PCI**	**2-PCI**
Yeast	0.403	0.419	0.056	0.401	0.056	0.371	0.354
HDN	0.407	0.417	0.033	0.407	0.033	0.390	0.373
BA	0.307	0.321	0.000	0.313	0.000	0.248	0.230
ER	0.247	0.258	0.007	0.249	0.007	0.186	0.179
max-min	0.160	0.161	0.056	0.158	0.056	0.204	0.194

Table 6| Vulnerability against simultaneous target attack by degree, betweenness, eccentricity, eigenvector, PCI, 2-PCI, closeness and k-shell of all networks. All values are rounded to the third digit.

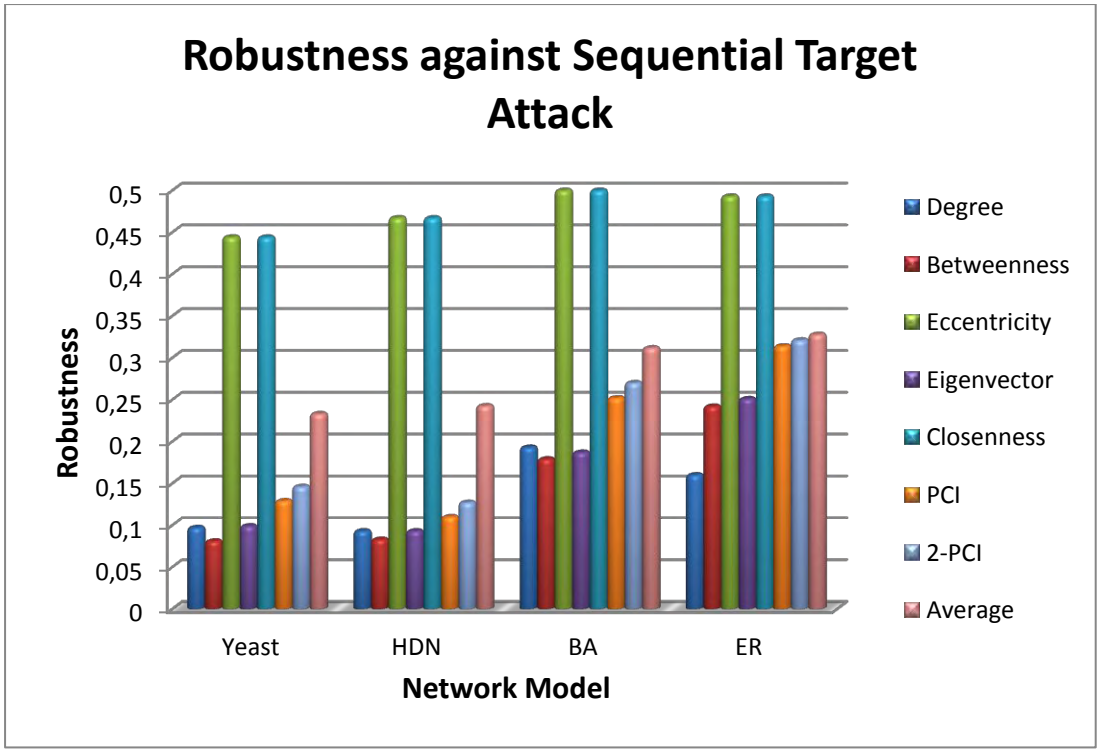


Figure 37| All network models and their robustness against Sequential Target Attacks. Average Robustness is also shown.

4.3.2| Random Attack

In order to study the ability of a biological complex system to maintain its throughput under random node attack and to reach reliable conclusions for the average case, we performed several rounds of random attacks on each network. The outcomes are shown in Figure 38. We are able to compare the results without any concerns about inconsistency since robustness is calculated in percentage and largest component size is normalized.

