

PAPER • OPEN ACCESS

## Characterization of Essential Protein in Aging Based on Network Analysis

To cite this article: H Syukor and A B Sakhinah 2019 *J. Phys.: Conf. Ser.* **1366** 012026

View the [article online](#) for updates and enhancements.



**IOP | ebooks™**

Bringing you innovative digital publishing with leading voices to create your essential collection of books in STEM research.

Start exploring the [collection](#) - download the first chapter of every title for free.

# Characterization of Essential Protein in Aging Based on Network Analysis

H Syukor<sup>1</sup> and A B Sakhinah<sup>1</sup>

<sup>1</sup>School of Mathematical Science, Faculty of Science and Technology, National University of Malaysia, 43600, Bangi, Selangor Darul Ehsan.

**Abstract.** The aging process is a process experienced by all living organisms. For human beings, this process occurs gradually starting at a young age. It is believed that proteins in some parts of the human brain play an important role in the aging process. Therefore, the protein interaction network for the aging process is aimed at obtaining characterization of essential proteins. Samples of proteins are obtained from *ArrayExpress*. Subsequently, data from Data of Interacting Protein (DIP) is used to obtain the connected proteins to generate network interactions for the aging process. A total of 38 networks are produced by age and gender, however we only focus on 24 networks that belong to 2 age classes. Using information from *GeneAge*, essential proteins for the aging process are extracted from all generated networks. Thereafter, five measurements were used to characterize all the essential proteins which are Degrees, Closeness Centrality, Betweenness Centrality, Local Clustering Coefficient and Number of Triangles. The results showed four out of five measurements were able to provide the main characteristics of essential proteins in which Degrees, Betweenness Centrality, Local Clustering Coefficients and Number of Triangles. All the characteristics of this protein are compared to age group and gender. Each age group and gender assigned different major protein markers. Therefore, a unique protein, which is the essential protein that “stands out” from other essential proteins based on every measurement in each class age and gender are identified. This analysis shows that there are several unique essential proteins to men and women that each have a distinctive characteristic in the protein interaction network for the aging process.

## 1. Introduction

Aging is a process experienced by every living organism. It can be defined as a slow process of losing the ability to maintain the normal functions possessed in humans [1]. Typically, the aging process is divided into two, namely physical change and mental change. Examples of physical changes commonly encountered by the elderly are wrinkles on the face and the failure to stand upright due to missing some bones as age increasing [2] while examples of mental changes are the loss of ability to remember something well [3].

In the aging process, it is believed that the protein has a bearing in effect throughout this process. In a study conducted by Irizarry [4],  $\beta$ -amyloid protein is one of the factors causing dementia, which is a memory loss that is considered as one of the aging process. In addition, the study conducted by Kato and his colleagues [5] showed that there was a change in the number of S-100 $\beta$  protein when the aging process occurred, that is found to affect the cerebral cortex where it is a large part of the brain that functions to process information from inputs received before producing output. Therefore, to



understand the aging process, studies on protein interactions in the human brain are aimed at providing information on protein characterization throughout the aging process.

The main purpose of this study is to obtain a protein interaction network that involves in the aging process in humans. Based on the resulting protein-protein interaction (PPI) network, this study has three objectives as follows:

- Identify the characterization of essential proteins in the aging process using network analysis.
- Compare essential protein features between age classes and gender.
- Identify the essential protein that is unique in the series based on the characterization of the essential proteins.

## 2. Problem Statement

Aging process can be seen in most of the organisms physically and the visible effects can be seen in humans [6]. In the human brain, aging can be detected in one's mental changes [7]. Therefore, the study of gene expression in the proper aging process is aimed to find out the causes and effects of the aging process. According to Faisal and Milenković [8], studies on gene expression have certain limitations and interaction of gene interactions are not considered. The alternative to this deficiency is to study the protein interactions involved in the aging process [8].

This study encompasses protein-protein interactions (PPI) network for the aging process. Using a protein sample obtained from several parts of the brain, the protein interaction network is produced by age and gender. Subsequently, the essential proteins obtained and characterization for these essential proteins are measured by several indicator such as degree, Betweenness Centrality (BC), Closeness Centrality (CC) and Local Clustering Coefficient (LCC) and number of triangles. The study continued with the identification of the unique essential protein which give a different reading for every measure tested on these essential proteins.

