



Molecular Network Analysis in Rabies Pathogenesis Using Cooperative Game Theory

Razieh Bostani, Mehdi Mirzaie*

Department of Applied Mathematics, Faculty of Mathematical Sciences, Tarbiat Modares University, Tehran, Iran

*Corresponding author: Mehdi Mirzaie, Department of Applied Mathematics, Faculty of Mathematical Sciences, Tarbiat Modares University, Tehran, Iran. P.O.Box: 14115-134; Tel: +98-21-82884267; Fax: +98-21-82884267; Email: mirzaie@modares.ac.ir

Background: Recently, many researchers from different fields of science have been used networks to analyze complex relational big data. The identification of which nodes are more important than the others, known as centrality analysis, is a key issue in biological network analysis. Although, several centralities have been introduced degree, closeness, and betweenness centralities are the most popular. These centralities are based on the individual position of each node and the cooperation and synergies between nodes have been ignored.

Objectives: Since in many cases, the network function is a consequence of cooperation and interaction between nodes, classical centralities were extended to a group of nodes instead of only individual nodes using cooperative game theory concepts. In this study, we analyze the protein interaction network inferred in rabies disease and rank gene products based on group centrality measurements to identify the novel gene candidates.

Materials and Methods: For this purpose, we used a game-theoretic approach at three scenarios, where the power of a coalition of genes assessed using different criteria including the neighbors of genes in the network, and predefined importance of the genes in its neighborhood. The Shapley value of such a game was considered as a new centrality. In this study, we analyze the network of gene products implicates rabies. The network has 1059 nodes and 8844 edges and centrality analysis was performed using CINNA package in R software.

Results: Based on three scenarios, we selected genes among the highest Shapley value that had low ranking from classical centralities. The enrichment analysis among the selected genes in scenario 1 indicates important pathways in rabies pathogenesis. Pair-wise correlation analysis reveals that changing the weights of nodes at different scenarios can significantly affect the results of ranking genes in the network.

Conclusions: A prior knowledge about the disease and the topology of the network, enable us to design an appropriate game and consequently infer some biological important nodes (genes) in the network. Obviously, a single centrality cannot capture all significant features embedded in the network.

Keywords: Centrality analysis; Cooperative game theory; Protein interaction network; Shapley value

1. Background

Network analyses have been widely used in different fields of science such as biology, physics, and social sciences. In many real networks, some specific nodes play more important roles than the other nodes (1). Centralities are functions which assign a real number to each node based on its special characteristic and accordingly rank network nodes. Different centralities have been introduced, meanwhile, degree centrality, closeness centrality, and betweenness centrality are the popular detailed measurements. Degree centrality (2) is a simple centrality that counts the number of edges for

each node. Closeness centrality (3) quantifies the mean distance from a node to all other nodes of the network. Betweenness centrality (4) measures the number of times a node lies on the shortest path between the other nodes.

Centrality analysis of biological networks including co-expression, signaling and protein-protein interaction networks were successfully used to identify important genes or proteins in the very complex system of pairwise relations in the network (5). Application of centrality measures for the analysis of biological networks can be found in the papers (6-10). In classical centralities,

the cooperation and synergies between nodes are not considered while in some cases the function of the network is highly dependent on the cooperation and interaction between groups of nodes (11). Everett and Borgatti introduced the concept of group centrality as an extension of classical centrality measures that consider a group of nodes instead of only individual nodes (12). As an example, consider the sample network of 9 nodes in **Figure 1A**.

Let nodes represent important places in a city and edges are communications routes between them. How can we determine two police stations, so that they can cover most of the network? Choosing 1 and 5 with highest degrees is an option that covers 6 nodes, while 5 and 7 support the whole network. This example shows that, although 7 has lower degree centrality than 1, joint degree centrality of {5, 7} is a more appropriate approach for the important node selection.

Recently, Cesari et al. applied game theory concept and the Shapley value to assess the relevance of each gene in interaction with the others in the co-expression network (13, 14). We used similar ideas to analyze the protein-protein interaction network of gene products implicate rabies constructed by Jamalkandi et al (15).

2. Objective

In this paper, we briefly review game-theoretic network centrality and apply three scenarios on the protein-

protein interaction network inferred in rabies disease. Each scenario tries to capture some topological features of the network and presents some new gene products that cannot be identified by other methods.

