

대사 경로 시각화를 위한 레이아웃 알고리즘 연구

송은하*, 용승림**

A Study on layout algorithm for metabolic pathway visualization

Eun-Ha Song*, Seunglim Yong**

요약

생명체 내에서 나타나는 네트워크 중 하나인 대사 경로는 화합물의 상호 관계를 그래프를 통해 표현된다. 대사 경로는 본질적인 복잡성 때문에 대사 경로 내의 흐름을 한 눈에 알 수 있도록 가시화하여 보여 주는 도구가 반드시 필요하다. 이러한 가시화 도구의 경우 노드수가 증가할수록 에지 교차가 기하급수적으로 증가하는 문제가 있다. 따라서 유전체 수준의 대사 경로를 연구하기 위해서는 대사 경로 그래프 레이아웃 상에 나타나는 에지 교차를 줄이는 것이 시각화의 매우 중요한 부분이다. 본 논문은 생물학에서의 대사 경로에 대한 시각화를 위한 2-계층을 이용한 대사 경로 레이아웃 알고리즘을 제안한다. 대사 경로의 구조적 특징을 고려하여 대사 경로 그래프에 존재하는 주요 컴포넌트인 환형 컴포넌트 또는 연결성 높은 노드를 찾아 상위 계층에 레이아웃 하고 나머지 부분그래프는 하위 계층에 레이아웃 한다. 제안한 대사 경로 레이아웃 알고리즘은 유연성 있는 대사 경로 분석의 기초가 된다.

▶ Keywords : 대사 경로, 시각화, 레이아웃 알고리즘

Abstract

In metabolomics, metabolic pathway is represented by well-displayed graph. Metabolic pathways, especially, have a complex binding structure, which makes the graphical representation hard to visualize. There is a problem that edge crossings exponentially increase as the number of nodes grows. To apply automatic graph layout techniques to the genome-scale metabolic flow of metabolism domains, it is very important to reduce unnecessary edge crossing on a metabolic pathway layout. we proposed a metabolic pathway layout algorithm based on 2-layer layout. Our algorithm searches any meaningful component existing in a pathway, such as circular components,

•제1저자 : 송은하 •교신저자 : 용승림

•투고일 : 2013. 3. 9, 심사일 : 2013. 3. 29, 게재확정일 : 2013. 4. 10.

* 이즈텍(ISTech)

** 인하공업전문대학 컴퓨터시스템과(Dept. of Computer Systems and Engineering, Inha Technical college)

highly connected nodes, and the components are drawn in upper layer. Then the remaining subgraphs except meaningful components are drawn in lower layer by utilizing a new radial layout algorithm. It reduces ultimately reduced the number of edge crossings. This algorithm is the basis of flexible analysis for metabolic pathways.

▶ Keywords : Metabolic pathway, visualization, layout algorithm

I. 서 론

생물체 내에서는 생명 활동을 유지하기 위해서 매우 많은 기능들이 수행되고 있다. 이러한 기능들은 세포 내외에 존재하는 단백질 및 여러 물질들 간의 복잡한 네트워크를 이루고 있다. 생물체 내에서 나타나는 네트워크는 그 기능과 참여하는 물질에 따라서 크게 대사 경로(metabolic pathway), 조절 경로(regulatory pathway), 신호 전달 경로(signal transduction pathway)의 세 가지로 구분된다[1]. 생물체 내에서 나타나는 네트워크 중 가장 기본적인 대사 경로에 대한 정보는 계속 새로운 정보를 반영하고 있으며, 연구를 수행하는 생물학자들은 KEGG, EcoCyc, WIT 등의 데이터베이스 내에 저장된 대사 경로 정보를 검색하여 필요한 정보를 얻어낸다. 검색된 결과는 대사 경로의 본질적인 복잡성 때문에 대사 경로의 흐름을 한 눈에 쉽게 이해하고 파악할 수 있는 가시화 도구가 반드시 필요하다[2, 3, 4, 5, 6, 7, 8, 9, 10].

경로 정보를 시각적인 정보로 표현하기 위한 대사 경로 시각화 시스템은 복잡한 대사 경로의 구조를 한 눈에 쉽게 파악할 수 있어 실제 연구를 수행하는 생물학자들에게는 필수적인 요소이다. 현재까지 개발된 대부분의 가시화 도구 및 알고리즘은 2차원(2D) 평면에 대사 경로를 레이아웃 하는 방법을 사용한다. 그러나 이러한 필요성에 의해 개발된 기존의 시스템들은 아직 많은 문제점들을 가지고 있다.