## 3. Materials and Methods

### 3.1 Data Preparation

In this study, the data were obtained from three different sources. The first is gene expression data [9]. This data contains gene expression selected from four parts of the brain: Hippocampus (HC), Superior Frontal Gyrus (SG), Postcentral Gyrus (PCG) and Entorhinal Cortex (EC). Sample of this gene expression contains 172 probes each representing the expression of the gene from the protein taken from the part of the above-mentioned brain. Data from *ArrayExpress* consists of 63 individuals aged between 20 - 99 years old as a sample [9]. For each probe, 38909 proteins with different IDs were tested to see the presence of the protein in each of the existing probes. The ID for this protein is referred from UniprotID, which provides a list of proteins with different features and functions [10]. These protein IDs are unique to differentiate one from the others. These 172 probes are then divided according to gender and ages. Then, only the active gene expression will be selected by using majority vote rule as proposed by Lee and his teammates [11].

The next data is protein interaction data for *homo sapiens*. The source of this data is from the Database of Interacting Protein (DIP) [12]. The content of this data is a pair of two proteins that possibly connecting each other. This data is important to obtain the interacting protein before building the PPI network. Those active gene expression from different ages and gender are referred to DIP to obtain static PPI network before further network analysis run over every network.

Finally, data for essential protein for aging process is taken from *GeneAge* [13]. It is a website that provides gene database from multiple sources for various uses and research. This data is important to differentiate which protein in the network is an essential protein. Therefore, the characterization of these essential proteins will be determined afterwards.

After obtaining essential protein from the entire network, the analysis is made differently according to gender and age class. Petry [14] states that age class can be divided into three groups, which are young (18-35 years), middle-aged (36-55 years) and elderly (>55 years). An addition has been made by dividing elderly class into two, which are elder (56-72 years) and senior citizen (>72 years). This separation is done to see whether there is a different characterization of essential protein among elderly before average human lifespan at 72 years [15] with those after average human lifespan.

### 3.2 Network Analysis

Basically, network can be measured globally and locally. This PPI network for aging process did not have a significance difference between ages globally [8]. Therefore, for global analysis, these networks are compared among gender and the measurement used is average density, given by:

$$D = \frac{2E}{N(N-1)} \quad (1)$$

where  $E$  is the number of edges and  $N$  is the number of nodes in the entire network [16].

For local topology, five different measurements used, which are as stated below:

1. Degree (D). Degree of a node is one of the easiest measures and determined by calculating the number of edges connecting one node with the other.
2. Closeness Centrality (CC). The closeness of a node referred as how close a node with other neighbouring nodes. This can be calculated using the formula as follows:

$$CC(v) = \frac{N}{\sum_{i \in V} d(v, v_i)} \quad (2)$$

where  $N$  is the number of nodes connecting to node  $v$  and  $d(v, v_i)$  describes as shortest path connecting node  $v$  to any node  $v_i$ .

3. Betweenness Centrality (BC). This centrality indicates how “often” a node is located in between of a shortest path of two nodes  $s$  and  $t$ . BC has multiple calculations and for this study, the formula proposed by Brandes [17] is used, which is:

$$\sum_{w: v \in P_s(w)} \left( \frac{\sigma_{sv}}{\sigma_{sw}} + \sum_{t \in V/\{w\}} \frac{\sigma_{sv}}{\sigma_{sw}} \cdot \frac{\sigma_{st}(w)}{\sigma_{st}} \right) \quad (3)$$

where  $P_s(w) = \{u \in V: \{u, v\} \in E; d_G(s, u) = d_G(s, v) + \omega(u, v)\}$  [17].

4. Local Clustering Coefficient (LCC). It has several definitions which one considered as a ratio of total number of pairs of neighbors of a node  $v$  that are connected with the total number of pairs of neighbors of  $v$  [18]. Besides, it is also defined as the probability that two nodes with common neighbors are connected [19]. An LCC of a node  $v$  can be computed using this formula:

$$LCC(v) = \frac{2E(v)}{k_v(k_v - 1)} \quad (4)$$

where  $E(v)$  is the number of edges between neighbors of  $v$  and  $k_v$  indicates degrees of a node  $v$ .