3. Materials and Methods

In game theory, cooperative game studies the behavior of players that can form coalitions to attain higher or lower scores than the sum of each of those players working individually. In fact, cooperation game theory addresses the synergy among groups of players. More formally, a cooperative game consists of a set of players $N = \{1, 2, \dots, n\}$ and a characteristic function $v: 2^N \rightarrow \mathbb{R}$ that assigns a real value to each coalition $S \subseteq N$ representing its value (pay off) (16). It assumes that $v(\emptyset) = 0$ and $v(S \cup T) \geq v(S) + v(T)$ for any coalition S and T such that $S \cap T = \emptyset$. A cooperative game denoted by $G = (N, v)$.

The big challenge is how to fairly allocate and divide the total payoff of $v(N)$ among the players. A solution concept is a vector $\theta(v) = (\theta_1(v), \dots, \theta_n(v))$ that represents the allocation to each player. In fact $\theta_i(v)$ is the value or power of i-th player in the game. Shapley value is one of the most popular solution concepts in cooperative game theory that assigns a single value for each player, which is the average of all marginal contributions to all possible coalitions and comply with the following principles which are known as Shapley principles (17):

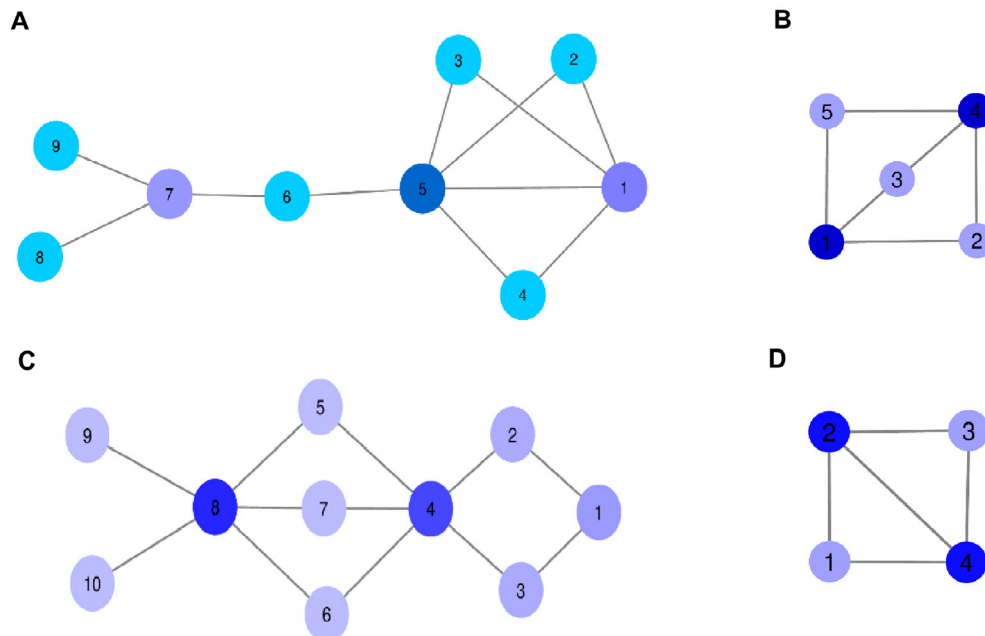


Figure 1. Sample network. A) The nodes 1 and 5 have highest degree centrality while nodes 5 and 7 cover the whole network. B) The nodes 4 and 8 have a similar degree, but node 8 attains higher Shapley value than 4. C) The hub nodes (nodes 5 and 1) are not directly connected to each other. D) The hub nodes (nodes 2 and 4) are directly connected to each other. The color intensity of each node represents the value of the Shapley value.

a- Efficiency: all payoffs have to be completely distributed among the players' i.e. $\sum_{i \in N} v(N)$

b- Symmetry: this principle indicates that if two players played equal roles in the payoff obtained by the coalition, they should enjoy that payoff to the same extent. That is, for two players i and j and for any coalition such as S consisting of i and j , if $v(S \cup \{i\}) = v(S \cup \{j\})$, then $\theta_i(v) = \theta_j(v)$.

c- Dummy axiom: if there exists a player i for any coalition S without i , such that $v(S) = v(S \cup \{i\})$, then $\theta(v) = 0$. That is, a player whose presence and absence in a coalition is the same, will have no share in the coalition payoff.

d- Additivity: if u and v were characteristic functions, then $\theta(v+u) = \theta(v) + \theta(u)$.