본 논문에서는 대사 경로의 구조적인 특징을 파악하고 이를 반영한 2.5차원 대사 경로 레이아웃 알고리즘을 설계하고 구현한다. 2.5차원 대사 경로 레이아웃 알고리즘은 단순히 3차원 공간상에 대사 경로 그래프를 표현하는 것이 아니라 그래프 이론에 기반하여 대사 경로의 구조적인 특징을 분석하고 이를 반영한 레이아웃 알고리즘이다[11, 12, 13, 14]. 3차원 상에 평행한 2개의 계층(layer)을 두어 대사 경로 그래프를 계층화하여 레이아웃 한다. 따라서 표현영역을 3차원으로

확장함으로써 대사 경로의 구조 분석을 용이하게 하고 기존 2차원 레이아웃과 비교하여 에지 교차를 줄이는 결과를 얻는다. 2장에서는 기존 연구 내용을 살펴보고 3장에서는 제안하는 레이아웃 알고리즘을 설명한다. 4장에서는 실험 결과를 제시하고 5장에서 결론을 맺는다.

II. 관련 연구

대사 경로는 본질적인 복잡성 때문에 대사 경로 내의 흐름을 한 눈에 쉽게 이해하고 파악할 수 있는 대사 경로 가시화 도구가 반드시 필요하다. 기존의 2차원 대사 경로 레이아웃 알고리즘들은 다음과 같다.

2.1 EcoCyc에 적용된 알고리즘

*Escherichia coli*의 전체 지놈과 대사 경로 및 조절 경로 등을 검색할 수 있는 데이터베이스 시스템인 EcoCyc에서는 그래프의 위상(topology)을 환형, 계층적, 선형, 복합형의 네 가지로 나누어 각각 '환형 레이아웃 알고리즘', '계층적 레이아웃 알고리즘', '선형 레이아웃 알고리즘(linear layout algorithm)'을 적용한다. 복합형인 경우에는 '트리 레이아웃 알고리즘(tree layout algorithm)'을 적용하여 전체 대사 경로에 대한 레이아웃 결과를 얻는다. 이 알고리즘은 '트리 레이아웃 알고리즘'을 적용할 때, 슈퍼 노드 내의 노드들의 실제적 배치를 고려하지 않기 때문에, 발생할 에지 교차를 미리 예측하여 제거하는 것이 불가능하므로 전체적인 레이아웃 결과가 적절하지 않을 수 있다. 또한 실제 EcoCyc에서 제공하는 이미지들은 매우 작은 대사 경로들만을 표현하고 있고 큰 대사 경로를 표현해주고 있지 않다는 문제점이 있다[15].

2.2 환형·계층적인 특징에 기반을 둔 알고리즘

이 알고리즘은 M. Y. Becker와 I. Rojas가 제안한 알고리즘으로, 대사 경로 그래프의 위상을 환형 컴포넌트가 없는 경우, 그래프 자체가 환형 컴포넌트인 경우, 그리고 이 두 가

지 경우에 모두 속하지 않을 경우의 세 가지로 나누어 각각에 적합한 레이아웃 알고리즘을 적용한다[16]. 이 알고리즘은 화합물 그래프를 이용하여 대사 경로 그래프를 표현하고 있다. 화합물 그래프로 대사 경로 그래프를 표현하게 되면 화합물 간의 다(多) 대 다(多) 관계를 표현하기 위해 하이퍼에지가 나타난다. 이러한 하이퍼에지는 레이아웃 결과에 잦은 에지 교차를 발생시킴으로써 가독성을 떨어뜨린다는 문제점이 있다.

2.3 W.Li와 H.Kurata의 알고리즘

W. Li와 H. Kurata는 force directed 알고리즘의 문제점을 보완한 레이아웃 알고리즘을 제안하였다. force directed 알고리즘은 노드 사이에는 척력, 에지에는 인력이 작용하면서 서로 결합된 역계(force system)로 표현하고, 모든 척력과 인력 작용은 평가함수(cost function)인 에너지함수(energy function)에 의해 표현되도록 하였다. 그러나 척력이 작용하는 노드들은 흩어져 있고, 인력이 과하게 작용하는 노드는 거리가 너무 짧은 등 균일하고 촘촘한 레이아웃 결과를 얻기에 적당하지 않았다[17]. W. Li와 H. Kurata는 네트워크의 위상 구조에 기반한 새로운 평가함수를 정의하고 평가함수에 따른 여러 개의 후보 레이아웃을 선정 후 시뮬레이티드 어닐링(simulated annealing) 과정을 통해 휴리스틱한 방법으로 최종 레이아웃 결과를 얻음으로써 노드 간의 관계를 그래프의 위상에 의해 결정하여 전체 네트워크의 구조를 잘 표현한다. 하지만 이 알고리즘은 일반적인 네트워크를 레이아웃 하기 위해 설계된 것으로 대사 경로 그래프에 적용하기에 부적절하다.