5. Number of triangles, which indicates the total number of triangles formed in a cluster of a node  $v$ .

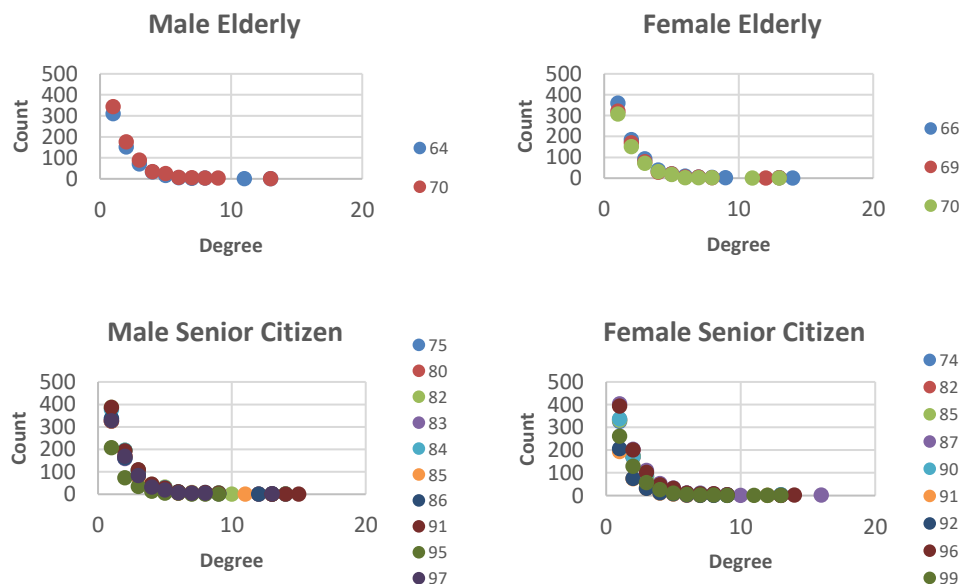
#### 4. Results and Discussion

The first part of the result is focusing on the network globally. The average network density of male and female in two classes of elderly stated as follows:

**Table 1:** Average network density according to gender and age classes

Age class	Male (%)	Female (%)
Elderly	0.3036	0.3082
Senior citizen	0.3092	0.3292

From the table, it is clearly stated that PPI network for female has a greater density compared to male, which means the interaction between proteins in PPI network for aging process in female is greater compared to male. Even though the density is different across classes, with close insight to the degree distribution we can see that all the proteins in the networks follows power law distribution. This means that as the degree of a protein increases, the number of proteins with such value are decreasing. Thus, we can say that only selective proteins have a high value of degree.



**Fig 1:** Graph of degree distribution of essential proteins in male and female elderly and senior citizen

As we can see, most proteins have a degree of less than 10 and below. By looking at the results obtained, we found out that in all networks, the proteins that have a high degree are the same proteins in which their degree are more than 10. Therefore, those proteins can be considered as “unique” proteins due to the high connection with the proteins in the neighborhood. Those “unique” proteins are P19387, P35222 and Q9Y4K3.

For the second part, the characterization of essential protein observed in local topological PPI network. Thus, three analyses have been done upon all the proteins involved in the aging process, which are the relationships between:

- Betweenness Centrality (BC) and Closeness Centrality (CC)
- Local Clustering Coefficient (LCC) and Degree (D)
- Number of Triangle and Local Clustering Coefficient (LCC)

The results of all three analyses will be described in each section.