It can be proven that Shapley value is unique and is as follows(18):

$$\theta_i(v) = \sum_{S \subseteq N, i \in S} \frac{(|S|-1)!(n-|S|)!}{n!} [v(S) - v(S - \{i\})] \quad (1)$$

Let $\langle N, E \rangle$ be a network consists of a set of nodes N , and set of edges between nodes E . In this paper, we study the protein interaction network of infected cells to elucidate the rabies implicated signal transduction network. Here, E is a set of gene products and the set of edges E describes the interaction among them. The network constructed using meta-analyzing whole-transcriptome microarray datasets of the CNS infected by strain CVS-11, and integrating them with interactome data using computational and statistical methods (15).

According to Cesari et al. (13, 14), we consider the coalitional game $G = (N, v)$, where N is the set of gene products in the network and the characteristic function v is as follows:

A worth to each coalition of gene products $S \subseteq N$ is defined by:

$$v(S) = \sum_{j \in (S \cup Nei(S))} a_j \quad (2)$$

Where $Nei(S)$ is the gene products that are directly connected to some nodes in S . a_j represents the predefined importance of the gene product in the network. We will analyze the network using three scenarios. In the first scenario, all nodes are under similar conditions without any precondition and therefore, for all nodes a_j is equal to 1. In the second one, we assign weights to nodes with an n -tuple vector (n is the number of nodes in the network). These weights generally indicate the importance of the nodes in the network and can be assigned using different experimental and computational approaches. Here, we assign weights using the following procedure.

At first, the network was clustered using the Cluster ONE

algorithm, and then the number of nodes in each cluster was considered as the weight of that cluster nodes.

Cesari et al. (14) have shown that the Shapley value of this coalitional game can be calculated according to the following equation:

$$\theta_i(v) = \sum_{j \in (\{i\} \cup Nei(i))} \frac{a_j}{d_j+1} \quad (3)$$

where a_j is the weighting vector and d_j is the number of neighbors of each node. Naturally, all weights in the first scenario would be 1 and all weights in the second method would be equal to the number of nodes in each cluster to which the node belongs. According to equation (3), a node connected to a high number of low degree nodes attains a high Shapley value.

In the third scenario, the value of each coalition was defined as the number of coalition members plus their neighbors with more than one distance as described by Michalak et al (16). In this case, the characteristic function is defined by:

$$v(S) = size(\{i : \exists j \in S \mid distance(i, j) \leq k\}) \quad (4)$$

and the Shapley value is calculated as follows (16):

$$\theta_i(v) = \sum_{j \in (\{i\} \cup Nei(i,k))} \frac{1}{1 + deg(j,k)} \quad (5)$$

where $Nei(i,k)$ is the neighbors of node i at distance k and $deg(j,k)$ is the degree of node j at distance k .

Example 1: Consider the sample network represented in **Figure 1A**. According to the first scenario, the value of each coalition is a number of all unique nodes reachable from the coalition in at most one step. For example $v(\{1,5\}) = |\{1,5\}| + |\{2,3,4,6\}| = 6$. According to equation 3, the Shapley value vector is:

$$\theta(v) = \left(\frac{82}{60}, \frac{42}{60}, \frac{42}{60}, \frac{42}{60}, \frac{102}{60}, \frac{45}{60}, \frac{95}{60}, \frac{45}{60}, \frac{45}{60} \right)$$

Nodes v_5 and v_7 attain the highest Shapley value. The degree, closeness and betweenness centralities are as follows:

Degree = (3, 1, 2, 2, 5, 2, 3, 1, 1)

Closeness = (0.44, 0.4, 0.42, 0.42, 0.61, 0.57, 0.47, 0.33, 0.33)

Betweenness = (0.01, 0, 0, 0, 0.69, 0.53, 0.46, 0, 0)

Therefore, the Shapley value gives different ranking to 5 and 7 compared with classical centralities.

Note that the Shapley value of a node that interacts with low degree nodes is higher than the one whose neighbors have degrees greater than 1. Consider the following example.

