

3 계층의 2.5차원 대사경로 레이아웃 알고리즘

송은하*, 용승림**

3-layer 2.5D Metabolic pathway layout algorithm

Eun-Ha Song*, Seunglim Yong**

요약

화합물의 상호 관계를 그래프를 통해 표현하는 대사 경로는 본질적인 복잡성 때문에 대사 경로 내의 흐름을 한 눈에 알 수 있도록 가시화하여 보여 주는 도구가 반드시 필요하다. 또한 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 3-계층을 이용한 대사 경로 레이아웃 알고리즘을 제안한다. 대사 경로의 구조적 특징을 고려하여 노드수가 증가하여도 에지 교차가 기하급수적으로 증가하는 문제를 해결하기 위하여 연결성 높은 노드와 환형 컴포넌트를 중앙계층에 위치시키고 나머지 부분 그래프를 상위와 하위 계층에 레이아웃 하도록 한다. 실험을 통해 에지 교차수가 줄어들음을 확인할 수 있다.

▶ Keywords : 대사경로, 시각화, 레이아웃 알고리즘

Abstract

Metabolic pathway, represented by well-displayed graph, have a complex binding structure, which makes the graphical representation hard to visualize. 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 3-layer layout. Our algorithm searches any meaningful component existing in a pathway, such as circular components, highly connected nodes, and the components are drawn in middle layer. Then the remaining subgraphs except meaningful components are drawn in upper and lower layer by utilizing a new radial layout algorithm. It reduces ultimately reduced the number of edge crossings. Our algorithm solve the problem that edge crossings exponentially increase as the number of nodes grows.

▶ Keywords : Metabolic pathway, visualization, layout algorithm

•제1저자 : 송은하 •교신저자 : 용승림

•투고일 : 2013. 3. 19, 심사일 : 2013. 4. 12, 게재확정일 : 2013. 4. 30.

* 이즈텍(IStech)

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

I. 서론

대사 경로에 대한 정보는 생물체 내에서 나타나는 네트워크 중 가장 기본적인 정보로, 연구를 수행하는 생물학자들은 KEGG, EcoCyc, WIT 등의 데이터베이스 내에 저장된 대사 경로 정보를 검색하여 필요한 정보를 얻어낸다. 그러나 대사 경로는 본질적인 복잡성 때문에 대사 경로의 흐름을 한 눈에 이해하기 어려워 경로 정보를 시각적인 정보로 표현하기 위한 방법이 요구된다.

대사 경로 시각화 시스템은 생물학자들에게 복잡한 대사 경로의 구조를 한 눈에 쉽게 파악할 수 있도록 하여 실제 연구를 수행하는 데 필수적인 요소가 되고 있으며 이를 위한 많은 연구가 진행되고 있다[1, 2, 3, 4, 5, 6, 7].

본 논문에서는 대사 경로의 구조적인 특징을 파악하고 이를 반영한 3계층의 대사 경로 레이아웃 알고리즘을 설계하고 구현한다. 레이아웃 알고리즘은 단순히 그래프 이론에 기반하여 대사 경로의 구조적인 특징을 분석하고 이를 반영하여 3개의 계층에 적절히 레이아웃 한다. 3차원 상에 평행한 3개의 계층(layer)을 두어 대사 경로 그래프를 계층화함으로써 대사 경로의 구조 분석을 용이하게 하고 에지 교차를 줄여 대사 경로의 가독성을 높이는 결과를 얻는다.

II. 기존 연구

대사 경로는 본질적인 복잡성 때문에 대사 경로 내의 흐름을 한 눈에 쉽게 이해하고 파악할 수 있는 대사 경로 가시화 도구가 반드시 필요하다. 2차원 상에서의 레이아웃 알고리즘과 함께 3차원 상에서 복잡한 대사 경로를 표현하기 위한 알고리즘들이 연구되고 있다.

2.1 K. Wegner와 U. Kummer의 알고리즘

K. Wegner와 U. Kummer가 제안한 대사 경로 레이아웃 알고리즘은 대사 경로 그래프 상에서 가장 큰 환형 컴포넌트 대신 서로 노드를 공유하는 작은 사이즈의 환형 컴포넌트를 찾아서 레이아웃 해줌으로써 전체적인 레이아웃 결과를 얻는다[8].

