CHAPTER 5

CONCLUSION AND DISCUSSION

In this thesis, we constructed the real human (*Homo sapiens*) PPI network, with high confident combine score yielded the property of scale-free network and made the connected network. This network had few numbers of high interaction proteins and a large number of small interaction proteins. Later on, we investigated attributes of each protein such as the average degree, global clustering coefficient, degree of correlation and gamma in the power-law form to determine the property of scale-free network and identified the disordered protein from *Online Mendelian Inheritance in Man* (OMIM) database.

In addition, we discovered the attribute degree divided by the average degree, $\dfrac{k}{<k>}$ and attribute sign of degree correlation, $sign(R)$ were significant and related to the property of disordered proteins affecting to scale-free network. These attributes were also related to the property of proteins affecting to scale-free network. These two attributes were selected to develop our new measures.

The purpose of this work was to develop the new measure of disordered proteins affecting to the scale-free network, $M_{SF\ Dsp} = \dfrac{k_i}{<k>} + 2.22 \cdot sign(R_i)$ and the new measure of identifying proteins affecting to the property of scale-free network, $M_{SF} = \dfrac{k_i}{<k>} + 1.75 \cdot sign(R_i)$. After that, we received the performance value AUC of approximately 0.92 and 0.94 in $M_{SF\ Dsp}$ and M_{SF}, respectively. These values were acceptable and showed our measures are good performance.

To investigate the impact of disordered proteins in the scale-free network structure, all disordered proteins were removed from the scale-free network. Then, the network lost the property of scale-free. Also, with various number of removing nodes (either disordered proteins or any proteins, in comparison), this also showed

that disordered proteins significantly affect to the property of scale-free network as well.

The advantage of this work is to validate the behavior of disordered proteins that related to the property of scale-free and also develop our new measures, $M_{SF\,Dsp}$ and M_{SF}. The new measures of $M_{SF.Dsp}$ and M_{SF} were useful to reduce the process of knocking out each protein and some neighbors of the protein to investigate the condition of power-law degree distribution for identifying the significant proteins affecting to scale-free network. The method of correlation measure (PCC) analysis was proposed to investigate the influential attributes and coefficients for developing new measures.

Since in this work we investigated only three attributes: the degree divided by average degree $\dfrac{k}{<k>}$, the clustering coefficient divided by the global clustering coefficient $\dfrac{c}{<c>}$ and sign of degree correlation $sign(R)$, to develop our measures, it might be more attributes to develop the measures. In future work, we will expand to develop the measures by using the technique of machine learning or trying to find more attributes for the better outcomes.

Moreover, the case of developing the measure $M_{SF\,Dsp}$ with attribute $\dfrac{k}{<k>}$ and attribute $\dfrac{c}{<c>}$, $M_{SF\,Dsp} = 2.00 \cdot \dfrac{k_i}{<k>} + \dfrac{c_i}{<c>}$ showed that the value of AUC was 0.72. In addition, the value AUC of another measure, $M_{SF} = 2.23 \cdot \dfrac{k_i}{<k>} + \dfrac{c_i}{<c>}$ was 0.78. This situation meant that the performance in measures with attribute $\dfrac{c}{<c>}$ were less than the performance in our measures.