A Grid Layout Algorithm for Biological Networks by Utilizing Biological Attributes

Kaname Kojima  
kaname@ims.u-tokyo.ac.jp

Masao Nagasaki  
masao@ims.u-tokyo.ac.jp

Enua Jeong  
eajeong@ims.u-tokyo.ac.jp
Mitsuru Kato  
mitsuru@ims.u-tokyo.ac.jp
Satoru Miyano  
miyano@ims.u-tokyo.ac.jp

Human Genome Center, Institute of Medical Science, University of Tokyo, 4-6-1 Shirokane-dai, Minato-ku, Tokyo 108-8639, Japan

Keywords: grid layout, graph layout, biopathway, time complexity, search space, biological attribute

1 Introduction

Clearly visualized biological pathways provide a great help in understanding biological systems. However, manual drawing of large-scale pathways is a time consuming task. Thus far, various layout algorithms aiming at biological networks were proposed, here we introduce one of grid layout based approaches termed SCCB-grid layout [1].

In the grid layout, the positioning of nodes are limited to grid points, and a cost function to minimize is defined by reflecting edge-edge crossings, node-edge crossings and distance between nodes. Since even finding the layout with the minimum edge-edge crossings is NP-hard [2], extant algorithms repeatedly update the layout by moving one node under greedy search strategy, and a locally optimal layout is obtained after convergence.

[1] argued the necessity of simultaneous movements of multiple nodes at each updating step from the practical trials, then proposed a method that considers swapping nodes as well as moving a node without the increase of the time complexity. In addition, they focused on the fact that nodes having the same biological attribute tend to be aligned in manually drawn layouts, as a result introduced a new function that handles this property. These two ideas were implemented as new grid layout algorithm termed SCCB-grid layout algorithm. The next section describes SCCB-grid layout algorithm briefly and shows resulting layouts of the algorithm.

2 Method and Result

Given a graph \( G = (V, E) \) with nodes \( V \) and edges \( E \), a layout \( L = (V, E, U, P) \) of \( G \) consists of the underlying graph \( G \), grid points \( U \) and a function \( P : V \to U \) such that \( P(v_\alpha) \neq P(v_\beta) \) for any two distinct nodes \( v_\alpha, v_\beta \in V \). We first introduce following functions to define a cost function:

- \( \text{Cross}_{e_i,e_j}(L) \): a binary function that returns 1 if an edge \( e_i \) crosses with an edge \( e_j \) and 0 otherwise.
- \( \text{Cross}_{v_i,v_j}(L) \): a binary function that returns 1 if an edge \( e_j \) crosses with a node \( v_i \) and 0 otherwise.
- \( \text{Distance}_{v_i,v_j}(L) \): a function that returns \( w_{v_i,v_j} \cdot \text{md}(v_i, v_j) \), where \( w_{v_i,v_j} \) is the weight to the couple of nodes \( v_i \) and \( v_j \), and \( \text{md}(v_i, v_j) \) is the Manhattan distance between \( v_i \) and \( v_j \).
- \( \text{Combo}_v(L) \): a function that returns negative cost for a vertex \( v \) depending on how vertices having the same attribute as \( v \) are aligning around \( v \). Its detailed definition is in [1].
By using the above functions, the layout cost $C(L)$ of $L$ is defined as follows:

$$C(L) = W_{ee} \sum_{e_i, e_j \in E} \text{Cross}_{e_i, e_j}(L) + W_{ne} \sum_{v \in V, e \in E} \text{Cross}_{v, e}(L) + W_{dc} \sum_{v_i, v_j \in V} \text{Distance}_{v_i, v_j}(L) - W_{cs} \sum_{v \in V} \text{Combo}_v(L),$$

where $W_{ee}$, $W_{ne}$, $W_{dc}$ and $W_{cs}$ are called respectively edge-edge crossing weight, node-edge crossing weight, distance cost weight and combo score weight, which are used for adjusting the effect of each factor.

At each step, SCCB-grid layout calculates costs of all layouts that can be generated by moving one of all nodes to one of all vacant points or swapping all pairs of nodes. The layout with the minimum cost is selected as a starting layout for the next step. After convergence, a locally optimal layout is obtained. Since calculating all possible adjacent layouts requires high time complexity in a naïve way, SCCB-grid layout stores the cost differences from possible movements at the previous step in a two-dimensional array termed $\Delta$ matrix. By using these cost differences in $\Delta$ matrix, we succeed in reducing the time complexity to $O(|V|^2 + |E|^2 + |U| \cdot \deg(v_\beta) \cdot (|V| + |E|))$ at each step, where $\deg(v_\beta)$ is the degree of the node moved at the previous step. See [1] for the details of the search strategy.

For the evaluation, we create a pathway model of the endothelial cell from the biological literature in [3]. We randomly map nodes in the pathway model to grid points (Figure 2(a)) and apply SCCB-grid layout to the randomly generated layout. Figure 2(b) shows its resulting layout.

References