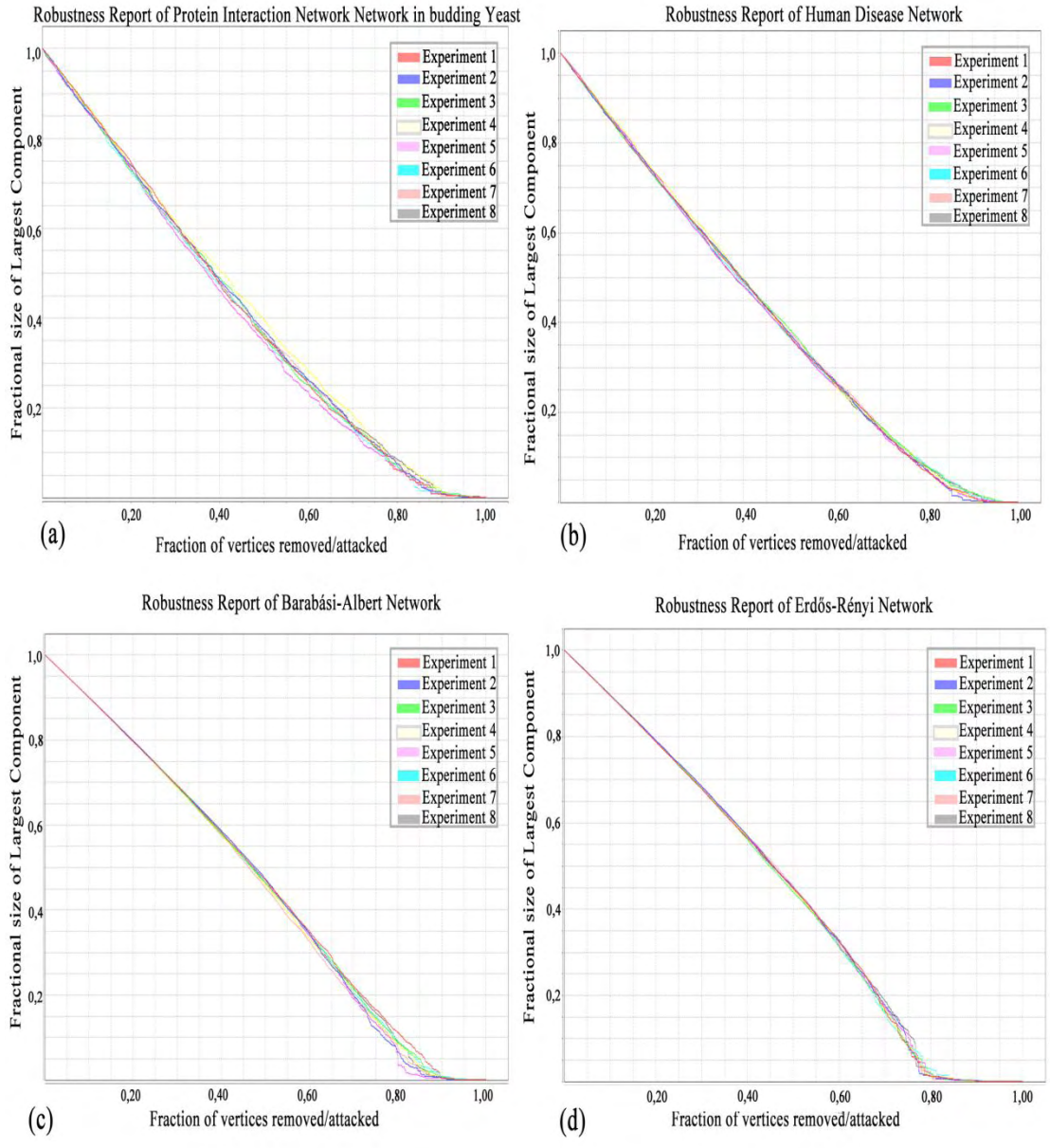


Figure 38| Robustness against simultaneous target attack of (a) the Protein Interaction Network in budding Yeast (b) the Human Disease Network, (c) Barabási–Albert Network and (d) the Erdős-Rényi Network

One can observe that the largest component reduces its size gradually and uniformly in all four network models and in each experiment robustness has a very similar behavior. In the Yeast Protein Interaction Network the size of the largest component shrinks to half when slightly more than 40% of the nodes are removed randomly. Approximately the same results appear in the Human Disease Network. To understand the results for the Yeast Protein Interaction Network we should have in mind that Protein Interaction Networks have an inhomogeneous structure and a great tolerance to random failures of their parts, as mentioned in [15]. Therefore a random deletion of nodes, does not highly affect the overall network structure.