Example 2: A sample network with 10 nodes is shown in **Figure 1B**. Nodes 4 and 8 have a similar degree and symmetric position in the network. Suppose again that all the nodes have the same prior weight equal to 1, i.e. $a_i = 1, \forall i \in N$. The Shapley value vector for this network is as follows:

$$\theta(v) = \left(\frac{6}{6}, \frac{5}{6}, \frac{5}{6}, \frac{11}{5}, \frac{4}{6}, \frac{4}{6}, \frac{4}{6}, \frac{13}{6}, \frac{4}{6}, \frac{4}{6}\right)$$

Node 8 that connected to nodes who themselves have a low degree gets higher Shapley value than the node 4. Therefore, removing the node with highest Shapley value would split the network to a maximum number of connected components and isolated nodes.

Example 3: Consider two sample networks in **Figures 1C and 1D**. In both cases, hub nodes have three neighbors. Shapley value for **Figures 1C and 1D** are $\theta(v) = \left(\frac{15}{12}, \frac{10}{12}, \frac{10}{12}, \frac{10}{12}, \frac{15}{12}\right)$ and $\theta(v) = \left(\frac{14}{12}, \frac{10}{12}, \frac{10}{12}, \frac{14}{12}\right)$, respectively. The hubs in the network in **Figure 1C** are not directly connected to each other while in **Figure 1D** hubs are neighbors and this decreased their Shapley value. Since in Shapley value the number of neighbors is very important, whether the hubs are neighbors or not is very effective in Shapley value. In the network we will analyze, the average of the distance between hubs is approximately 2. In order to diminish the effect of hubs in the Shapley value, we used scenario 3.

4. Results

In this study, we analyze the network of gene products implicate rabies. The network constructed by Jamalkandi et.al (15) by integrating the results of meta-analyzing whole transcriptome microarray data sets of the CNS infected by strain CVS-11 with interaction data using the computational method. In fact, they reconstructed a weighted protein-protein interaction network of infected cells based on differentially expressed genes to clarify the rabies-implicated signal transduction network. The weights are STRING combined scores (19). We selected the edges with a score greater than 0.4 to strictly filter weak and false positive interactions. The giant component of the network was obtained using CINNA package (20) with 1059 nodes and 8844 edges. In what follows, we analyze the network using three scenarios introduced in the Materials and Methods section.

4.1. First Scenario

In the first scenario, a group of gene products consists of its members plus their neighbors. The classical centralities including degree, closeness and betweenness and the Shapley value of associated game were calculated. The correlation coefficient, scatter plot, and distribution of centrality values between pairs of centralities are shown in **Figure 2**. Betweenness and Shapley value are highly correlated, meanwhile, the other pairwise correlations are not significantly correlated.

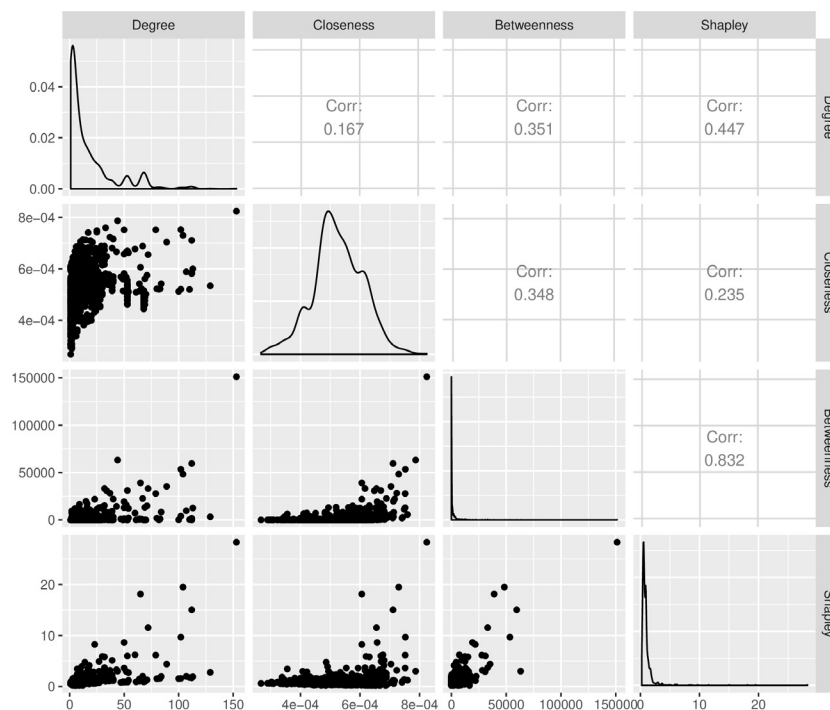


Figure 2. Scatter plots, distributions and the correlation coefficient between centralities in the first scenario.