1. Relationship between BC and CC.

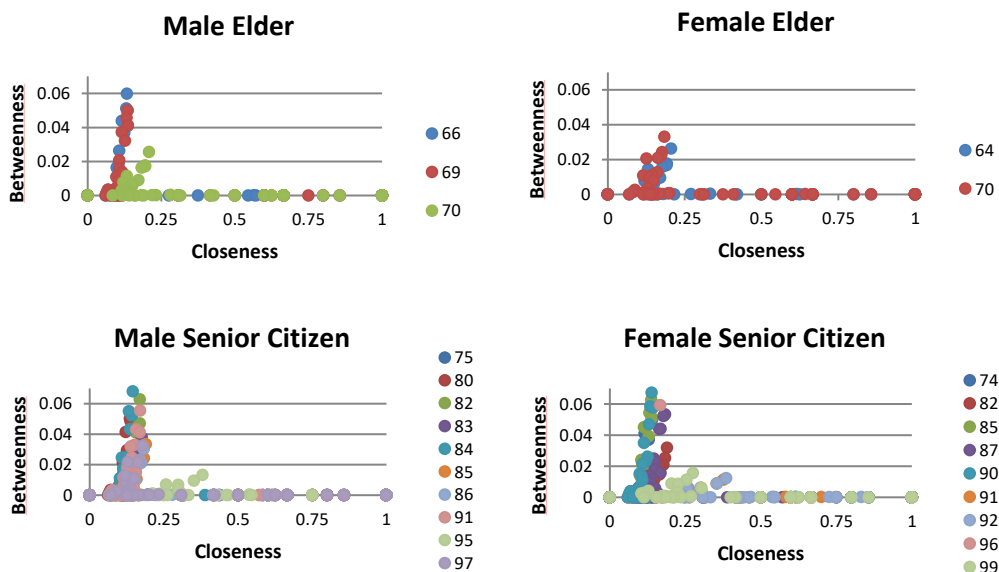


Fig 2: Graph of BC versus CC for essential protein in male and female elderly and senior citizen

From the result, all the graphs give the same pattern as we can see that most of the proteins are located within 0 to 0.3 for CC value and these proteins have a various range of BC. We can also see that the correlation of the centralities is negatively weak (near to 0). By applying the interpretation of social network analysis made by [20] into this PPI network, we can say about the value of centralities of each protein can be described as follows:

Table 2: Characteristics of proteins behavior that related to CC and BC

	Low Betweenness Centrality	High Betweenness Centrality
Low Closeness Centrality	In the network, these proteins either did not interact with other protein or located at the tail of the paths.	The protein monopolized the connection from several proteins to many other proteins.
High Closeness Centrality	Network exists in many paths. The proteins are near to other proteins and so are the others.	This kind of protein did not exist in our PPI network for aging. If it does, it means that the protein possessed with a lot of connection and important in the entire network.

We consider low value of CC is below 0.5 and low value of BC is 0.02. To determine the “uniqueness” of a protein, the distribution of the reading for both centralities should be taken into account. For closeness centrality, almost 70% of the protein has low closeness value, suggesting a closeness centrality is not a good measure for measuring the “uniqueness” of a protein because protein should be close to each other to show that the protein is important in the PPI network. Although there are some proteins give a high reading of CC, however the protein has a value of 0 for betweenness centrality indicating that the main protein is not an intermediary to two other proteins in a connecting path. For betweenness centrality, there are some proteins that give a high readability even though the reading for its closeness centrality is low. Therefore, there are unique

characteristics of these proteins and therefore BC is a good measure in determining the “uniqueness” of proteins in the PPI network for aging process.

As a result, we found out that the proteins with a high value of betweenness centrality is the protein with ID P20226, P25963, Q01094, Q14186 and P41229. These proteins may have different values of BC in each network according to age and gender, yet these are among the highest.

Apart from centralities, other measurement that is usually used in PPI network analysis is the strength of the clusters in the network. With clustering, we can tell the topology of the network has a strong cluster or not. The objective of clustering is usually to determine if the network can be divided into sub networks (clusters). From the clusters, we can differentiate the function of the respective clusters because biological network tends to split into different clusters according to its functionality. Thus, for the next two analyses, we focused on the clustering of a node which includes the reading of Local Clustering Coefficient, Degree and number of triangles.