In the case of BA and ER once the fraction of vertices removed exceeds about 45%-50% the size of the largest component is halved. The explanation for this is that when performing the node deletion, high centralized nodes are distributed throughout the network and the possibility to delete only these nodes is not very high. Instead, what is most likely to occur is a very random removal and therefore not only the central nodes are deleted but also those that are more "isolated". Consequently, the network is relatively robust against random failures of its components.

Reviewing the results, we believe that the disadvantage of performing random attacks on biological complex networks is that the vulnerability of the network under consideration cannot be exposed in any case since the results of each attack are broadly consistent with every other random attack performed on the same network. It seems that there is no clear case in which fragility of the network under consideration is higher than others.

Robustness Report

Random Attack

Network	First Attack	Second Attack	Third Attack	Fourth Attack	Fifth Attack	Sixth Attack	Seventh Attack	Eighth Attack
Yeast	0.381	0.382	0.380	0.393	0.370	0.375	0.381	0.395
HDN	0.394	0.394	0.397	0.398	0.392	0.394	0.398	0.394
BA	0.463	0.454	0.458	0.454	0.451	0.458	0.450	0.457
ER	0.435	0.437	0.433	0.436	0.436	0.435	0.436	0.440
max-min	0.082	0.072	0.078	0.061	0.081	0.083	0.069	0.063

Table 7| Robustness against simultaneous target attack by degree, betweenness, eccentricity, eigenvector, PCI, 2-PCI, closeness and k-shell of all networks. All values are rounded to the third digit.

Table 7 summarizes the results from the robustness measured during several random attacks on each network. The values in the last row show that each round of random attacks performed has very similar results for all networks; hence the difference from the maximum and the minimum values found is very small. Moreover, the lack of fragility of the networks against random node removal is also confirmed from the values of Table 8. Vulnerability appears to be relatively small in all network models and in all attacks performed on them.

Vulnerability Report								
Random Attack								
Network	**First Attack**	**Second Attack**	**Third Attack**	**Fourth Attack**	**Fifth Attack**	**Sixth Attack**	**Seventh Attack**	**Eighth Attack**
Yeast	0.119	0.118	0.120	0.107	0.130	0.125	0.119	0.105
HDN	0.106	0.106	0.103	0.102	0.108	0.106	0.102	0.106
BA	0.037	0.046	0.042	0.046	0.049	0.042	0.050	0.043
ER	0.065	0.063	0.067	0.064	0.064	0.065	0.064	0.060
max-min	0.082	0.072	0.078	0.061	0.081	0.083	0.069	0.063

Table 8| Vulnerability against simultaneous target attack by degree, betweenness, eccentricity, eigenvector, PCI, 2-PCI, closeness and k-shell of all networks. All values are rounded to the third digit.

Figure 39 shows the maximum and minimum values of robustness we encountered during the experiments and Figure 40 shows the number of nodes and the k-shells found in each network.