III. 대사 경로 시각화를 위한 레이아웃 알고리즘

대사 경로 그래프를 두 개의 계층으로 나누어 레이아웃하기 위해 주어진 대사 경로 그래프를 상위계층, 하위계층에 레이아웃 할 개체로 나눈다. 3차원 상의 2계층은 각각 상위 계층(upper layer), 하위 계층(lower layer)이라 하고 각각의 계층에 상위계층 레이아웃 알고리즘과 하위계층 레이아웃 알고리즘을 적용한다. 대사 경로 그래프가 입력으로 들어올 때 그 그래프에 환형 컴포넌트의 존재 유무에 따라 [그림1]과 같이 해당 계층에 레이아웃 하기 위하여 각 알고리즘을 적용한다.

환형 컴포넌트가 존재하는 경우에는 대사 경로 그래프상에 존재하는 환형 컴포넌트를 모두 찾고, 환형 컴포넌트를 제외한 나머지 부분을 연결 성분으로 나눈다. 찾아낸 모든 환형

컴포넌트에는 '상위 계층 레이아웃 알고리즘(upper-layer layout algorithm)'을, 환형 컴포넌트를 제외한 나머지 연결 성분에는 '하위 계층 레이아웃 알고리즘(lower-layer layout algorithm)'을 적용하여 레이아웃 하여 결과를 얻는다.

환형 컴포넌트가 존재하지 않는 경우에는 대사 경로 그래프에 존재하는 차수가 6 이상인 모든 노드를 찾아 연결성 높은 노드로 정의한다. 연결성 높은 노드들은 상위 계층 레이아웃 알고리즘을 적용하고, 나머지 부분은 다시 연결 성분으로 나누어 하위 계층 레이아웃 알고리즘을 적용한다.

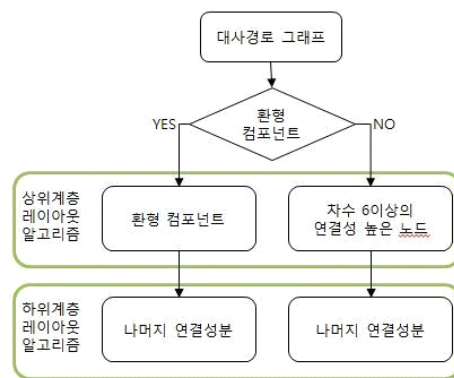


그림 1. 대사 경로 레이아웃 알고리즘
Fig 1. metabolic pathway layout algorithm

3.1 상위계층 레이아웃 알고리즘

입력으로 주어진 대사 경로 그래프에 존재하는 주요 컴포넌트인 환형 컴포넌트, 연결성 높은 노드들은 3차원상의 두 개의 계층 중 상위 계층에 레이아웃 된다.

3.1.1 환형 컴포넌트가 존재하는 경우

대사 경로 그래프의 환형 컴포넌트는 각각을 하나의 슈퍼노드로 간주한다. 슈퍼노드란 여러 개의 관련 있는 정점을 하나의 노드로 간주하는 것으로, 각각 일정 크기를 갖는 원으로 간주하여 원의 반지름을 사이즈로 갖는다. 슈퍼노드화된 환형 컴포넌트들은 그 컴포넌트의 반지름과 중심 좌표값, 그리고 다른 환형 컴포넌트와의 사이에 존재하는 에지 수에 대한 정보를 갖는다.

슈퍼노드들은 에지의 수에 따라 '중앙 슈퍼노드(center super node)'와 '외부 슈퍼노드(outer super node)'의 두 그룹으로 나눈다. '중앙 슈퍼노드'는 슈퍼노드들 중 환형 컴포넌트 사이에 존재하는 에지가 가장 많은 슈퍼노드로 정하고, 중앙 슈퍼노드를 제외한 나머지 슈퍼노드를 '외부 슈퍼노드'로 정한다.

슈퍼 노드의 그룹이 모두 정해지면, 중앙 슈퍼 노드는 상위 계층의 중심에 레이아웃 하고, 외부 슈퍼 노드들은 중앙 슈퍼 노드를 둘러싸는 원을 따라 방사 형태로 레이아웃 한다. 방사 형태의 레이아웃은 슈퍼노드들을 루트를 중심으로 하여 트리 형태의 구조를 갖는 방사모양으로 위치시키면 된다[4].

