Poster B-32

A Classification for Conceptually Layouting Protein-Protein Interaction Networks

Authors:
SunLee Bang (*Electronics & Telecommunications Research Institute*)
JaeHun Choi (*Electronics & Telecommunications Research Institute*)
JongMin Park (*Electronics & Telecommunications Research Institute*)
SooJun Park (*Electronics & Telecommunications Research Institute*)

Short Abstract: To better understand interactions and protein function in the protein-protein interactions network, we nearly layout the proteins that do same function with keeping physical structure using annotated information of the protein. With this method, we can conceptually analyze the network and visualize it faster than existing graph drawing algorithms

Long Abstract:

A protein performs inherent functions, with variously interacting with several other proteins to play a particular biological role in a living cell [1]. To understand the biological events, the interactions are viewed as a network, which consists of nodes and edges among them[2]. In the network, proteins and their interactions are represented as nodes and edges respectively. Although the many researches are proposed to layout them, most of them take only optimal topologies for networks into consideration. For optimally visualizing the topology of proteins included in the network, Osprey [3] employs a multilevel algorithm for the force-directed placement (MFDP), which is developed for drawing large graphs typically.

The MFDP, improving a force-directed placement (FDP) algorithm, arranges a network balanced by forces having an influence on nodes and edges. The arrangement first merges the set of nodes through a multilevel process, and then uses the FDP algorithm in the expanding process to reduce the layout time [4]. To nearly lay nodes which belong to the same classification out, we improved choosing an initial node, coarsening graph, expanding graph of the MFDP. First, an initial node that is a basis of coarsening a graph is chose as the following order. In the PPI network, proteins that have many relationships with other proteins are more important, therefore we chose a node which has the most neighbor in a connected graph as the initial node. If there are nodes which have the same neighbor count, we chose a node which has the least sum of the neighbor node's degree to prevent intensively clustering with a node. We create a replaced node which merges every neighbor of the initial nodes at a time until the coarsened graph downs to 3 nodes. This merging method reduces processing steps in the graph coarsening. In expanding graph, we set partition points as merged node number for a replaced node. The points evenly located on the circle of natural spring force constant in diameter, and expanded nodes are located to the points. If the expanded node has relationship with nodes of the pre-step, then the node is located to the near point with nodes of the pre-step. If not, the expanded node is located to an empty point.

In addition, the MFDP nearly place merged nodes together, so that we first merge nodes that are annotated by GO[5] terms of same hierarchy to conceptually understand the PPI

network. GO terms consisted of hierarchy of three categories (BP: Biological Process, CC: Cellular Component, MF: Molecular Function). To lay proteins by function out, we classify protein nodes according to BP and MF categories. Levels of annotated GO terms for a protein node are different, so that the terms are converted to terms of an equal level. Proteins, having same GO terms, are handled as uniform classification. Considering the classification, choosing the initial node and coarsening graph are expanded as following. First, the node that has the most neighbor is chose as the initial node among annotated nodes by GO terms. If nodes more than two are selected then annotated node by more GO terms is selected in the first. In the graph coarsening, protein nodes are annotated with several GO terms in a category. So, we coarsen a graph with merging nodes when there is intersection among annotated terms of nodes. The merged nodes are annotated by union among terms to preserve annotation information of nodes. In case to consider together BP and MF categories, we set the weight whose range is from 0.5 to 1 to the merged node. Nodes that have same GO terms for all of two categories are set as weight value of 1. Nodes that have same GO terms for just one category are set as less than 1 weight through intersection of GO terms. This weight is applied to calculate force in the FDP algorithm of expanding graph. After getting the coarsest graph through above process for a specific level of GO, we repeat the graph coarsening process until the level downs to 1.

References

[1] P. Uetz and R. L. Jr. Finley, "From protein networks to biological systems," FEBS Letters, Vol. 579 No. 8, pp. 1821-1827, 2005.
[2] P. Uetz, T. Ideker and B. Schwikowski, "Visualization and Integration of Protein-Protein Interactions," E. Golemis, (ed.) Protein-Protein Interactions - A Molecular Cloning Manual. Cold Spring Harbor Laboratory Press, pp. 623-646.
[3] B. J. Breitkreutz, C. Stark and M. Tyers, "Osprey: a network visualization system," Genome Biology, Vol. 4, Issue 3, Article R22, 2003
[4] C. Walshaw, "A Multilevel Algorithm for Force-Directed Graph-Drawing," Journal of Graph Algorithms and Applications, Vol. 7, No. 3, pp. 253-285. 2003.
[5] Gene Ontology, http://www.geneontology.org