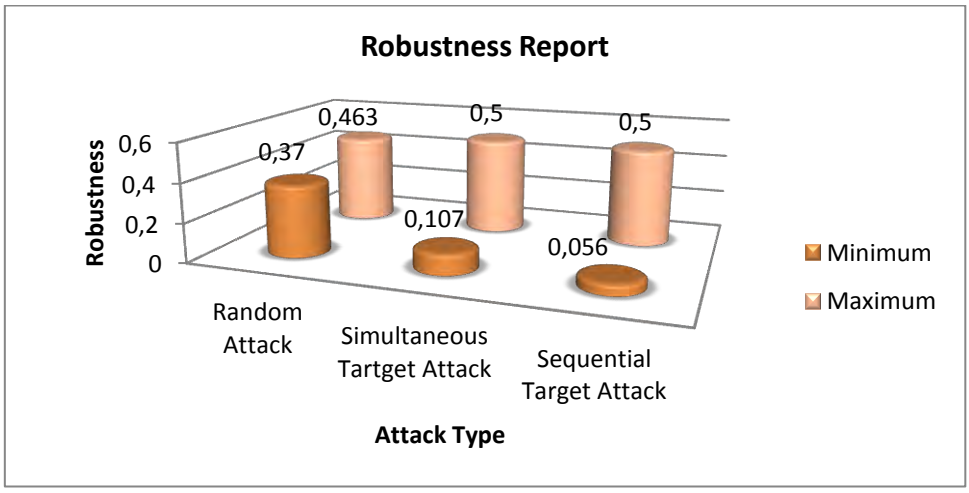


Figure 39| All networks along with their characteristics

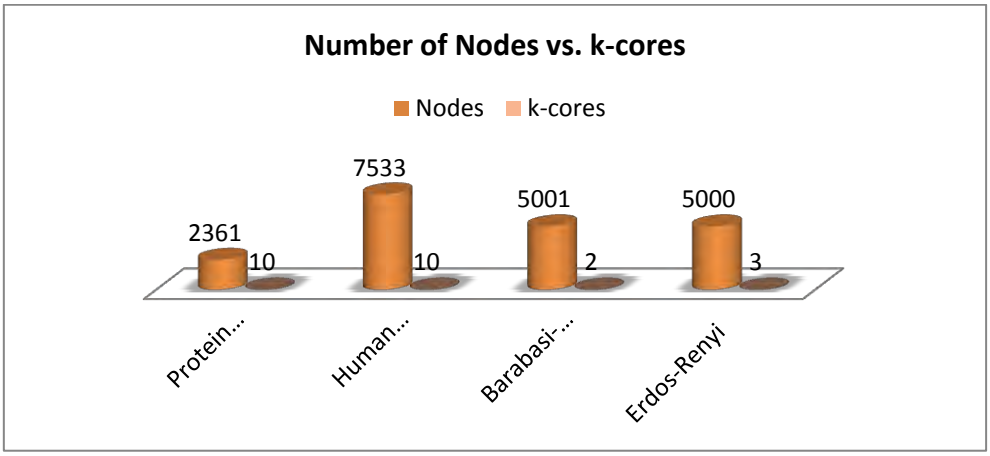


Figure 40| All networks along with their characteristics

5 Discussion

Many complex systems can be represented by complex networks. Complex network representation has found a number of applications in areas as bioinformatics, graph visualization, Internet mapping projects, sociology and distributed system analysis. Scientists originating from each discipline focus their interests on defining the robustness of the complex network since this will highlight the ability of a network to maintain its total throughput under node and link removal. Specifically, a robust network maintains to keep its components strongly connected and makes the network degradation very hard. The less robust a network is the more vulnerable it is against failures of its components. We consider three types of network attacks: the random attack, where nodes are removed randomly; sequential target attacks, where nodes are deleted according to the descending order of their centrality measure which is recalculated in each round; the simultaneous target attack, where nodes are deleted according to the descending order of their centrality measure which is calculated only initially.

This work endeavors to determine the robustness and vulnerability of biological complex networks towards targeted and random failures, using, besides from the centralities that have already been examined in previous work, the μ-Power Community Index to find out the most important nodes in the network. The obtained results described extensively in the above section lead us to the following summarized conclusions:

An interesting result is that PCI and 2-PCI metrics have in each attack type almost the same impact on the network. Particularly, when trying to attack a biological network by descending or random order of their Power Community Index, of class one and two for sequential target attacks and of class two for simultaneous target attacks, the robustness and vulnerability that is retrieved in each case approximates better the average values of robustness and vulnerability than the other centrality metrics. We consider as average value of robustness the average value of robustness of all other metrics except robustness value μ-PCI. We believe that 2-PCI can be considered as an approximation tool of average robustness and vulnerability for each of the two targeted attack types.

Another significant outcome is that regardless of the number of nodes the network contains and regardless of the number of k-cores it consists removing the nodes simultaneously targeted to the k-shell they belong we do not achieve any maximum or minimum value of robustness compared to the other measures. Deleting the most “central” nodes in the latter case, captures better the robustness in case of empirical biological networks and exposes better their vulnerability in case of synthetic ones.

