

2-계층 레이아웃을 이용한 2.5차원 대사 경로 드로잉

(2.5D Metabolic Pathway Drawing based on 2-layered Layout)

송은하[†] 함성일^{**} 이상호^{***} 박현석^{****}
(Eun-Ha Song) (Sung-il Ham) (Sang-Ho Lee) (Hyun-Seok Park)

요약 대사체학은 대사 경로 네트워크를 통해 생명 활동을 이해하고자 하는 분야로서, 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 논문에서는 대사 경로의 구조적인 특징에 기반한 3차원 공간 상에 대사 경로 그래프를 레이아웃해 주는 모듈을 설계, 구현하였다. 2-계층 레이아웃을 이용하여 대사 경로 그래프를 계층적으로 레이아웃함으로써 표현영역을 3차원으로 확장시키고 기존의 2차원 레이아웃 알고리즘 적용시 빈번히 나타나는 에지 교차의 수를 감소시키는 결과를 얻었다.

키워드 : 생물정보학, 대사 경로, 그래프 드로잉, 에지 교차, 가독성, 2-계층 레이아웃

Abstract Metabolomics interprets an organism as a network of functional units and an organism is represented by a metabolic pathway i.e., well-displayed graph. So a software tool for drawing pathway is necessary to understand it comprehensively. These tools have 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, it is very important to reduce unnecessary edge-crossing on a metabolic pathway layout.

In this paper, we design and implement 2.5D metabolic pathway layout modules. Metabolic pathways are represented hierarchically by making use of the '2-layered layout algorithm' in 3D. It enhances the readability and reduces unnecessary edge-crossings by using 3D layout modules instead of 2D layout algorithms.

Key words : bioinformatics, metabolic pathway, graph drawing, edge-crossing, readability, 2-layered-layout

† 학생회원 : 이화여자대학교 컴퓨터공학과
ehsong@ewhain.net
** 학생회원 : 연세대학교 건축공학과
archispace@yonsei.ac.kr
*** 종신회원 : 이화여자대학교 컴퓨터공학과 교수
shlee@ewha.ac.kr
**** 정회원 : 이화여자대학교 컴퓨터공학과 교수
neo@ewha.ac.kr
논문접수 : 2008년 12월 3일
심사완료 : 2009년 8월 27일

Copyright©2009 한국정보과학회 : 개인 복제이나 교육 목적인 경우, 이 저작물의 전체 또는 일부에 대한 복사본 혹은 디지털 사본의 제작을 허가합니다. 이 때, 사본은 상업적 수단으로 사용할 수 없으며 첫 페이지에 본 문구와 출처를 반드시 명시해야 합니다. 이외의 목적으로 복제, 배포, 출판, 전송 등 모든 유형의 사용행위를 하는 경우에 대하여는 사전에 허가를 얻고 비용을 지불해야 합니다.

정보과학회논문지 : 소프트웨어 및 응용 제36권 제11호(2009.11)

1. 서론

생물정보학(bioinformatics)은 다루는 데이터의 종류에 따라 서열, 구조, 발현, 네트워크 등으로 분류된다. 세포내에서는 생명 활동의 유지를 위해서 세포 내외에 존재하는 단백질 및 여러 물질들 간의 복잡한 네트워크를 이루고 있다[1]. 이러한 네트워크를 이해하는 것은 네트워크 내에 존재하는 각각의 단백질과 화합물들의 역할을 이해하는 데에 중요할 뿐만 아니라, 이와 관계된 질병의 원인을 알아내고 이에 적합한 약을 개발하는 데에도 기본이 되는 정보라 할 수 있다.

생물체 내에서 나타나는 이러한 네트워크 중 가장 기본적인 대사 경로(metabolic pathway)는 화합물이 효소

(단백질)에 의해 다른 화합물로 변화하는 과정을 보여주는 네트워크로, 본질적인 복잡성 때문에 대사 경로의 흐름을 한 눈에 쉽게 이해하고 파악할 수 있는 가시화 도구가 반드시 필요하다. 그러나 대사 경로를 자동으로 레이아웃 해주는 시스템의 경우 노드수가 증가할수록 에지 교차가 기하급수적으로 증가한다. 이러한 에지 교차는 가독성을 저해시키는 원인이 된다. 따라서, 유전체 수준(genome scale)에서의 대사 경로간의 관계(cross-talk) 등을 살펴보고 연구하기 위해서는 대사 경로 레이아웃 상에 나타나는 에지 교차수를 줄이면서 단계적으로 복잡한 대사 경로 레이아웃을 단순화시키는 작업이 필요하다.