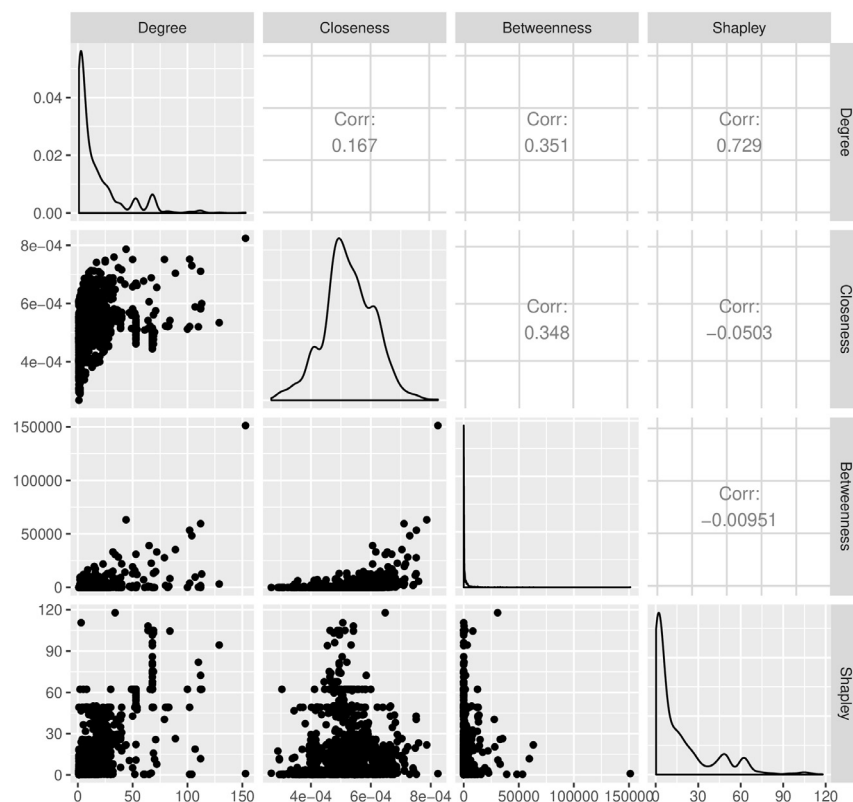


Figure 3. Scatter plots, distributions and the correlation coefficient between centralities in the second scenario.

In order to assess the effect of removing the top-ranked nodes on the network, the top twenty nodes according to betweenness centrality and Shapley value centrality were selected and removed from the network. In the first case (removing the nodes with highest betweenness centrality), the network was split into 106 components; the giant component with 909 nodes and the remaining components include 11, 5, 3, 2 nodes, and 94 single nodes. But, the network obtained by the removing the top 20 nodes according to Shapley value centrality was split into 166 components; the giant component with 953 nodes and the remaining components consist of 5, 4, 3, 2 nodes, and 151 single nodes. Therefore, high Shapley value nodes are able to interact directly with the maximum number of other nodes in the network and their removal would split the network into several connected components with few gene products or even isolated nodes.

In order to elucidate the beneficial role of the Shapley value centrality, we selected the genes among the highest Shapley value centrality ranks that had a low ranking from classical centralities. Three of the 40 top Shapley value-ranked genes, including TPK1, BMP2, and SMAD4 were not among the 80 top-ranked genes from classical centralities (degree, closeness, and betweenness).