주어진 대사 경로 그래프에 대해 환형 컴포넌트가 존재하지 않는 경우는 전체 대사 경로 그래프에 계층적 레이아웃 알고리즘을 적용하여 레이아웃 결과를 얻는다. 반면 환형 컴포넌트가 존재하는 경우는 전체 대사 경로 그래프가 하나의 환

형 컴포넌트로 이루어진 경우와 그렇지 않은 경우 즉, 환형 컴포넌트를 포함하는 복잡한 구조의 경우로 나뉜다. 대사 경로 그래프가 하나의 환형 컴포넌트로 이루어진 경우에는 전체 그래프에 환형 레이아웃 알고리즘을 적용한다. 이 알고리즘은 대사 경로 그래프에 존재하는 작은 사이즈의 사이클 특히 노드를 공유하는 사이클을 찾아서 함께 레이아웃 해 줌으로써 전체적인 레이아웃 결과에 나타나는 에지 교차를 줄였고, 생물학자에게 익숙한 레이아웃 결과 얻을 수 있다.

2.2 Y. Yang이 제안한 3차원 레이아웃 시스템

Y. Yang은 3차원 공간 상에서 그래프 레이아웃, 컴퓨터 그래픽스, 가상현실을 결합하여 상호작용이 가능한 대사 경로 네트워크 시각화 시스템을 개발하였다[9]. 대사 경로 네트워크는 여러 개의 대사 경로로 구성되어 있고 그 노드의 수가 수천 개에 이른다. 따라서 전체 대사 경로 네트워크를 레이아웃 하기 보다는 부분적으로 레이아웃 하여 그 결과를 보여주도록 하였다. 사용자는 대사 경로 네트워크에 포함되는 모든 반응 중에서 관심 있는 반응(ROI, Reaction of Interest)을 찾고 반응의 중심이 되는 노드(focus node)를 선택할 수 있다. 중심 노드가 선택되면 중심 노드를 기반으로 연관된 반응들이 3차원 공간 상에 표현된다. 이와 같은 시스템은 가상 환경이 제공되는 태블릿 PC를 이용하여 사용자 인터페이스를 통해 수행된다.

2.3 I. Rojdestvenski의 3차원 레이아웃 시스템

I. Rojdestvenski는 VRML을 이용하여 가상현실 속에 대사 경로 네트워크를 시각화해 주는 시스템을 제안하였다[10, 11]. 제안한 시스템은 XML-compliant MNV (Metabolic Network Visualizer) 언어를 기반으로 개발되었으며, MNV 언어 표준 및 파서, MNV를 VRML으로 변환하는 번역기(translator), HTML기반 사용자 그래픽 인터페이스를 이용한 경로 생성기(interactive pathway constructor) 그리고 기존 데이터베이스에 저장된 경로 데이터 변환기(converter) 등 4가지 모듈로 구성되어 있다. 이러한 주요 모듈을 통해 대사 경로 그래프를 3차원 상의 XY평면 또는 3차원 공간에 레이아웃 하며 주어진 대사 경로 네트워크를 레이아웃 하기 위해 기존 스프링 임베딩 알고리즘을 수정하여 대사 경로 네트워크를 표현하였다.

이 시스템은 대사 경로를 가상현실이라는 환경에 표현함으로써 대사 경로를 다양한 각도에서 관찰할 수 있고 다양하게 나타낼 수 있다는 3차원 공간의 장점을 최대한 이 용해 대사 경로를 표현한다. 하지만 대사 경로 그래프에서 보조 화합물

도 하나의 노드로 표현하므로 여러 반응(reaction)에 참여하는 보조 화합물의 경우 연결성이 높은 노드가 되어 전체적으로 에지 교차수를 증가시키는 결과를 초래한다.

III. 3계층의 대사경로 레이아웃 알고리즘

3-계층을 이용하여 대사 경로 그래프를 레이아웃 하기 위해서 대사경로 그래프를 각각 상위 계층, 중앙 계층, 하위 계층의 세 그룹으로 나누어 2.5차원 공간상에 계층적으로 레이아웃 한다. 제안하는 각 계층의 레이아웃 알고리즘은 다음과 같다.

3.1 계층 분할 알고리즘

대사 경로 그래프의 계층을 구분하기 위하여 먼저, 입력으로 들어온 대사 경로 그래프에 대해 환형 컴포넌트의 존재 여부를 확인한다. 환형 컴포넌트가 존재하는 경우, [그림 1]과 같이 대사 경로 그래프상에 존재하는 환형 컴포넌트를 모두 찾고, 환형 컴포넌트를 제외한 나머지 부분을 연결 성분으로 나눈다. 환형 컴포넌트가 존재하지 않는 경우에는 대사 경로 그래프에 존재하는 차수가 6이상인 모든 연결성 높은 노드를 찾고, 연결성 높은 노드들을 제외한 나머지 부분을 연결 성분으로 나눈다. 환형 컴포넌트 또는 차수가 6 이상인 연결성 높은 노드는 중앙 계층에 레이아웃하게 되고, 연결성분은 상위 또는 하위 계층에 레이아웃하게 된다.