2. Relationship between LCC and D for male and female of elders.

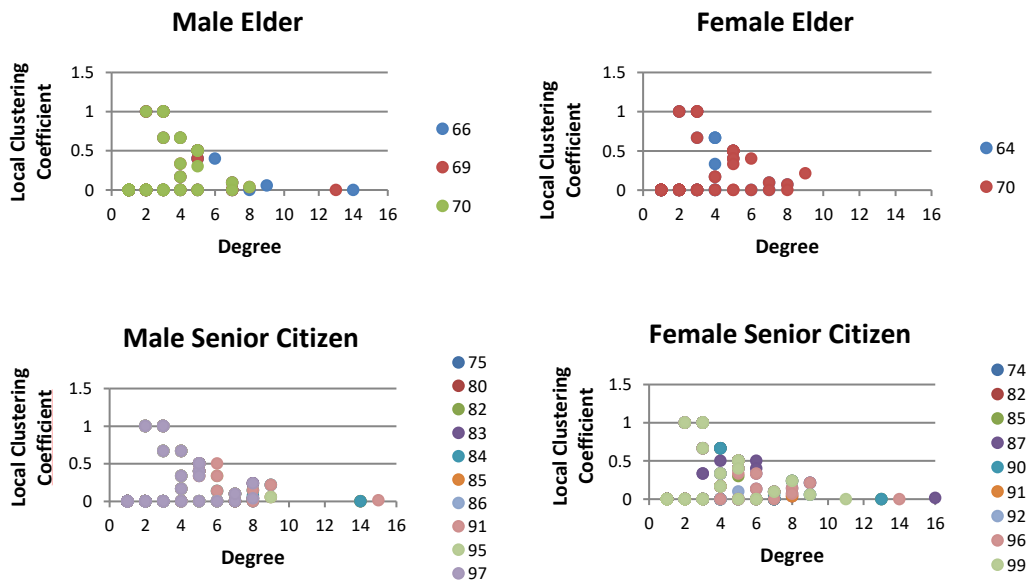
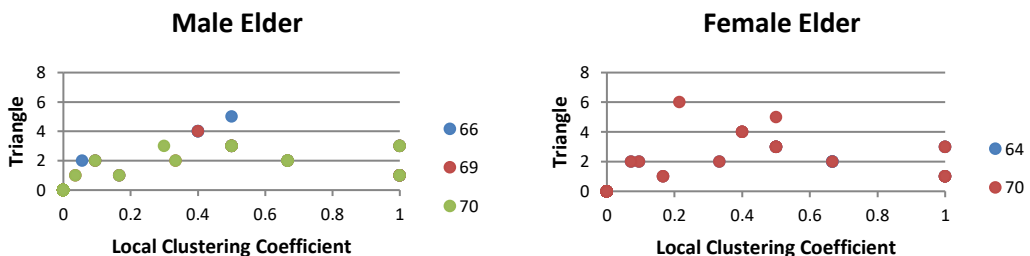
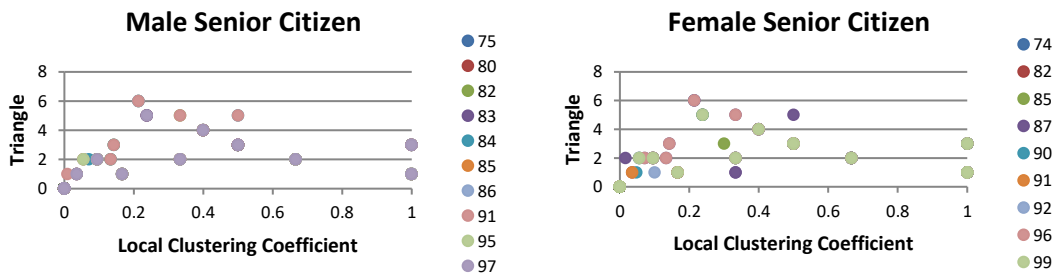


Fig. 3: Graph of LCC versus D for essential proteins in male and female senior citizen.