외부 슈퍼 노드를 중앙 슈퍼 노드의 외부 원에 겹치지 않고 레이아웃하기 위해서는 외부 원의 사이즈를 결정하고, 각 외부 원의 위치를 지정하는 두 과정을 수행해야 한다. 중앙 슈퍼 노드와 외부 원은 [그림 2]와 같다.

1) 외부 원의 사이즈

외부 원의 사이즈는 중앙 슈퍼 노드와 외부 슈퍼 노드의 반지름으로 계산한다. 중앙 슈퍼 노드를 C_{center} , 외부 슈퍼 노드를 C_1, C_2, \dots, C_n 이라 하고, 외부 슈퍼 노드 C_i 의 반지름을 r_i 이라 정의할 때, 외부 원의 반지름(R_{outer})을 구하는 식은 식(1)과 같다.

$$R_{outer} = \max \left\{ R_{center} + r_{\max}, \frac{\sum_{i=1}^n r_i}{\pi} \right\} \quad (1)$$

r_i : 외부 슈퍼 노드 C_i 의 반지름
 r_{\max} : r_i 의 최대값

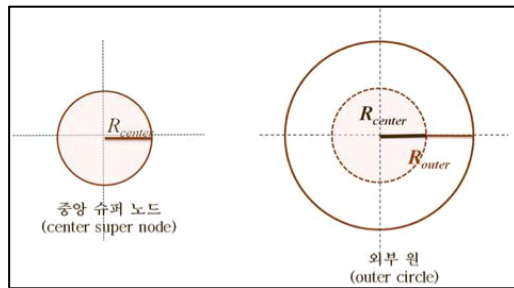


그림 2. 중앙 슈퍼 노드와 외부 원
 Fig 2. center supernode and outer circle

2) 외부 원의 위치 지정

외부 슈퍼 노드들은 위치를 결정하기 위해서 각 외부 슈퍼 노드의 무게 중심(center of mess)을 구한다. 외부 슈퍼 노드의 무게 중심은 중앙 슈퍼 노드와 해당 외부 슈퍼 노드와 연결되어 있는 노드들의 평균 좌표값으로부터 구한다. 계산된 외부 슈퍼 노드의 무게 중심에 따라 외부 슈퍼 노드의 위치가 결정된다.

외부 슈퍼 노드가 하나 존재하는 경우에는 외부 원의 둘레

의 점들 중 무게 중심과 가장 가까운 점을 계산하여 외부 슈퍼 노드를 위치시킨다.

외부 슈퍼 노드가 2개 이상인 경우에는 각 외부 슈퍼 노드의 무게 중심과 가까운 초기 위치를 선택한 후, 외부 슈퍼 노드들이 보기 좋게 레이아웃 되도록 위치를 조정하여 레이아웃 한다.

외부 원의 반지름의 길이가 $R_{outer} = \frac{\sum_{i=1}^n r_i}{\pi}$ 의 값으로 선택되는 경우는 [그림 3]과 같이 외부 슈퍼노드가 겹쳐서 레이아웃 될 수 있기 때문에 위치를 조정해야 한다.

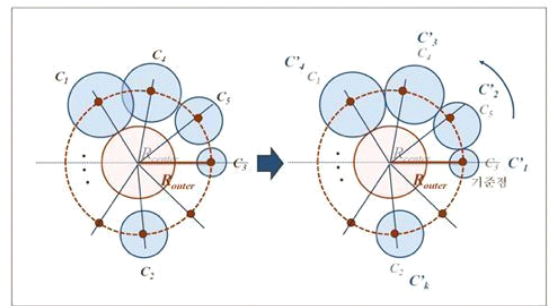


그림 3. 외부 슈퍼 노드의 레이아웃 위치
 Fig 3. outer supernode layout

외부 슈퍼 노드의 초기 위치는 외부 슈퍼 노드의 개수 n 에 따라 외부 원의 둘레를 n 개로 나눈 위치에 해당되는 점으로 한다. 초기 위치가 정해지면, 각 외부 슈퍼 노드는 외부 슈퍼 노드의 레이아웃 위치 각도(positioning angle)를 계산하여 x축을 기준으로 반시계방향으로 위치를 조정한다.