For simultaneous target attacks, deleting nodes by degree is a superior method to expose the fragility of the network under node attack, since biological complex networks contain many hub nodes which are crucial for their functionality and play a vital role in the network structure. If the goal is to highlight the robustness then closeness centrality should be preferred to arrange the nodes in order of their importance.

For sequential target attacks, robustness of a biological complex network can be better retrieved when using closeness or eccentricity centrality to measure the importance of each node. If we want to exhibit how fragile a network under node attack is we may simply use betweenness centrality; one will only need to remove a fraction of nodes less than 30% to achieve the dissolution of the network structure.

A random attack, as our results confirm, does not show any particular trend to a certain metric value in order to determine either the robustness or the vulnerability of the network. All centralities tend to have the same behavior against random failures and the networks appear to be relatively tolerant to random node removal.

As an extension to our study, 2-PCI value can be examined as a approximation tool for robustness of biological complex networks; specifically, we have to define how precisely it approaches the average robustness compared to other centrality measures and how much deviation from the real robustness value can be expected.

Contributions

I would like to thank the Department of Electrical and Computer Engineering of University of Thessaly for providing me the Daemons for MapReduce tasks and their technical support stuff for the overall help. Special thanks to CERTH (CEnter for Research & Technology Hellas) of Volos for giving me the chance to work there and develop my k-core decomposition (MapReduce) algorithm. Finally, I would like to thank the anonymous reviewers for their suggestions and comments.

References

[1] He X, Zhang J, "Why Do Hubs Tend to Be Essential in Protein Networks?", **PLoS Genet** , vol.2, no. 6, 2006, Available at: http://www.plosgenetics.org/article/info:doi/10.1371/journal.pgen.0020088

[2] A. Sydney, C. Scoglio, M. Youssef, P. Schumm ," Characterizing the Robustness of Complex Networks", **Networking and Internet Architecture**, vol.3, 2009. Available at: http://arxiv.org/pdf/0811.3272.pdf

[3] S. Zhou, R. J. Mondragon "Redundancy and Robustness of the AS-level Internet topology and its models", IEE Electronic Letters, vol. 40, no. 2, pp. 151-15, 2004. Available at: http://arxiv.org/abs/cs/0402026

[4] Bothner, Matthew and White, Harrison C. and Smith, Edward (Ned), "A Model of Robust Positions in Social Networks", **American Journal of Sociology**, 2010. Available at SSRN: http://ssrn.com/abstract=1646446

[5] S Iyer, T Killingback , B Sundaram , Z Wang, "Attack Robustness and Centrality of Complex Networks", **PLoS ONE**, vol.8, no.4, 2013. Available at: http://www.plosone.org/article/info%3Adoi%2F10.1371%2Fjournal.pone.0059613

[6] N. Dimokas, D. Katsaros, L. Tassiulas, Y. Manolopoulos, "High performance, low complexity cooperative caching for wireless sensor networks", **ACM Wireless Networks**, vol. 17, no. 3, pp. 717–737, 2011. Available at: http://delab.csd.auth.gr/papers/WINET11dktm.pdf

[7] V. Nicosia, R. Criado, M. Romance, G. Russo, V. Latora, "Controlling centrality in complex networks", Scientific Reports, vol.2, 2012, Available at: http://dx.doi.org/10.1038/srep00218

[8] J.I. Alvarez-Hamelin, L. Dall' Asta, A. Barrat, A. Vespignani, "K-core decomposition of Internet graphs: Hierarchies, self-similarity and measurement biases", **Networks and Heterogeneous Media**, vol. 3, no. 2, pp. 371–393, 2008. Available at: http://www.cpt.univ-mrs.fr/~barrat/NHM.pdf

[9] A. Montresor, F. De Pellegrini, D. Miorandi, "Distributed k-core decomposition", **IEEE Transactions on Parallel and Distributed Systems**, vol. 24, no. 2, pp. 288–300, 2013. Available at: http://www.congas-project.eu/sites/www.congas-project.eu/files/publications/(2012)%20dKcore_TransParall_0.pdf