3. Relationship between number of triangle and LCC for male and female of elders.





**Fig. 4:** Graph of triangles versus LCC for essential proteins in male and female senior citizen

For the resulting clusters, graphs for each age group and gender are almost identical. Based on the reading, we can identify certain characteristics of the networks:

- For protein having zero readings for both LCC and number of triangles, it means that those proteins did not have a neighbor that are connecting to each other.
- For protein having a reading of 1 for LCC, it means that the neighbors to that protein are connecting to each other, forming a triangle. From the graph, we can see that proteins with LLC =1 have degree either 2 or 3 and with the number of triangles of 1 and 3 respectively. The connection for these clusters is shown as below:



**Fig 5:** Proteins with LLC=1 (red) that have a degree of 2 and number of triangles of 1 (left) and degree of 3 with number of triangles of 3 (right)

Overall, the LCC for analysis in section 2 and 3 comprehend that the essential proteins are not really in a strong cluster, because most of the proteins that have a strong LCC value has a small degree. The cluster is considered strong when there are proteins that interact within have a high value of degree as well as LCC. This can be seen in the network (see Appendix 6.2) where the structure of this PPIN for aging are more to tree-like instead of cluster-like.

Even though the overall network has a weak clustering, we can still further the analysis by identifying the clusters locally that gives a constant reading throughout different networks according to age and gender. These proteins can also be classified as “unique” proteins. This is due to a statement made by Faisal and Milenkovic [8] stating that the essential proteins usually undergo a dynamic change in terms of interaction as age continues. So, by using LCC, we present the results of the study by showing that even if age increases, some essential proteins do not constantly change their interaction instead some of them has a constant interaction in clusters.

As a result, there are five proteins each for men and women who provide consistent reading for Local Cluster Coefficient and number of triangles, as shown in the tables below:

**Table 3:** List of “unique” proteins in clustering for both male and female

Male	Female
P01100	P00553
P00116	P01100
P08047	P00116
P09619	P08047
P12965	P12965
P25445	P78527
P60484	Q00163
P61244	
P78527	

## 5. Conclusions

This study encompasses the characterization of essential proteins using network analysis in which can be describe using centralities as well as clustering. For centrality, we used three different types of centralities which are Degrees, Closeness and Betweenness. Based on analysis, we can concur that CC did not give a meaningful information since most of the proteins have a low value of closeness with their neighborhood proteins. Thus, we determined the “unique” protein based on degrees and BC.

In terms of clustering, we decided to see the relationships between LCC, degree and number of triangles. Although some of the proteins are not remarkable in terms of centralities but maybe several of them give meaningful information in clustering. Thus, by looking at the clusters of every single network by age and gender, we managed to pinpoint exactly 5 essential proteins for both male and female that gives a constant reading. Hereby we conclude all 9 proteins for male and female respectively that we contemplated as “unique” proteins.

**Table 4:** List of “unique” essential proteins female from the measurement of characterization

Measurement	ID protein	
	Male	Female
Degree	P19387	P19387
	P35222	P35222
	Q9Y4K3	Q9Y4K3
Betweenness	P20226	P20226
	P25963	P25963
	P41229	P41229
	Q01094	Q01094
	Q14186	Q14186
Clustering	P01100	P00553
	P00116	P01100
	P08047	P00116
	P09619	P08047
	P12965	P12965
	P25445	P78527
	P60484	Q00163
	P61244	
	P78527	

From the table, the “unique” proteins are similar for both male and female except one protein from each gender which is P60484 for male and P12004 for female. Therefore, in aging process, these two proteins are believed to be functioning better at one gender compared to the other, in term of clustering.

The study of aging computationally benefits medical fields where by using network analysis it can help drug targeting especially multi-targeting that has proven gives a better result [20]. Besides, by understanding biological network especially in aging can also helps in drug design [21] because the different type of networks can give different signalling and pathways of the protein interaction therefore different techniques required in designing and targeting the drug to the specific proteins [22].