4.2. Second Scenario

In this scenario, a weight vector was calculated that assigns a score for each node. For this purpose, at first, the network was divided into clusters using the ClusterONE algorithm (21). Genes can belong to multiple clusters and the weight of each gene in the network was defined as the maximum order of clusters it belongs to. The pair-wise correlations between any two of the centralities are shown in **Figure 3**. In this scenario, the correlation coefficient between degree centrality and Shapley value centrality was raised to 0.729, and the correlation coefficient between betweenness and Shapley value decreased to zero. Similar to the first scenario, we picked the genes among the highest Shapley value centrality ranks that had a low ranking from classical centralities. Four of the 40 top Shapley value-ranked genes, including IFI44, SLFN5, RTP4, and IRGM2 were not among the 130 top-ranked genes from classical centralities (degree, closeness, and betweenness). The top twenty nodes according to Shapley value centrality were selected and removed from the network. The network was split into 43 components; the giant component with 986 nodes and the remaining components consist of 5, 3, and 2 nodes, and 35 single nodes. The number of connected components compared to scenario 1 was

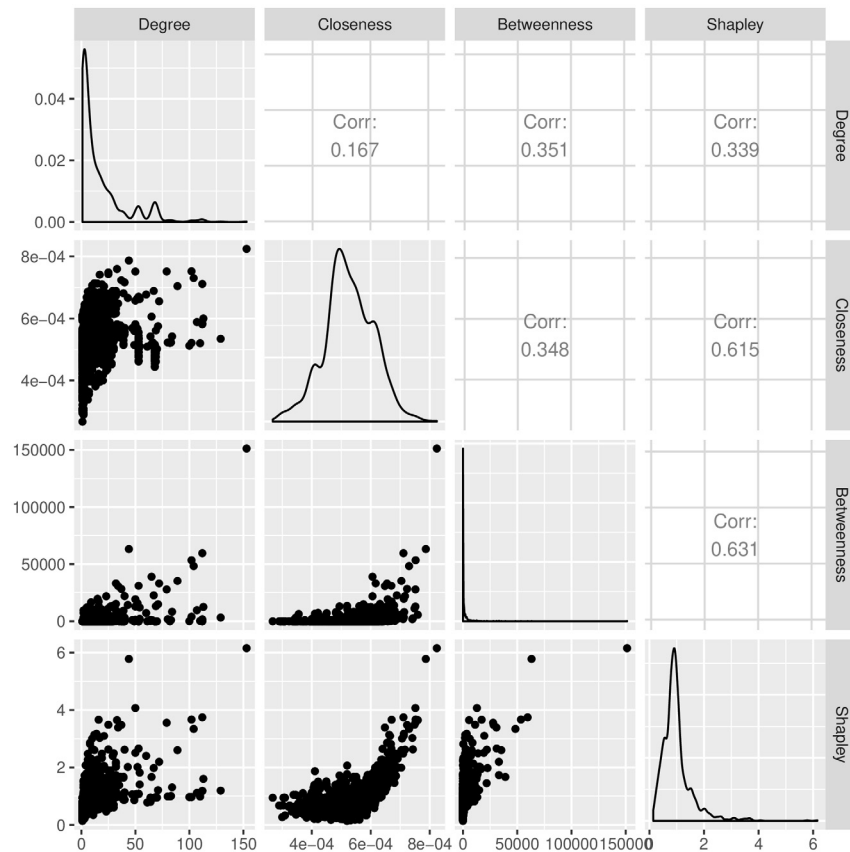


Figure 4. Scatter plots, distributions and the correlation coefficient between centralities in the third scenario.

Table 1. The ranking of three selected gene in scenario 1 according to different centrality measures.

Gene Name	Shapley value rank	Degree rank	Closeness rank	Betweenness rank	The average degree of its neighbors
TPK1	16	389	721	82	5.06
BMP2	32	465	966	113	5.09
SMAD4	36	262	111	120	16.52

significantly decreased and this could be a consequence of the weight of nodes based on the order of maximum cluster.

4.3. Third Scenario

The average of distances between hubs in the network was approximately 2. Therefore, in the third scenario, the value of each coalition was defined as the number of coalition members plus their neighbors with more than one distance. In fact, each coalition consists of nodes, their neighbors and neighbor of its neighbors. We expect that this coalition definition, reduce the correlation coefficient between degree centrality and Shapley value (**Fig. 4**).

The correlation coefficient of this centrality with closeness and betweenness centralities was about 0.60,

while as we expected at this scenario the correlation coefficient between degree and Shapley value decreased. Two of the 40 top Shapley value-ranked genes, including MAML1 and CITED2, were not among the 80 top-ranked genes from classical centralities (degree, closeness, and betweenness).