[10] S. Carmi, S. Havlin, S. Kirkpatrick, Y. Shavitt, E. Shir, "A model of Internet topology using k-shell decomposition", **Proceedings of the National Academy of Sciences (PNAS)**, vol. 104, no. 27, pp. 11150-11154, 2007. Available at: http://www.eng.tau.ac.il/~shavitt/pub/PNAS07.pdf

[11] A. Gursoy, O. Keskin and R.Nussinov, "Topological properties of protein interaction networks from a structural perspective", **Biochemical Society Transactions**, vol. 36, no. 6, 2008. Available at: http://www.ncbi.nlm.nih.gov/pubmed/19021563

[12] H. Jeong, B. Tombor, R. Albert, Z. N. Oltvai & A.-L. Barabási, **Nature**, vol. 407, pp. 651-654, 2000, Available at: http://www.nature.com/nature/journal/v407/n6804/full/407651a0.html

[13] D. Koschutzki, F. Schreiber, "Comparison of Centralities for Biological Networks", **Proc German Conf Bioinformatics (GCB'04)**, vol.53, 2004. Available at: http://citeseerx.ist.psu.edu/viewdoc/summary?doi=10.1.1.2.3286

[14] S. Wuchty, P. F. Stadler, "Centers of complex networks", **Journal of Theoretical Biology**, vol. 223, no.1, pp. 45-53, 2003. Available at: http://www.sciencedirect.com/science/article/pii/S0022519303000717

[15] H. Jeong, S. P. Mason, A.-L. Barabási & Z. N.Oltvai, "Lethality and centrality in protein networks",**Nature**, vol. 411, pp. 41-42, 2001. Available at: http://www.nature.com/nature/journal/v411/n6833/abs/411041a0.html

[16] K.-I. Goh , M. E. Cusick, D. Valle, B. Childs, M. Vidal, A.-L. Barabási, "The human disease network", Proceedings of the National Academy of Sciences (PNAS), vol.104, no. 21, pp.8685-8690, 2007. Available at: http://www.pnas.org/content/104/21/8685.full

[17] L. da F. Costa, F. A. Rodrigues and A. S. Cristino, "Complex networks: The key to systems biology", **Genetics and Molecular Biology**, vol.31, no. 3, pp. 591-601, 2008. Available at: http://www.scielo.br/pdf/gmb/v31n3/a01v31n3

[18] http://en.wikipedia.org/wiki/Biological_network

[19] http://en.wikipedia.org/wiki/Complex_network

[20] http://cophy-wiki.informatik.uni-koeln.de/index.php/Graph

[21] http://en.wikipedia.org/wiki/Barab%C3%A1si%E2%80%93Albert_model

[22] http://en.wikipedia.org/wiki/Erd%C5%91s%E2%80%93R%C3%A9nyi_model

[23] http://en.wikipedia.org/wiki/Betweenness_centrality

[24] http://en.wikipedia.org/wiki/Centrality

[25] http://en.wikipedia.org/wiki/Eccentricity_(graph_theory)

[26] http://socnetv.sourceforge.net/docs/analysis.html

[27] http://vlado.fmf.uni-lj.si/pub/networks/data/bio/Yeast/Yeast.htm

[28] http://www.barabasilab.com/pubs/CCNR-ALB_Publications/200705-14_PNAS-HumanDisease/Suppl/index.htm

[29] http://djjr-courses.wikidot.com/soc180:eigenvector-centrality

Figure References

Figure 1| http://bioinfow.dep.usal.es/coexpression/network.jpg

Figure 2| http://fiehnlab.ucdavis.edu/staff/grapov/grapov-metabolic-network-jpg.png

Figure 3| http://optimizationandanalytics.files.wordpress.com/2013/01/complex-network-structure.png

Figure 4| http://www.nature.com/srep/2013/130827/srep02517/fig_tab/srep02517_F5.html

Figure 5| http://homepages.ius.edu/rwisman/C455/html/notes/AppendixB4/B4-2.gif

Figure 6| http://www.nature.com/scitable/topicpage/genome-wide-association-studies-and-human-disease-788

Figure 10| http://assets.20bits.com/misc/low-degree.png

Figure 11|http://upload.wikimedia.org/wikipedia/commons/6/60/Graph_betweenness.svg