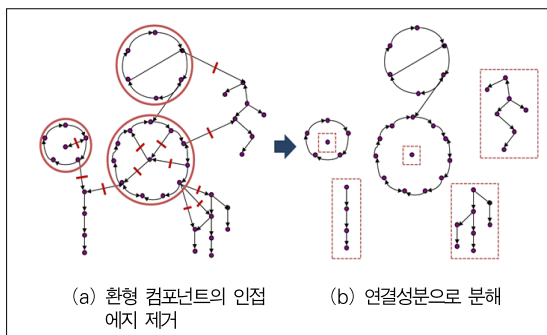


그림 1. 환형 컴포넌트와 연결성분
Fig 1. Circular component and connected component

연결 성분은 중앙 계층에 레이아웃 되는 컴포넌트 사이의 에지 수를 고려하여 상위, 하위의 두 계층 그룹으로 나눈다. [그림 2]와 같이 연결 성분과 에지 수가 큰 연결 성분부터 하나씩 그룹에 추가하여 두 그룹의 에지 수의 합이 최대한 같아지도록 나눈다. 이는 계층과 계층 간에 존재하는 에지의

수를 되도록 균형 있게 분배하여 전체 레이아웃에 있어서 덜 복잡하고 보기에 좋은 결과를 얻기 위함이다.

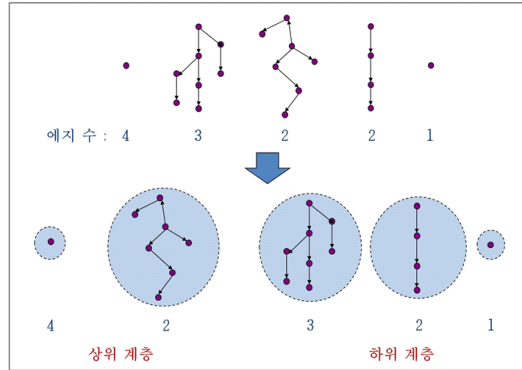


그림 2. 연결성분의 그룹핑
Fig 2. grouping of connected components

계층 분할을 위한 알고리즘은 알고리즘1과 같다.

```

LayoutPathway(Pathway graph)
{
  IF any cycle exists in graph
  FIND all cycles ( $N_c$ ) in graph
  decompose graph -  $N_c$  into connected components  $Comps$ 
  group  $Comps$  into  $Upper\_Comps, Lower\_Comps$ 
  LayoutCenter ( $N_c$ )
  LayoutUpper ( $Upper\_Comps$ )
  LayoutLower ( $Lower\_Comps$ )

  ELSE IF highly connected node exists in graph
  FIND all highly connected nodes ( $N_d$ ) in graph
  decompose graph -  $N_d$  into connected components  $Comps$ 
  group  $Comps$  into  $Upper\_Comps, Lower\_Comps$ 
  LayoutCenter ( $N_d$ );
  LayoutUpper ( $Upper\_Comps$ )
  LayoutLower ( $Lower\_Comps$ )
}
    
```

알고리즘 1. 계층 분할 알고리즘
algorithm 1. layer partition algorithm

3.2 중앙 계층 레이아웃 알고리즘

입력으로 주어진 대사 경로 그래프에 존재하는 주요 컴포넌트인 환형 컴포넌트, 연결성 높은 노드들은 중앙계층에 레이아웃 된다.

1) 슈퍼 노드의 정의

주어진 대사 경로 그래프는 알고리즘 2와 같이 중앙 슈퍼노드와 외부 슈퍼노드를 정의하게 된다.

환형 컴포넌트가 존재하는 경우는 찾아낸 환형 컴포넌트에는 '환형 레이아웃 알고리즘'을 적용하여 레이아웃 한 후, 각

각을 하나의 슈퍼 노드로 간주한다[12, 13]. 슈퍼 노드들 중 환형 컴포넌트 사이에 존재하는 예지가 가장 많은 슈퍼 노드는 '중앙 슈퍼 노드(C_{center})'로, 중앙 슈퍼 노드를 제외한 나머지 슈퍼 노드는 '외부 슈퍼 노드(C_i)'로 정한다.