5. Discussion

In the first scenario, a coalition consists of its members and their neighbors and as explained in the Materials and Methods section a node connected to low degree nodes attains a high Shapley value. **Table 1** compares the ranking of three selected genes in scenario 1 according to different centrality measures. For each of the three selected genes in the first scenario, the average degree of its neighbors is very low as we expected.

Table 2. The enriched KEGG pathways using three selected genes in scenario 1.

KEGG Pathway	Adjusted P-Value
TGF-beta signaling pathway	0.001
<u>Hippo signaling pathway</u>	0.001
Thiamine metabolism	0.011
<u>Pathways in cancer</u>	0.011
Basal cell carcinoma	0.027
Th17 cell differentiation	0.027
<u>Colorectal cancer</u>	0.027
Chronic myeloid leukemia	0.027
<u>Pancreatic cancer</u>	0.027
Adherens junction	0.027
AGE-RAGE signaling pathway in diabetic complications	0.027
Signaling pathways regulating pluripotency of stem cells	0.027
<u>Cell cycle</u>	0.027
<u>Wnt signaling pathway</u>	0.027
Apelin signaling pathway	0.027
FoxO signaling pathway	0.027
<u>Hepatocellular carcinoma</u>	0.027
<u>Hepatitis B</u>	0.027
<u>Gastric cancer</u>	0.027
Human T-cell leukemia virus 1 infection	0.034
<u>Cytokine-cytokine receptor interaction</u>	0.043

Underlined/bold pathways are reported by Azimzadeh Jamalkandi et al (15)

To undertake enrichment analysis among the selected genes in scenario 1 (TPK1, BMP2, and SMAD4) Enrichr (22, 23) was used based on KEGG 2019 human. The results are listed in **Table 2**.

In this paper, we analyze the protein-protein interaction network implicates rabies with the objective of identifying the novel gene candidates acting as intermediaries between hub nodes and leaf nodes. For this purpose, we used a game-theoretic approach at three scenarios, where the power of a coalition of genes assessed using different criteria including the neighbors of genes in the network, and predefined importance of the genes in its neighborhood and the Shapley value of such a game was considered as a new centrality. Pair-wise correlation analysis reveals that changing the weights of nodes at different scenarios can significantly affect the results of ranking genes in the network. Therefore, some prior knowledge about the disease and the topology of the network, enable us to design an appropriate game and consequently infer some biological important nodes (genes) in the network. Obviously, a single centrality cannot capture all significant features embedded in the network.

Acknowledgments

The author's special thanks go to Dr. Mohieddin Jafari from the University of Helsinki for his precious comments and assistance on this manuscript.

Author's contribution

M.M. conceived and designed the experiments: R.B. performed the experiments and analyzed the data. M.M. Wrote the paper.

Conflict of interest

The authors declared no conflict of interest.

References

- Ashtiani M, Salehzadeh-Yazdi A, Razaghi-Moghadam Z, Hennig H, Wolkenhauer O, Mirzaie M, et al. A systematic survey of centrality measures for protein-protein interaction networks. *BMC Syst Biol.* 2018;**12**(1):80. doi: 10.1186/s12918-018-0598-2.
- Nieminen J. On the centrality in a graph. *Scand. J. Psychol.* 1974;**15**(1):332- 6. doi: 10.1111/j.1467-9450.1974.tb00598.x.
- Bavelas A. Communication Patterns in Task-Oriented Groups. *J Acoust Soc Am.* 1950; **22**(6):725- 730. doi: 10.1121/1.1906679.
- Freeman LC. A set of measures of centrality based on between-

