The Topological Properties of Virus-Human Protein Interaction Networks

Jinying Tan¹, ²             Xiufen Zou¹, *

¹ School of Mathematics and Statistics, Wuhan University, Wuhan 430072, China
² College of Science, Huazhong Agricultural University, Wuhan 430070, China

Abstract In the research of biological networks, it is an important problem how to use the quantity to identify essential protein-protein interactions. In this paper, a new definition “representative value of networks”, denoted as “RV”, is presented and applied to analyze the topological properties of three protein-protein interaction networks. From calculation we find that the relationship between human protein-protein interactions is highly clustered, for example, the percentage of value of the top 20% proteins of representative value from high to low accounts for 71.96%. The virus-human protein interactions (i.e. virus attacks human) are selective. And the attack of the virus also has a high aggregation. These results show that the proposed new definition is reasonable and can be considered as an important index for analyzing the topological properties of biological networks.

Keywords Protein-protein interaction network; Representative value (RV); Virus; degree

1 Introduction

Protein-protein interaction (PPI) plays a central role in many biological processes [1, 2], which not only is the base of normal physiological processes such as DNA replication, transcription, translation, metabolism, signal transduction and cell cycle control [3, 4], but also plays an important role in pathological processes [5, 6, 7]. In the viral infection of the host, it is also essentially expressed as interactions of viral proteins and host proteins, inhibiting activity of host proteins or denaturizing host proteins.

With the development in the high-throughput protein interaction detection technology such as the Yeast Two-Hybrid (Y2H) technology [8] and the tandem affinity purification - mass spectrometry technique (TAP-MS) [9], the protein-protein interaction networks of many species as data are uncovered, allowing the understanding of the process of life activities from the system-level of the protein-protein interaction networks [10, 11, 12, 13]. Access to the interaction networks of virus proteins and host proteins can also enable us to understand the mechanism of virus infecting host from the network-level and find new ways to solve the problems such as the toxicity for the virus which became stronger or

* Corresponding author, Email: xfzou@whu.edu.cn
weaker, differences among species, and the identification of targets for treatment and tumor pathogenesis.

Calderwood et al. [14] undertook a study for the interaction network of the Epstein-Barr virus (EBV) proteins and human proteins, from a broader perspective, EBV genes can be classified into two evolutionary classes, a class called the "core" genes, which contains conservative genes of the herpes viruses and their sub-classes, and the other for "non-nuclear" genes. Prior to that, the interaction networks for viral proteins have been reported for vaccinia virus (VV) [15], Varicella-Zoster virus (VZV) and Kaposi's sarcoma-associated herpes virus (KSHV) [16] by the yeast two-hybrid system. B de Chassey, et al. [17] performed a proteome-wide mapping of interactions between hepatitis C virus (HCV) and human proteins by Y2H and literature mining, to provide a comprehensive view of the cellular infection.

Recent researches showed that the attacked proteins in the virus-Human protein interaction networks have "Hub" characteristics and the higher average connectivity, corresponding to play an important role in their life activities [14, 15, 16, 17, 18]. For the above different viral infection networks, it is obviously a significant problem to study whether have some common features in network structure for attacking the host proteins of viral proteins. To reflect some common features of these networks, in this study, we propose a new definition "representative value of networks" based on the degree of networks, denoted as “RV” and conduct a comparative analysis of viral infection in the networks.

2 Data sets and Methods

2.1 Data sets

In this study, we use three data sets, the human protein-protein interaction network which is composed of 44223 non-redundant PPIs between 9520 different human proteins, the HCV-human protein interaction network which contains 481 HCV-human protein interactions between 11 virus proteins and 338 human proteins and the EBV-human protein interaction network which includes 173 different EBV-human protein interactions between 40 different EBV proteins and 112 human proteins, which are adapted from the literature [17] and [14] respectively.

2.2 Methods

The function of a protein can not be completed independently in the cell. Any life course is collaboratively completed by many proteins and other molecules together. In the interaction network, each protein can be seen as a node and each interaction as an edge, which forms a linked network of protein interactions. Thus we can research the topological properties of protein interaction networks from the perspective of graphic theory [19, 20, 21, 22, 23, 24]. To better compare different virus-human protein interaction networks, we introduce a concept "representative value".

Protein-protein interaction network can be defined as a simple undirected graph \( G = (V, E) \), which has \( n \) vertices and \( m \) edges. The graph vertex represents a protein...
and the edge represents the interaction between two proteins.