[그림 4]와 같이 중앙 슈퍼 노드의 중심으로부터 외부 슈퍼 노드 C_i 에 그은 접선과 각 노드의 중심을 잇는 직선 사이의 각을 θ_{oi} 라 정의할 때 θ_{oi} 는 다음의 식(2)와 같이 계산된다.

$$\theta_{oi} = \sin^{-1} \left(\frac{r'_i}{R_{outer}} \right) \quad (2)$$

따라서 위치 각도 θ_i 는 식(3)과 같이 구할 수 있다.

$$\theta_i = \theta_{i-1} + (\theta_{oi} + \theta_{o(i-1)}) \quad (3)$$

θ_i : 외부 슈퍼노드 C_i 의 위2치 각도

외부 원의 반지름의 길이가 $R_{outer} = R_{center} + r_{\max}$ 의 값으로 선택되는 경우는 [그림 4]와 같이 외부 슈퍼 노드는

서로 인접하지 않는다. 따라서 외부 슈퍼 노드를 외부 원의 둘레에 레이아웃 해 줄 때 외부 슈퍼 노드 사이에 간격이 일정하도록 위치 각도를 계산한다.

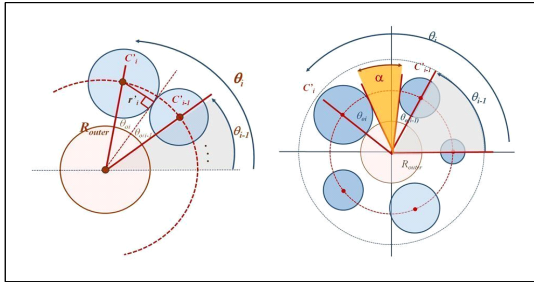


그림 4. 외부 슈퍼 노드의 위치 각도 조정
Fig 4. calculate positioning angle of outer supernode

3.1.2 연결성 높은 노드가 존재하는 경우

연결성 높은 노드가 존재하는 경우는 환형 컴포넌트가 존재하는 경우와 마찬가지로 연결성 높은 노드들을 방사 형태로 레이아웃 하게 된다.

대사 경로에 존재하는 모든 연결성 높은 노드들 중 에지 수가 가장 많은 노드를 중앙 슈퍼 노드로 정한다. 그리고 나머지 연결성 높은 노드를 외부 슈퍼 노드로 지정한다. 슈퍼 노드들은 중심 좌표값과 각 연결성 높은 노드 간에 존재하는 에지 수를 정보로 갖는다.

환형 컴포넌트가 존재하는 경우에서와 마찬가지로, 중앙 슈퍼 노드로 결정된 연결성 높은 노드는 상위 계층의 중심에 레이아웃 하고, 나머지 연결성 높은 노드들은 외부 원의 둘레에 레이아웃 하여 최종 레이아웃 결과를 얻는다.

3.2. 하위계층 레이아웃 알고리즘

대사 경로 그래프에 존재하는 주요 컴포넌트를 제외한 나머지 연결 성분들은 하위계층에 레이아웃 한다. 나머지 연결 성분들은 환형 컴포넌트가 존재하지 않는 트리 형태의 계층적인 구조를 가지거나 선형 구조를 가지게 된다. 연결 성분의 구조가 트리 형태인 경우에는 방사 레이아웃 알고리즘을 적용하고, 선형인 경우에는 환형 레이아웃 알고리즘을 적용하여 선형 구조를 원모양으로 배치하여 레이아웃 영역이 불필요하게 넓어지는 것을 방지한다. 각 연결 성분은 연결 성분을 포함하는 최소 환형 컴포넌트로 간주하고, 각각을 슈퍼 노드로 표현한다.

상위 계층에 레이아웃 된 주요 컴포넌트와의 에지 수를 계산하여 에지 수가 많은 순서로 슈퍼 노드

$C_1 \geq C_2 \geq \dots \geq C_n$ 를 정렬한다. 이 중 '중앙 슈퍼 노드'는 에지 수가 가장 많은 노드 C_1 으로 정하고 하위 계층의 중심에 레이아웃 한다. 나머지 슈퍼 노드 C_2, C_3, \dots, C_n 은 중앙 슈퍼 노드를 둘러싸는 내부 원 그룹(inner circular group)과 외부 원 그룹(outer circular group)으로 나누어 [그림 5]와 같이 계층적으로 레이아웃 한다.