- ness. *Sociometry*. 1977;**40**(1):35-41. doi:10.2307/3033543.
5. Carlson MR, Zhang B, Fang Z, Mischel PS, Horvath S, Nelson SF. Gene connectivity, function, and sequence conservation: predictions from modular yeast co-expression networks. *BMC Genomics*. 2006;**7**(1):40. doi: 10.1186/1471-2164-7-40.
 6. Azimzadeh S, Mirzaie M, Jafari M, Mehrani H, Shariati P, Khodabandeh M. Signaling network of lipids as a comprehensive scaffold for omics data integration in sputum of COPD patients. *BBA-MOL. Cell Biol. L*. 2015;**1851**(10):1383-93. doi:10.1016/j.bbali.2015.07.005.
 7. Mozhgani SH, Zarei-Ghobadi M, Teymoori-Rad M, Mokhtari-Azad T, Mirzaie M, Sheikhi M, et al. Human T-lymphotropic virus 1 (HTLV-1) pathogenesis: A systems virology study. *J Cell Biochem*. 2018;**119**(5):3968-79. doi: 10.1002/jcb.26546.
 8. Giorgi FM, Del Fabbro C, Licausi F. Comparative study of RNA-seq-and microarray-derived coexpression networks in *Arabidopsis thaliana*. *Bioinformatics*. 2013;**29**(6):717-24. doi: 10.1093/bioinformatics/btt053.
 9. Junker BH, Koschützki D, Schreiber F. Exploration of biological network centralities with CentiBiN. *BMC Bioinformatics*. 2006;**7**:219. doi: 10.1186/1471-2105-7-219.
 10. Jeong H, Mason SP, Barabási A-L, Oltvai ZN. Lethality and centrality in protein networks. *Nature*. 2001;**411**:41-42. doi: 10.1038/35075138.
 11. Galeotti A, Goyal S, Jackson MO, Vega-Redondo F, Yariv L. Network Games. *Rev Econ Stud*. 2010;**77**(1):218-244. doi: 10.1111/j.1467-937X.2009.00570.x
 12. Everett MG, Borgatti SP. The centrality of groups and classes. *J. Math. Sociol.* 1999;**23**(3):181-201. doi:10.1080/0022250X.1999.9990219.
 13. Cesari G, Algaba E, Moretti S, Nepomuceno JA. An application of the Shapley value to the analysis of co-expression networks. *Appl Netw Sci*. 2018;**3**:35. doi: 10.1007/s41109-018-0095-y.
 14. Cesari G, Algaba E, Moretti S, Nepomuceno JA. A game theoretic neighbourhood-based relevance index. *Complex Networks & Their Applications VI*. 2017:29-40. doi: 10.1007/978-3-319-72150-7_3.
 15. Azimzadeh Jamalkandi S, Mozhgani S-H, Gholami Pourbadie H, Mirzaie M, Noorbakhsh F, Vaziri B, et al. Systems biomedicine of rabies delineates the affected signaling pathways. *Front Microbiol*. 2016;**7**:1688. doi: 10.3389/fmicb.2016.01688.
 16. Michalak TP, Aadithya KV, Szczepanski PL, Ravindran B, Jennings NR. Efficient Computation of the Shapley Value for Game-Theoretic Network Centrality. *JAIR*. 2013;**46**:607-650. doi: 10.1613/jair.3806.
 17. Shapley LS. A value for n-person games. *Ann Math Stud*. 1953;**2**:307-317. doi:10.1515/9781400881970-018.
 18. Roth A. The Shapley Value: Essays in Honor of Lloyd S. Shapley. Cambridge: Cambridge University Press;1988. doi:10.1017/CBO9780511528446.
 19. Szklarczyk D, Franceschini A, Wyder S, Forslund K, Heller D, Huerta-Cepas J, et al. STRING v10: protein-protein interaction networks, integrated over the tree of life. *Nucleic Acids Res*. 2015;**43**:D447-452. doi: 10.1093/nar/gku1003.
 20. Ashtiani M, Mirzaie M, Jafari M. CINNA: an R/CRAN package to decipher Central Informative Nodes in Network Analysis. *Bioinformatics*. 2018;**35**(8):1436-1437. doi: 10.1093/bioinformatics/bty819.
 21. Nepusz T, Yu H, Paccanaro A. Detecting overlapping protein complexes in protein-protein interaction networks. *Nat Methods*. 2012;**9**:471-472. doi: 10.1038/nmeth.1938.
 22. Chen EY, Tan CM, Kou Y, Duan Q, Wang Z, Meirelles GV, et al. Enrichr: interactive and collaborative HTML5 gene list enrichment analysis tool. *BMC Bioinformatics*. 2013;**14**:128. doi: 10.1186/1471-2105-14-128.
 23. Kuleshov MV, Jones MR, Rouillard AD, Fernandez NF, Duan Q, Wang Z, et al. Enrichr: a comprehensive gene set enrichment analysis web server 2016 update. *Nucleic Acids Res*. 2016;**44**(W1):W90-W97. doi: 10.1093/nar/gkw377.