연결성 높은 노드가 존재하는 경우는 먼저 연결성 높은 노드 사이에 가장 예지수가 많은 노드를 '중앙 슈퍼 노드(C_{center})'로 지정하고, 나머지 연결성 높은 노드를 '외부 슈퍼 노드(C_i)'로 정한다.

```

LayoutCenter (components N)
{
  IF component N is  $N_c$ 
    CircularLayout(each circular component of  $N_c$ )
    set each circular component of  $N_c$  to super nodes
  ELSE IF component N is  $N_d$ 
    set HC node of  $N_d$  to super nodes

  count edges among super nodes
  set super node with the highest degree to  $C_{center}$ 
  set other super nodes to  $C_{outer} = \{C_i : i=1, \dots, n\}$ 
  ModifiedRadial1 ( $C_{center}, C_{outer}$ );
}
    
```

알고리즘2. 슈퍼노드 정의 알고리즘
algorithm2. supernode definition algorithm

2) 슈퍼 노드의 레이아웃

중앙 슈퍼 노드는 중앙 계층의 중심에 레이아웃 한다. 외부 슈퍼노드들은 중앙 슈퍼노드와 겹치지 않고 레이아웃하기 위해서 외부원의 반지름(R_{outer})을 계산하고 반지름을 원으로 하는 원 위에 위치하도록 한다. 외부원의 반지름은 중앙 슈퍼 노드의 반지름(R_{center})과 외부 슈퍼 노드의 반지름(r_i)으로부터 계산된다. 알고리즘 3은 중앙 슈퍼노드를 레이아웃하고 외부 슈퍼노드를 레이아웃하기 위한 외부 원의 크기를 계산하는 알고리즘이다.

```

ModifiedRadial1 ( $C_{center}, C_{outer}$ )
{
  layout  $C_{center}$  to Center on CL
   $R_{center} :=$  radius of  $C_{center}$ 
   $r_i :=$  radius of  $C_i$ 
   $r_{max} :=$  maximum among  $r_i$ 
   $n :=$  # of  $C_i$ 

  calculate  $R_{outer} := \max \{ R_{center} + r_{max}, \frac{\sum_{i=1}^n r_i}{\pi} \}$ 

  FOR EACH  $C_i$ 
    calculate center of mess of  $C_i$ 
    sort  $C_i$  by degree of node
  IF  $n = 1$ 
    FIND ( $x_{final}, y_{final}$ ) near to center of mess of  $C_{outer}$ 
  ELSE
     $C'_{outer} =$  Find_init_position ( $C_{outer}$ )
    Final_node_Layout ( $C'_{outer}$ )
}
    
```

알고리즘 3. 변형된 방사 레이아웃 알고리즘
algorithm3. modified radial layout algorithm

외부 원의 크기가 결정되고 나면 각 외부 슈퍼 노드의 무게 중심(center of mess)을 구하여 외부원상의 위치를 결정한다. 외부 슈퍼 노드의 무게 중심은 중앙 슈퍼 노드와 해당 외부 슈퍼 노드와 연결되어 있는 노드들의 평균 좌표값으로부터 구한다. 계산된 외부 슈퍼 노드의 무게중심에 따라 외부 슈퍼 노드의 위치가 결정된다.

외부 슈퍼 노드의 초기 위치는 외부 원의 둘레를 외부 슈퍼 노드의 개수 n 개로 나눈 위치에 해당되는 점으로 한다.

```

Find_init_position ( $C_{node}$ )
{
  init_position[i] : initial position of super node
   $n :=$  # of  $C_i$ 

  FOR  $i = 1, \dots, n$ 
    init_position[i].theta :=  $\frac{360}{n} \times i$ 
    init_position[i].set := false
  FOR EACH  $C_i$ 
    FOR  $i = 1 \dots n$ 
      IF init_position[i].set = false
        calculate distance init_position[i] & center of mess of  $C_i$ 
        IF min_distance > distance
          min_distance := distance
          init_position[i].set := true
    sort  $C_{node}$  by init_position
}
    
```