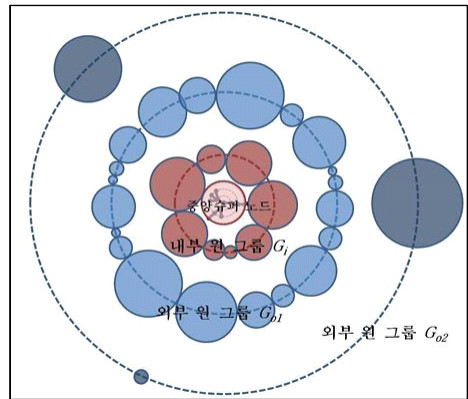


그림 5. 외부 원 그룹
Fig 5. group of outer circle

1) 내부 원 그룹

슈퍼 노드 C_2, C_3, \dots, C_n 중, 내부 원 그룹에 포함되는 슈퍼 노드는 식(4)를 만족하는 최소값 $k (< n)$ 의 존재 여부에 따라 결정된다.

$$R_{center} + r_{max} \leq rad_{inner} \cong \frac{\sum_{i=2}^{k+1} r_i}{\pi} \quad (4)$$

r_i : C_i 의 반지름, R_{center} : 중앙 슈퍼 노드의 반지름

식을 만족하는 최소값 k 가 존재하는 경우에는 k 개의 슈퍼 노드가 내부 원의 둘레에 서로 인접하여 레이아웃 되며, 내부

원의 반지름 R_{inner} 는 $\frac{\sum_{i=2}^{k+1} r_i}{\pi}$ 의 값이 된다.

최소값 k 가 존재하지 않는 경우는 모든 슈퍼 노드가 내부 원 그룹에 속하여 내부 원의 둘레에 레이아웃 될 수 있다. 따라서 내부 원의 반지름은 $R_{inner} = R_{center} + r_{max}$ 로 결정되며, 내부 원의 둘레에 레이아웃 되는 슈퍼 노드는 서로 인접하지 않고, 슈퍼 노드 간에 일정 간격을 유지하여 레이아웃 된다.

내부 원 그룹에 속하는 슈퍼 노드들의 위치는 상위 계층에

서 외부 슈퍼 노드로 지정된 환형 컴포넌트의 레이아웃 위치를 결정하는 방법과 동일하다.

2) 외부 원 그룹

내부 원의 반지름이 $R_{inner} = R_{center} + r_{max}$ 인 경우에는 모든 슈퍼 노드가 내부 원에 레이아웃 되므로, 외부 원 그룹의 슈퍼 노드는 내부 원 그룹의 최소값 k 가 $k < n - 1$ 인 경우에만 고려한다. 이 경우 전체 슈퍼 노드 중 k 개를 제외한 나머지 슈퍼 노드가 모두 레이아웃 될 때까지 필요한 외부 원을 단계적으로 반복 생성하여 해당 외부 원 그룹에 속하는 슈퍼 노드를 레이아웃 해야 한다. 하위 계층의 외부 원 그룹의 레이아웃 방법은 내부 원 그룹의 레이아웃 방법과 동일하게 수행될 수 있다.

이와 같은 과정을 모두 수행하여 모든 슈퍼 노드의 레이아웃 위치가 결정되면, 하위 계층에서의 최종 레이아웃 결과를 얻을 수 있다.

IV. 실험 결과

4.1 실험 및 결과

2-계층을 이용한 2.5차원 대사 경로 레이아웃 알고리즘을 적용하기 위해, 기존 실험과 마찬가지로 KEGG에 저장된 대사 경로 데이터를 사용하였다.

제안한 알고리즘의 수행 과정에 따라 입력으로 주어진 대사 경로 그래프 상에 환형 컴포넌트 또는 연결성 높은 노드가 존재하는 경우에만 2-계층 대사 경로 레이아웃 알고리즘이 적용되도록 하였다. 일반적으로 환형 컴포넌트 또는 연결성 높은 노드가 존재하지 않는 경우에는 대사 경로 그래프의 사이즈가 크지 않고 복잡한 구조를 갖지 않는다. 따라서 나머지 경우의 대사 경로 그래프에 대해서는 2차원 대사 경로 레이아웃 결과를 3차원 상에 그대로 보여주도록 하였다.