Let \( I_i = \{j \mid (v_j, v_i) \in E \} \) \((i = 1, 2, \ldots, n)\) and \(|I_i| = d_i\), i.e. the degree of node \( v_i \) is \( d_i \). If \(|I_i| = d_i\), then node \( v_i \) is claimed to vote to \( d_i \) nodes and the value of each vote of \( v_i \) is \( 1/d_i \). The representative value (RV) of node \( v_i \) is defined as \( RV_i = \sum_{j \in I_i} 1/d_j \). Obviously, the mean of representative value of all nodes is 1, i.e. \( \frac{\sum_{i=1}^{n} RV_i}{n} = 1 \).

![Fig.1 An example of a simple network.](image1)
![Fig.2 The deletion of P2.](image2)

We give a simple network as an example shown in Fig.1. The degree of P1, P2, P3, P4 and P5 is 3, 3, 3, 2 and 1, respectively. In order to reflect the various protein importance, we can vote and score to each protein: because the degree of P1 is 3, then the value of each vote of P1 is 1/3, and the value of each vote of P2, P3, P4 and P5 is 1/3, 1/3, 1/2 and 1/1, respectively. If \( P_i \) votes for \( P_j \), then \( P_j \) obtains the corresponding value of the vote. At the same time, the representative value of \( P_1 \) is \( P_2 + P_3 + P_4 = 1/3 + 1/3 + 1/2 = 1.1667 \). The rest may be deduced by analogy, the representative values of P2, P3, P4 and P5 are \( P_1 + P_3 + P_5 = 1/3 + 1/3 + 1/2 = 1.1667 \), \( P_1 + P_3 = 1/3 + 1/2 = 0.6667 \) and \( P_2 = 1/3 = 0.3333 \), respectively. The order of these nodes by "RV" is \( P_2 > P_1 = P_3 > P_4 > P_5 \), which is different from the order with the degree of nodes (\( P_2 = P_1 = P_3 > P_4 > P_5 \)). We can find that the node with the same degree possibly has different RV. In this simple network, it is obvious that three nodes P1, P2 and P3 have the same importance according to the degree. In fact, P2 is more important than P1 and P3 because if P2 is inhibited and lost its function, then the impact is not just P2 itself, and P5 will also be affected, whose function will be lost because at this time the only interaction with other nodes is cancel (Fig.2). Therefore, we could distinguish P2 from P1 and P3 by calculating the RV. In addition, this definition
also has an advantage, that is, the mean of the RV of all nodes in the whole network is always 1. When the virus proteins select to interact with the human proteins which have high representative value, then the mean of the RV of the sub-network of virus-human interaction will be higher than 1. Accordingly, a higher "RV" can be argued that the connectivity of the node in the network is more intense and the node is often more important.

3 Results

Here we calculate the RV of above three networks and list the names of the top 5 proteins of RV from high to low. We also calculate the percentage of total value of the top 20% proteins of RV from high to low and the results are shown in Table 1, Table 2 and Table 3, respectively.

The obvious feature of human protein-protein interaction network is highly clustered, for example, the percentage of value of the top 20% proteins of representative value from high to low accounts for 71.96%, which is similarity to the result of "Hub" characteristics [18].

From Table 1, the protein SLC2A4 has the highest RV, which is probably related with its structure and function. SLC2A4, a kind of protein involved in the glucose transport, is widespread in Skeletal and cardiac muscles, brown and white fat. The protein localizes primarily to the perinuclear region, undergoing continued recycling to the plasma membrane where it is rapidly reinternalized. The dileucine internalization motif is critical for intracellular sequestration. [25] From Table 2 and 3, we are lucky to find that HCV and EBV don’t interact with the human protein SLC2A4, defects in which may be a cause of noninsulin-dependent diabetes mellitus (NIDDM), or a cause of certain post-receptor defects in NIDDM. Otherwise, hepatitis and herpes disease would be more serous than we thought. [26, 27, 28]. The second highest RV protein ATXN1, whichLocates cytoplasm or nucleus, also expresses widely throughout the body. ATXN1 may be involved in RNA metabolism, Defects in ATXN1 are the cause of spinocerebellar ataxia type 1; also known as olivopontocerebellar atrophy I (OPCA I or OPCA1). Spinocerebellar ataxia is a clinically and genetically heterogeneous group of cerebellar disorders. [29, 30, 31].