## 6. Appendix

### 6.1 Dataset sample

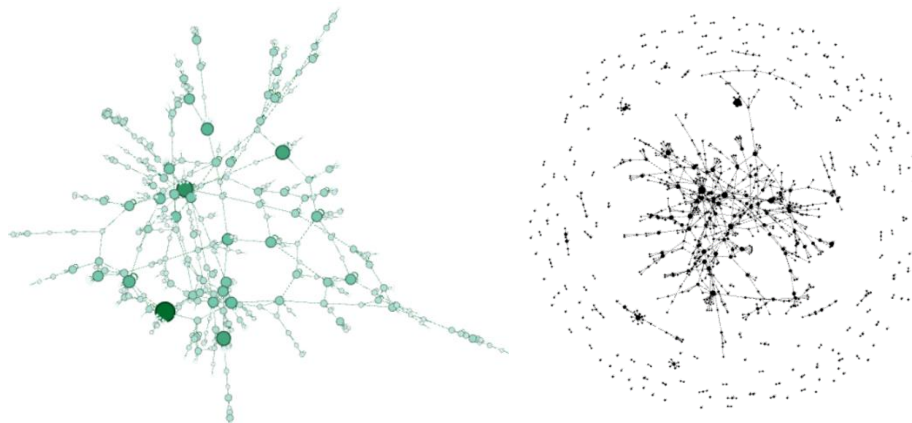
Below is the sample of pair of interacting proteins for male age 69. The ID of proteins are referred to Uniprot [10].

P20226	P09086		P09661	Q12874		P35221	P35222		P45983	P05412
P31946	P04049		P35568	P06213		P08238	P08238		P45983	P01112
P25490	Q09472		P29353	P06213		P19388	Q00403		Q9Y618	Q05516
P15884	Q02535		P01112	P04049		P01100	P20226		P19525	P19525
P02593	P35609		P02593	P17677		P09619	P42229		P02787	P02786
Q15628	P19438		P19438	P19438		P13861	P13861		P15692	P17948
Q16594	P49848		Q92793	P16220		P13861	P11137		P15692	P15692
Q00403	Q16594		P46939	Q13424		P41236	P36873		P04141	Q99062
Q15572	P20226		P11831	P35269		P05412	P28482		P18075	Q13873
P08047	Q00403		P07948	P22681		P45984	P05412		Q13873	Q13873
Q00403	P25490		P11802	Q96TE0		P05412	P53779		P36894	Q13873
P08047	P25490		Q00534	Q96TE0		P05412	P18848		P49238	P78423
P20226	Q15544		P11802	P30279		P18848	P01100		P36894	P36894
P06127	Q8NEB9		P30279	Q96TE0		P04049	Q02750		Q12888	P04150
P01112	Q13671		P20226	P05412		P29353	P00533		Q14192	Q13683
Q13671	P42655		P11166	P27797		P20936	P09619		Q14192	P26006
P42684	Q13671		P31323	P31323		P19174	P09619		Q14739	Q13185
P42768	P16333		P01111	P04049		P27635	P49768		P45973	Q14739
P00533	Q13596		P32780	P19447		Q9UQB3	P49768		P10276	Q15596
P25445	P25445		P20226	Q13487		P29323	P29323		Q00534	P55273

**Fig 6:** A list of some pairs of interacting proteins

### 6.2 Network sample

All those pairs of interacting proteins will eventually used to construct a single network. This is a part of PPI network for aging process of male age 69. This network consisting 951 nodes (proteins) and 1103 edges (interaction between proteins). The network is visualized according to the degree. The higher the degree, the larger the node size.