알고리즘 4. 슈퍼 노드의 초기 위치 설정 알고리즘
algorithm 4. initial position setting algorithm

초기 위치가 정해지면 외부 슈퍼노드들은 보기 좋게 x 축

을 기준으로 반시계방향으로 위치를 조정한다. 외부원의 반지

름이 $R_{outer} = \frac{\sum_{i=1}^n r_i}{\pi}$ 인 경우에는 외부 슈퍼 노드가 겹쳐서 레이아웃 되지 않도록 위치 각도(positioning angle)를 계산하여 위치를 조정하여 레이아웃하고, $R_{outer} = R_{center} + r_{max}$ 인 경우에는 외부 슈퍼 노드 사이에 간격이 일정하도록 위치를 조정하여 레이아웃 한다. 조절할 위치 각도를 계산하고 레이아웃 하는 알고리즘은 알고리즘 5 와 같다.

```

Final_node_Layout ( $C'_{node}$ )
{
  angle[i] : positioning angle of  $C_i$ 
   $C'_{node\_angle}[i]$  : angle of  $C'_{node}$ 
  angle[1]=0

  FOR  $i = 2$  to  $n$ 
     $C'_{node\_angle}[i] = \sin^{-1}(r'_i / R_{outer})$ 
    angle[i]=angle[i-1]+ $C'_{node\_angle}[i]$ + $C'_{node\_angle}[i-1]$ 
  FOR  $i = 2$  to  $n$ 
    IF  $R_{outer} = R_{center} + r_{max}$ 
      alpha =  $2(\pi - \sum_{j=1}^n angle[j]) / n$ 
      angle[i]=angle[i]+ alpha
    FOR  $i = 1$  to  $n$ 
      layout  $C_i \in C'_{node}$  with angle[i]
}
    
```

알고리즘 5. 슈퍼노드의 레이아웃 위치 결정 알고리즘
algorithm 5. final layout algorithm

알고리즘 5에서 r'_i 의 값은 [그림 3]와 같이 중앙 슈퍼 노드의 중심으로부터 외부 슈퍼 노드 C'_i 에 그른 점선과 원의 중심과의 거리이다.

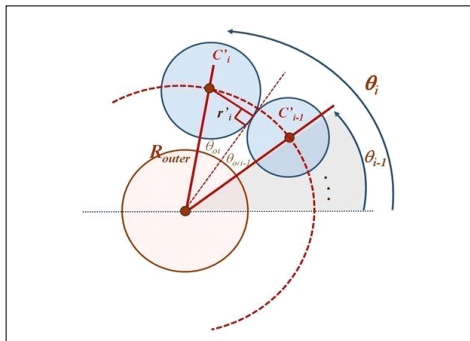


그림 3. 외부 슈퍼 노드의 위치 각도 조정
Fig 3. calculate positioning angle of outer supernode

3.3. 상위 · 하위 계층 레이아웃 알고리즘

대사경로 그래프에 존재하는 주요 컴포넌트를 제외한 나머지 연결 성분들은 상위 계층과 하위 계층의 두 계층에 나누어 레이아웃 한다.

```

LayoutUpper&Lower(Components Comps)
{
  FOR EACH Comp  $i \in$  Comps
    IF Comp  $i$  is a tree && |Comp  $i$ |>2
      IF Comp  $i$  is a linear && |Comp  $i$ |>3
        apply 'circular layout algorithm' to Comp  $i$ 
      ELSE
        RadialLayout (Comp  $i$ )
    set each component to circular super nodes  $C_i$ 
    ModifiedRadial2 ( $C_i$ )
}
    
```

알고리즘 6. 상위, 하위 계층의 레이아웃 알고리즘
algorithm 6. upper and lower layout algorithm

나머지 연결성분들은 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.

연결 성분은 중앙 계층에 레이아웃 되는 컴포넌트 사이와 의 에지 수가 큰 연결 성분부터 하나씩 상위 계층의 그룹과 하위 계층의 그룹에 추가하여 두 그룹의 에지 수의 합이 최대한 같아지도록 나눈다. 상위와 하위 계층으로 나뉜 슈퍼노드들은 각각 중앙 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 상위와 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 계층적으로 레이아웃 한다.

```

ModifiedRadial2 ( $C_s$ )
{
  count edges between  $C_i$  and center layer CL
  sort  $C_s$  by the number of edges between  $C_i$  and CL

  FOR component
  set  $C_i$  with the highest degree to  $C_{center}$ 
  layout  $C_{center}$  to Center on CL

  FIND minimum  $k$  such that  $R_{center} + r_{max} \leq rad_{inner} \cong \frac{\sum_{i=2}^{k+1} r_i}{\pi}$ 

  IF  $k < n$  &&  $k$  exists
  set  $R_{inner} \cong \frac{\sum_{i=2}^{k+1} r_i}{\pi}$ 

   $C'_{innerS} = \text{Find\_init\_position}(C_{innerS})$ 
  Final_node_Layout( $C_2, C_3, \dots, C_{k+1}$ )
  IF  $k < n-1$ 
  LayoutOuter( $C_{k+2}, C_{k+3}, \dots, C_n, k+2$ )
  ELSE
  set  $R_{inner} = R_{center} + r_{max}$ 
   $C'_{innerS} = \text{Find\_init\_position}(C_{innerS})$ 
  Final_node_Layout ( $C_2, C_3, \dots, C_n$ )
}
    
```