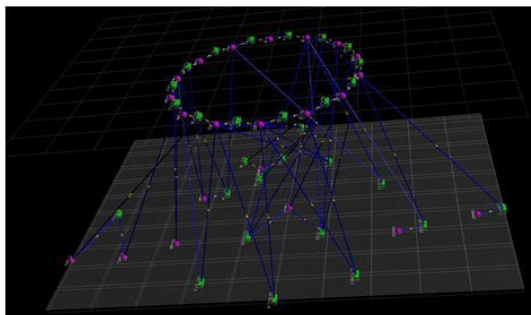


그림 6. 'Citrate cycle(TCA cycle)'에 대한 대사 경로 레이아웃 결과
Fig 6. metabolic pathway layout about 'Citrate cycle(TCA cycle)'

대사 경로 'Citrate cycle(TCA cycle)'는 본 논문에서 제안한 노드의 연결성을 고려한 2차원 대사 경로 레이아웃 알고리즘을 적용한 결과 에지 교차수를 13개로 줄였다. 이와 같은 레이아웃 결과는 [그림 6]과 같이 3차원 공간 상의 2-계층을 이용하여 좀 더 보기 좋게 바꿀 수 있다. 특히 각 계층의 중심을 일치시켜 레이아웃 결과가 어느 방향으로 치우치지 않게 대칭적으로 표현할 수 있다.

[그림 7]은 좀 더 복잡한 구조를 갖는 대사 경로 'Pyruvate metabolism'에 대한 레이아웃 결과이다. 'Pyruvate metabolism'은 노드수가 113개로 2차원 레이아웃 결과에서도 에지 교차수가 많이 존재하고, 넓은 영역에 걸쳐 레이아웃 되어 결과를 한눈에 보기 쉽지 않다. 그러나 제안한 레이아웃 알고리즘을 적용하여 에지 교차를 없애고, 레이아웃 영역도 줄어들면서 균형 있는 레이아웃 결과를 얻었다.

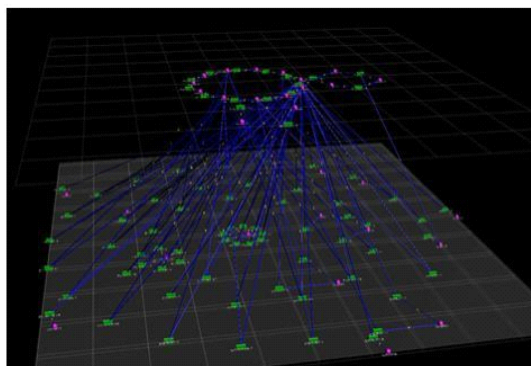


그림 7. 복잡한 구조를 갖는 'Pyruvate metabolism'에 대한 대사 경로 레이아웃 결과
Fig 7. metabolic pathway layout about 'Pyruvate metabolism'

V. 결 론

본 논문에서는 그래프 이론에 기반하여 대사 경로의 구조적인 특징을 분석하고 이를 반영한 대사 경로 레이아웃 알고리즘을 제안하였다. 3차원 상에 2개의 평행한 계층을 두고, 대사 경로의 구조적인 특징을 고려하여 입력으로 받은 대사 경로 그래프를 두 그룹으로 나누어 각각 다른 계층에 레이아웃 하도록 하였다. 각각의 계층에서는 새롭게 제안한 방식 레이아웃 알고리즘을 적용하여 각 컴포넌트를 레이아웃 하고, 2개 계층의 중심축을 일치시켜 각 계층의 컴포넌트를 중심을 기준으로 대칭적으로 표현함으로써 좀 더 균형 있는 결과를 얻었다. 또한 대사 경로를 계층적으로 표현함으로써 2차원 레이아웃 결과와 비교하여 예지 교차를 현저히 줄이면서 가독성을 높였다.

실험을 통해 2.5차원 레이아웃 모듈을 이용하여 대사 경로 그래프 상에 존재하는 주요 컴포넌트를 평행 이동시킨 결과 예지 교차수를 줄임과 동시에 전체 레이아웃 영역도 줄일 수 있었다. 뿐만 아니라 상위 계층과 하위 계층의 중심축을 일치시켜 각 계층의 컴포넌트를 중심을 기준으로 대칭적으로 표현함으로써 좀 더 균형 있고 보기에 좋은 레이아웃의 결과를 얻을 수 있었다. 특히 주요 컴포넌트인 환형 컴포넌트나 연결성 높은 노드를 제외한 나머지 연결 성분 간에는 예지가 존재하지 않고, 각 연결 성분은 기존의 방식 레이아웃 알고리즘을 적용하여 연결 성분 내에 예지 교차가 일어나지 않도록 하였다. 따라서 하위 계층에 레이아웃 결과에서는 예지 교차가 전혀 일어나지 않음을 확인하였다.

새롭게 제안한 대사 경로 시각화 레이아웃 알고리즘은 대사 경로의 구조적 특징을 고려하여 대사 경로를 단계적으로 세분화하고 3차원 공간상에 계층적으로 표현함으로써 대사 경로를 효과적으로 표현할 수 있다.