**Fig 7:** Sample of network (left) and full-scale network (right).

## Reference

- [1] P Constantinides. (1994). *General Pathobiology*. McGraw-Hill/Appleton & Lange.
- [2] P D Delmas, D Stenner, H W Wahner, K G Mann, & B Riggs. 1983. Increase in Serum Bone Gamma-Carboxyglutamic Acid Protein with Aging in Women. Implications for the Mechanism of Age-Related Bone Loss. *Journal of Clinical Investigation* 71(5): 1316.
- [3] M Perlmutter, C Adams, J Berry, M Kaplan, D Person. & F Verdonik. 1987. Aging and Memory. *Annual review of gerontology & geriatrics* 7(57-92).
- [4] M Irizarry, M Gurol, S Raju, R Diaz-Arrastia, J Locascio, M Tennis, B Hyman, J Growdon, S Greenberg, & T Bottiglieri. 2005. Association of Homocysteine with Plasma Amyloid B Protein in Aging and Neurodegenerative Disease. *Neurology* 65(9): 1402-1408.
- [5] K Kato, F Suzuki, R Morishita, T Asano, & T Sato. 1990. Selective Increase in S-100 $\beta$  Protein by Aging in Rat Cerebral Cortex. *Journal of neurochemistry* 54(4): 1269-1274.
- [6] G E Rodwell, R Sonu, J M Zahn, J Lund, J Wilhelmy, L Wang, W Xiao, M Mindrinos, E Crane & E Segal, 2004. A Transcriptional Profile of Aging in the Human Kidney. *PLoS biology* 2(12): e427.
- [7] T Lu, Y Pan, S-Y Kao, C Li, I Kohane, J Chan & B A Yankner. 2004. Gene Regulation and DNA Damage in the Ageing Human Brain. *Nature* 429(6994): 883-891.
- [8] F E Faisal, & T Milenković. 2014. Dynamic Networks Reveal Key Players in Aging. *Bioinformatics* 30(12): 1721-1729.
- [9] N C Berchtold, D H Cribbs, P D Coleman, J Rogers, E Head, R Kim, T Beach, C Miller, J Troncoso, & J Q Trojanowski. 2008. Gene Expression Changes in the Course of Normal Brain Aging Are Sexually Dimorphic. *Proceedings of the National Academy of Sciences* 105(40): 15605-15610.
- [10] T U Consortium. 2008. The Universal Protein Resource (Uniprot). *Nucleic acids research* 36(suppl 1): D190-D195.
- [11] M-L T Lee, F C Kuo, G Whitmore & J Sklar. 2000. Importance of Replication in Microarray Gene Expression Studies: Statistical Methods and Evidence from Repetitive Cdna Hybridizations. *Proceedings of the National Academy of Sciences* 97(18): 9834-9839.
- [12] I Xenarios, D W Rice, L Salwinski, M K Baron, E M Marcotte, & D Eisenberg. 2000. Dip: The Database of Interacting Proteins. *Nucleic acids research* 28(1): 289-291.
- [13] R Tacutu, T Craig, A Budovsky, D Wuttke, G Lehmann, D Taranukha, J Costa, V E Fraifeld & J O P De Magalhães. 2012. Human Ageing Genomic Resources: Integrated Databases and Tools for the Biology and Genetics of Ageing. *Nucleic acids research* 41(D1): D1027-D1033.
- [14] N M Petry. 2002. A Comparison of Young, Middle-Aged, and Older Adult Treatment-Seeking Pathological Gamblers. *The Gerontologist* 42(1): 92-99.
- [15] W H Organization. 2016. Life Expectancy Increased by 5 Years since 2000, Both Health Inequalities Persist. *World Health Statistics*
- [16] J Scott. 2017. *Social Network Analysis*. Sage.
- [17] U Brandes. 2001. *Journal of Mathematical Sociology* 25(2): 163-177
- [18] M Newman. 2010. *Networks: An Introduction*. Oxford university press.
- [19] S A Bakar, J Taheri, & A Y Zomaya. 2014. Characterization of Essential Proteins Based on Network Topology in Proteins Interaction Networks. *AIP Conference Proceedings*, pg. 36-42.
- [20] Zhang, J., & Luo, Y. (2017, March). Degree Centrality, Betweenness Centrality, and Closeness Centrality in Social Network. In 2017 2nd International Conference on Modelling, Simulation and Applied Mathematics (MSAM2017). Atlantis Press.
- [21] Zimmermann, G. R., Lehar, J., & Keith, C. T. (2007). Multi-target therapeutics: when the whole is greater than the sum of the parts. *Drug discovery today*, 12(1-2), 34-42.
- [22] Csermely, P., Agoston, V., & Pongor, S. (2005). The efficiency of multi-target drugs: the network approach might help drug design. *Trends in pharmacological sciences*, 26(4), 178-182.
- [23] Korcsmáros, T., Szalay, M. S., Böde, C., Kovács, I. A., & Csermely, P. (2007). How to design multi-target drugs: target search options in cellular networks. *Expert opinion on drug discovery*, 2(6), 799-808.