알고리즘 7. 새로운 방사 레이아웃 알고리즘
algorithm 7. new radial layout algorithm

1) 내부 원 그룹

슈퍼 노드 C_2, C_3, \dots, C_n 중, 알고리즘 7에서와 같이

$R_{center} + r_{max} \leq rad_{inner} \cong \frac{\sum_{i=2}^{k+1} r_i}{\pi}$ 의 식을 만족하는 최소값

$k(k < n)$ 의 존재 여부에 따라 내부 원의 반지름 R_{inner} 는 $\frac{\sum_{i=2}^{k+1} r_i}{\pi}$ 의 값 또는 $R_{center} + r_{max}$ 로 정해진다. 내부 원 그룹에 속하는 슈퍼 노드들의 위치는 중앙 계층에서 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.

```

LayoutOuter( $C_{outerS}$ , int start_index)
{
   $r_{outer\_max}$ : maximum radius among  $C_{outerS}$ 

  WHILE start_index  $\leq n$ 
  FIND minimum  $l$  such that  $R_{inner} + r_{outer\_max} \leq rad_{outer} \cong \frac{\sum_{i=start\_index}^l r_i}{\pi}$ 
  IF  $l \leq n$  &&  $k$  exists
  set  $R_{outer} \cong \frac{\sum_{i=start\_index}^l r_i}{\pi}$ 
  Final_node_Layout( $C_{start\_index}, C_3, \dots, C_l$ )
  start_index =  $l+1$ 
   $R_{inner} = R_{outer}$ 
  ELSE
  set  $R_{outer} = R_{inner} + r_{outer\_max}$ 
  Final_node_Layout ( $C_{start\_index}, C_3, \dots, C_n$ )
}
    
```

알고리즘 8. 외부 원 그룹의 레이아웃 알고리즘
algorithm 8. outer circular group layout algorithm

IV. 실험 결과

본 절에서는 KEGG에 저장된 80여개의 대사 경로 데이터에 대해 3-계층을 이용한 2.5차원 대사 경로 레이아웃 알고리즘을 적용하여 그 레이아웃 결과를 얻고, 각 계층에 나타나는 에지 교차수를 분석하였다.

[그림 4]는 보조 인자 및 비타민의 대사(Metabolism of Cofactors and Vitamins) 카테고리에 속하는 대사 경로 'One carbon pool by folate'에 대해 3계층 레이아웃 알고리즘을 적용한 결과이다. 대사 경로 'One carbon pool by folate'는 40개의 노드 수를 포함하고 있으며, 3계층 레이아웃 알고리즘을 적용하여 레이아웃 한 결과 각 계층에 에지 교차가 전혀 일어나지 않았으며 각 계층에서 컴포넌트들이 중심축을 기준으로 방사 형태로 잘 레이아웃 되었음을 알 수 있다.

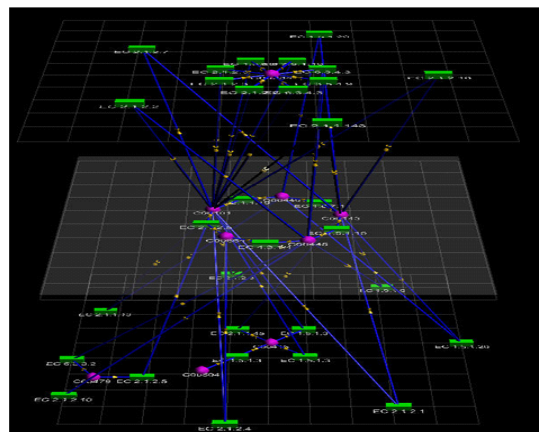


그림 4. 대사 경로 'One carbon pool by folate' 3-계층 레이아웃 결과
Fig 4. metabolic pathway layout about 'One carbon pool by folate'

[그림 5]는 사이즈가 크고 좀 더 복잡한 구조를 갖는 대사 경로 'fatty acid metabolism'에 대한 3-계층 레이아웃 결과이다. 대사 경로 'fatty acid metabolism'와 같이 노드수를 많이 포함하는 대사 경로의 경우 기존의 알고리즘을 이용하여 대사 경로를 레이아웃 하면 노드가 한쪽 계층으로 치우치면서 좀 더 복잡하게 표현되는 경향이 있다. 3-계층을 이용해서 레이아웃 한 결과 137개 노드에 대한 계층 간 평균 에지 길이는 [표 1]에서와 같이 2 계층으로 레이아웃 할 때보다 계층 간 깊이의 1.71배에서 1.3배로 짧아졌다.