참고문헌

- [1] Deville, Y., D. Gilbert, J. Helden, and S. Wodak, "An Overview of Data Models for the Analysis of Biochemical Pathways", Proceedings of International Workshop on Computational Methods in System Biology, pp. 174, 2003.
- [2] BioCarta Team, Biocarta: Charting Pathways of Life, <http://www.biocarta.com>, 2001.
- [3] Christian Klukas and Falk Schreiber, "Dynamic Exploration and Editing of KEGG Pathway Diagrams, Bioinformatics", Vol. 23 No. 3, pp. 344-350, 2007.
- [4] Demir, E., O. Babur, U. Dogrusoz, A. Gursoy, G. Nisanci, R. Cetin-Atalay, and M. Ozturk, "PATIKA: An Integrated Visual Environment for Collaborative Construction and Analysis of Cellular Pathways, Bioinformatics", Vol. 18, No. 7, pp. 996-1003, 2002.
- [5] Holford, M., N. Li, P. Nadkarni, and H. Zhao, "VitaPad: Visualization Tools for the Analysis of Pathway Data", Bioinformatics, Vol. 21, No. 8, pp. 1596-1602, 2005.
- [6] Hu, Z., E. S. Snitkin, and C. DeLisi, "VisANT: an Integrative Framework for Networks in Systems Biology", Briefings in Bioinformatics, 2008.
- [7] Iragne, F., M. Nikolski, B. Mathieu, D. Auber, and D. Sherman, "ProViz: Protein Interaction Visualization and Exploration", Bioinformatics, Vol. 21, No. 2, pp. 272-274, 2005.
- [8] Rzhetsky, A., I. Iossifov, T. Koike, M. Krauthammer, P. Kra, M. Morris, H. Yu, P. A. Duboué, W. Weng, W. J. Wilbur, V. Hatzivassiloglou, and C. Friedman, "GeneWays: A System for Extracting, Analyzing, Visualizing, and Integrating Molecular Pathway Data", J. Biomed. Informat., Vol. 37, No. 1, pp. 43-53, 2004.
- [9] Saraiya, P., C. North, and K. Duca, "Visualizing Biological Pathways: Requirements Analysis, Systems Evaluation and Research Agenda", Information Visualization, Vol. 4, No. 3, pp. 191-205, 2005.
- [10] Wiese, K. C., and C. Eicher, "Graph Drawing Tools for Bioinformatics Research: An Overview", Proceedings of the 19th IEEE Symposium on Computer-Based Medical Systems, 2006.
- [11] Douglas B. West, Introduction to Graph Theory, Prentice Hall, pp.232-245, 1996.

- [12] Eades, P., "A Heuristic for Graph Drawing", Congressus Merantium, Vol. 42, pp. 142-160, 1984.
- [13] Friedrich, C. and F. Schreiber, "Flexible Layering in Hierarchical Drawings with Nodes of Arbitrary Size", Proceedings of the 27th Australasian Conference on Computer Science, Vol. 26, 2004.
- [14] Purchase, H.C., "Which Aesthetic has the Greatest Effect on Human Understanding?", Proceedings of the 5th International Symposium on Graph Drawing, pp. 249-261, 1997.
- [15] Karp, P. D., and S. Paley, "Automated Drawing of Metabolic Pathways", Third International Conference on Bioinformatics and Genome Research, pp. 225-238, 1994.
- [16] Becker, M. Y., and I. Rojas, "A Graph Layout Algorithm for Drawing Metabolic Pathways", Bioinformatics, Vol. 17, No. 5, pp. 461-467, 2001.
- [17] Weijiang, L., and H. Kurata, "A Grid Layout Algorithm for Automatic Drawing of Biochemical Networks", Bioinformatics, Vol. 21, No. 9, pp. 2036-2042, 2005.

저 자 소 개



송 은 하
 1996: 성심대학교 수학과 이학사.
 1999: 이화여자대학교
 컴퓨터공학과 공학석사.
 2002: 한국과학기술연구원 연구원
 2011: 이화여자대학교
 컴퓨터공학과 공학박사
 현재: 이즈텍(ISTech)
 관심분야: 컴퓨터공학, 알고리즘,
 그래프이론
 Email : Ehsong@ewhain.net



용 승 림
 1998: 이화여자대학교
 전자계산학과 공학사.
 2000: 이화여자대학교
 컴퓨터공학과 공학석사.
 2006: 이화여자대학교
 컴퓨터공학과 공학박사
 현 재: 인하공업전문대학
 컴퓨터시스템과 교수
 관심분야: 컴퓨터공학, 알고리즘,
 정보보안
 Email : slyong@inhac.ac.kr