YWHAG, UBQLN4, etc., also the characteristics of the protein in the body exists in most organizations, to interact with many proteins, some biological features common to complete, and therefore high RV value [31, 32]. Of course, low RV does not mean that the protein is not important, only because of its structure or its expression in some exceptional cases to complete certain specific features with other proteins together. The protein PGCP, a kind of plasma glutamate carboxypeptidase, for example, which RV is only 0.0012, up-regulated in the majority of hepatitis C virus-associated hepatocellular carcinoma [33]. Perhaps the low RV could also be the false negative test result because the interaction with other proteins is not detected in these experiments.

The virus-human protein interactions (i.e. virus attacks human) are selective, and the virus protein interacts often more easily with the high RV of human protein (such as human protein interaction with HCV, the mean of RV for HCV-human
### Table 1: Results of HIV-human Protein Interaction Networks

<table>
<thead>
<tr>
<th>Protein</th>
<th>Value</th>
<th>0.0025</th>
<th>0.0012</th>
<th>0.0102</th>
<th>0.0025</th>
<th>0.0012</th>
<th>0.0102</th>
<th>Value</th>
<th>0.0025</th>
<th>0.0012</th>
<th>0.0102</th>
</tr>
</thead>
<tbody>
<tr>
<td>TRAP1</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>CD79A</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>BCL6</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
</tbody>
</table>

### Table 2: Results of HCV-human Protein Interaction Networks

<table>
<thead>
<tr>
<th>Protein</th>
<th>Value</th>
<th>0.0025</th>
<th>0.0012</th>
<th>0.0102</th>
<th>0.0025</th>
<th>0.0012</th>
<th>0.0102</th>
<th>Value</th>
<th>0.0025</th>
<th>0.0012</th>
<th>0.0102</th>
</tr>
</thead>
<tbody>
<tr>
<td>TRAP1</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>CD79A</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>BCL6</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
</tbody>
</table>

### Table 3: Results of Influenza Protein Interaction Networks

<table>
<thead>
<tr>
<th>Protein</th>
<th>Value</th>
<th>0.0025</th>
<th>0.0012</th>
<th>0.0102</th>
<th>0.0025</th>
<th>0.0012</th>
<th>0.0102</th>
<th>Value</th>
<th>0.0025</th>
<th>0.0012</th>
<th>0.0102</th>
</tr>
</thead>
<tbody>
<tr>
<td>TRAP1</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>CD79A</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>BCL6</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
</tbody>
</table>

### Table 4: Results of FEPV-human Protein Interaction Networks

<table>
<thead>
<tr>
<th>Protein</th>
<th>Value</th>
<th>0.0025</th>
<th>0.0012</th>
<th>0.0102</th>
<th>0.0025</th>
<th>0.0012</th>
<th>0.0102</th>
<th>Value</th>
<th>0.0025</th>
<th>0.0012</th>
<th>0.0102</th>
</tr>
</thead>
<tbody>
<tr>
<td>TRAP1</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>CD79A</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>BCL6</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
</tbody>
</table>

These tables represent the topological properties of virus-human protein interaction networks, including the percentage of proteins from high to low and their corresponding ranks.
The attack of the virus also has a high aggregation, such as human protein interaction with HCV, of which 20% of the protein its RV accounts for 74.40%, and interaction with the EBV, of which 20% of the protein its RV accounts for 71.75%.

### 4 Conclusions

From these results, we can see that the purpose of the protein-protein interaction is strong. When the virus invades the body, the virus proteins often interact with the human proteins which have high RV, and inhibit the activity of these host proteins or alter their activity. Thus this causes to degrade their corresponding functions of these proteins. If the proteins which have high RV and tend to play more important role in normal physiological processes lose their functions, the body is possibly in the disease state. In conclusion, the proposed new definition is reasonable and can be considered as an important index for analyzing the topological properties of biological networks.

### Acknowledges

We thank Professor Zishu Pan at College of Life Sciences of Wuhan University for the helpful suggestions. We thank anonymous reviewers for manuscript comments. This work was supported by Chinese National Natural Science Foundation (Grant no. 30771597) and Self-research Project of Wuhan University (Grant no. 1081001).

### References


[7] Selkoe D J. The cell biology of beta-amyloid precursor protein and presenilin in