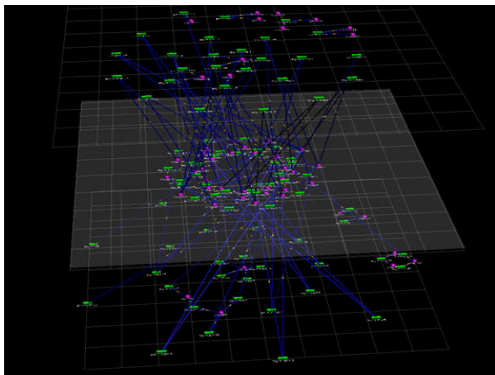


그림 5. 대사 경로 'fatty acid metabolism' 3-계층 레이아웃 결과
Fig 5. metabolic pathway layout about 'fatty acid metabolism'

[표 1]은 계층과 계층 간 평균 에지 길이를 정리한 것이다. 계층 간 평균 에지 길이는 계층 간 깊이의 배수로 표현하였다. 대부분의 대사 경로 그래프에 대해 3-계층을 이용한 레이아웃 알고리즘을 적용한 경우 2-계층을 이용한 레이아웃 알고리즘 적용 결과에 비해 계층 간 평균 에지 길이가 짧아짐을 알 수 있다. 특히, 노드의 사이즈가 100이상인 대사 경로 그래프에 대해서는 계층 간 평균 에지 길이가 확연히 줄어들어 레이아웃의 복잡함을 감소시켰다.

표 1. 계층 간 평균 에지 길이
table 1. average edge length between layers

대사 경로 그래프		계층 간 평균 에지 길이	
		2-계층 레이아웃	3-계층 레이아웃
Carbohydrate Metabolism	Glycolysis_Gluconeogenesis	1.83	1.76
	Citrate cycle (TCA cycle)	1.61	1.28
	Pentose phosphate pathway	1.59	1.55
	Pentose and gluconate interconversions	1.52	1.42
	Fructose and mannose metabolism	1.68	1.36
	Galactose metabolism	1.74	1.7
	Starch and sucrose metabolism	2.34	1.61
	Aminosugars metabolism	1.48	1.44
	Nucleotide sugars metabolism	1.72	1.65
	Pyruvate metabolism	1.73	1.42
	Glyoxylate and dicarboxylate metabolism	1.73	1.55
	Propanoate metabolism	1.83	1.69
	Butanoate metabolism	1.72	1.65
	C5-Branched dibasic acid metabolism	1.25	1.25
	Inositol phosphate metabolism	1.63	1.51

Lipid Metabolism	alpha-Linolenic acid metabolism	4.2	4.09
	Androgen and estrogen metabolism	1.7	1.52
	Bile acid biosynthesis	1.74	1.78
	C21-Steroid hormone metabolism	1.57	1.41
	Ether lipid metabolism	1.43	1.49
	Fatty acid biosynthesis	2.27	1.41
	Fatty acid elongation in mitochondria	1.29	1.46
	Fatty acid metabolism	1.71	1.3
	Glycerolipid metabolism	1.49	1.55
	Glycerophospholipid metabolism	1.67	1.3
Nucleotide Metabolism	Synthesis and degradation of ketone bodies	1.21	1.1
	Purine metabolism	2.92	2.11
Amino Acid Metabolism	Pyrimidine metabolism	2.1	1.59
	Alanine and aspartate metabolism	1.88	1.85
	Arginine and proline metabolism	1.67	1.32
	Cysteine metabolism	1.55	1.45
	Glutamate metabolism	1.7	1.48
	Glycine, serine and threonine metabolism	1.77	1.47
	Histidine metabolism	3.83	3.2
	Lysine degradation	1.51	1.36
	Methionine metabolism	1.73	1.68
	Phenylalanine metabolism	1.79	1.69
	Phenylalanine, tyrosine and tryptophan biosynthesis	1.64	1.56
	Tryptophan metabolism	3.03	2.05
	Tyrosine metabolism	2.18	1.71
	Urea cycle and metabolism of amino groups	1.46	1.33
	Valine, leucine and isoleucine biosynthesis	1.43	1.38
	Valine, leucine and isoleucine degradation	1.62	1.48

V. 결론

본 논문은 대사 경로 그래프의 표현 영역인 3차원 공간의 장점을 최대한 활용할 수 있는 '3-계층을 이용한 2.5차원 대사 경로 레이아웃 알고리즘'을 제안하였다. 3차원 공간상에 상위 계층, 중앙 계층, 하위 계층 3개의 평행한 계층을 두고, 대사 경로 그래프를 세 그룹으로 나누어 레이아웃 하였다. 먼저, 대사 경로의 구조적인 특징을 고려하여 대사 경로에 포함되는 환형 컴포넌트 또는 연결성 높은 노드를 찾아 중앙 계층에 레이아웃 한다. 다음으로 주요 컴포넌트를 제외한 나머지 부분을 연결 성분으로 나누고, 연결 성분을 두 그룹으로 나누어 각각 상위 계층과 하위 계층에 레이아웃 하였다. 이 때, 상위 계층과 중앙 계층 그리고 중앙 계층과 하위 계층 사이에 존재하는 에지 수를 균형 있게 분배하기 위해서 연결 성분들과 중앙 계층 간의 에지 수를 고려하여 연결 성분을 두 그룹으로 나누었다. 두 그룹으로 나누어진 연결 성분은 각각 상위 계층과 하위 계층에 새로 제안한 방식 레이아웃 알고리즘을 적용하여 중심을 둘러싸는 내부 원 및 외부 원 둘레에 레이아웃 하였다.

대사 경로 그래프에 대해 계층을 확장한 3-계층을 이용한 레이아웃 알고리즘을 적용한 결과, 각각의 계층 내에서의 레이아웃이 덜 복잡함을 알 수 있었다. 특히 사이즈가 큰 대사 경로 그래프의 경우 그래프를 세 그룹으로 나누어 레이아웃함으로써 그래프의 노드가 한 계층에 많이 치우쳐서 레이아웃 되는 문제를 해결할 수 있다. 또한 에지 교차수도 줄이고, 각

계층에 레이아웃 되는 영역도 줄임으로써 가독성을 높일 수 있어 여러 개의 대사 경로가 포함되는 대사 경로 네트워크를 레이아웃 하는 방법으로 사용하기에 적합하다.

향후 이와 같은 대사 경로 레이아웃 알고리즘을 통해 많은 양의 대사 경로정보 데이터들을 쉽게 이해하고 파악할 수 있도록 유연성 있고 유지관리가 용이한 통합시스템을 구축하고 활용하는 연구가 필요하다.

참고문헌

- [1] Christian Klukas and Falk Schreiber, "Dynamic Exploration and Editing of KEGG Pathway Diagrams", *Bioinformatics*, Vol. 23 No. 3, pp. 344-350, 2007.
- [2] 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.
- [3] 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.
- [4] Hu, Z., E. S. Snitkin, and C. DeLisi, "VisANT: an Integrative Framework for Networks in Systems Biology", *Briefings in Bioinformatics*, 2008.
- [5] 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.
- [6] 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.
- [7] 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.
- [8] Wegner, K., and U. Kummer, "A New Dynamical Layout Algorithm for Complex Biochemical Reaction Networks", *BMC Bioinformatics*, Vol. 6, 2005.
- [9] Yang, Y., E. S. Wurtele, C. Cruz-Neira, and J. A. Dickerson, "Hierarchical Visualization of Metabolic Networks using Virtual Reality", *Proceedings of the 2006 ACM international conference on Virtual reality continuum and its applications*, pp. 377-381, 2006.
- [10] Rojdestvenski, I., "Metabolic Pathways in Three Dimensions", *Bioinformatics*, Vol. 19, No. 18, pp. 2436-2441, 2003.
- [11] Rojdestvenski, I., and M. Cottam, "Visualizing Metabolic Networks in VRML", *Proceedings of the Sixth International Conference on Information Visualisation*, 2002.
- [12] Becker, M. Y., and I. Rojas, "A Graph Layout Algorithm for Drawing Metabolic Pathways", *Bioinformatics*, Vol. 17, No. 5, pp. 461-467, 2001.
- [13] Karp, P. D., and S. Paley, "Automated Drawing of Metabolic Pathways", *Third International Conference on Bioinformatics and Genome Research*, pp. 225-238, 1994.

저 자 소 개



송 은 하
1996: 성심대학교 수학과 이학사.
1999: 이화여자대학교
컴퓨터공학과 공학석사.
2002: 한국과학기술연구원 연구원
2011: 이화여자대학교
컴퓨터공학과 공학박사.
현재: 이스텍(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