현재까지 개발된 대부분의 가시화 도구 및 알고리즘은 2차원(2D) 평면에 대사 경로를 레이아웃하는 방법을 사용한다. 그러나 2차원 상에서는 복잡한 대사 경로 그래프의 위상(topology)을 표현하는 데 있어서 표현영역이 제한적이고, 노드수가 증가할수록 에지 교차가 기하급수적으로 증가할 뿐만 아니라 대사 경로를 다른 각도(viewpoint)로 관찰할 수 없기 때문에 대사 경로를 표현하고 분석하는 데 한계가 있다. 이러한 측면에서 3차원(3D) 공간에서는 복잡한 대사 경로 그래프를 표현하는 데 공간적으로 제한적이지 않고, 2차원 레이아웃 결과에서 빈번히 나타나는 에지 교차를 줄임으로써 가독성을 높일 수 있으며, 대사 경로를 여러 각도에서 관찰함으로써 유연성 있는 대사 경로 구조 분석이 가능하다. 따라서 3차원 대사 경로 그래프 레이아웃 알고리즘 및 시스템은 효과적으로 대사 경로를 드로잉하고, 이러한 레이아웃 결과를 바탕으로 대사 경로의 구조를 체계적으로 분석하기 위한 중요한 도구가 된다.

본 논문에서는 2차원 평면에서의 제한적인 표현을 해결하고 좀 더 유연성 있는 2.5차원 대사 경로 레이아웃 모듈을 설계하고 구현한다. 2.5차원 레이아웃은 대사 경로의 구조적 특징을 고려하고 2-계층 레이아웃(2-Layered layout)을 이용하여 3차원 공간상에 대사 경로 그래프의 구조를 계층화하여 레이아웃한다. 따라서 표현영역을 3차원으로 확장함으로써 대사 경로의 구조 분석을 용이하게 하고 기존 2차원 레이아웃과 비교하여 에지 교차를 줄이는 결과를 얻는다.

2. 기존 연구

2.1 2차원상의 대사 경로 레이아웃 알고리즘

기존에 개발된 2차원 대사 경로 레이아웃 알고리즘으로는 EcoCyc에 적용된 알고리즘, M. Y. Becker와 I. Rojas가 제안한 대사 경로 레이아웃 알고리즘 등을 들 수 있다[2,3]. 기존 알고리즘들의 경우는 대사 경로 그래프의 환형·계층적인 특징에 기반하여 대사 경로 그래

프를 나누어 레이아웃한다. 그러나 이러한 알고리즘의 경우 노드와 에지수가 증가할 경우 에지 교차가 기하급수적으로 증가하여 전체적인 레이아웃 결과가 항상 적절하지 않다는 단점이 있다. 또한 화합물 간의 다(多)대 다(多)관계를 표현하기 위해 하이퍼에지가 빈번히 나타난다. 이러한 하이퍼에지는 레이아웃 결과에 잦은 에지 교차를 발생시킴으로써 가독성을 떨어뜨린다는 문제점이 있다.

기존 연구의 문제점을 보완하면서 대사 경로 그래프가 '척도 없는(scale-free) 네트워크'라는 특징을 고려하여, 여러 형태의 컴포넌트가 복합적으로 존재하는 그래프도 그려줄 수 있는 개선된 대사 경로 드로잉 알고리즘을 제안하였다. 척도 없는 네트워크 구조를 고려한 경우 노드 증가에 따른 에지 교차 증가 현상을 완화시켰고 특히, 사이클이 존재하지 않는 대사 경로 그래프의 경우 더욱 효과적인 결과를 얻었다. 하지만 그 표현에 있어 제한적이고 특히 복잡한 구조를 갖는 대사 경로의 경우 에지 교차도 기하급수적으로 증가하여 가독성을 떨어뜨린다.

2.2 3차원 대사 경로 가시화 시스템

2차원 상에서의 대사 경로 레이아웃 알고리즘이 제한된 영역에 레이아웃 결과를 나타내고 여러 각도에서의 대사 경로를 관찰이 불가능하다는 한계점을 해결하기 위해 3차원 상에 대사 경로를 레이아웃해주는 시스템들이 제안되었다.

Y. Yang은 3차원 공간상에서 그래프 레이아웃, 컴퓨터 그래픽스, 가상현실을 결합한 상호작용(interactive)이 가능한 대사 경로 네트워크 시각화 시스템 개발하였다[4]. 이는 대사 경로 네트워크를 시각화하는데 있어 이론적으로 그 구조를 분석하고 이를 고려한 새로운 레이아웃 알고리즘을 제안하고 적용한 것이 아니라 단순히 가상현실에 표현함으로써 다양한 각도에서 대사 경로를 관찰할 수 있다는 것만을 강조하고 있다.

I. Rojdestvenski도 VRML을 이용하여 가상현실 속에 대사 경로 네트워크를 시각화해주는 시스템을 제안하였다[5]. 하지만 이 시스템도 마찬가지로 그래프 이론에 기반하여 대사 경로의 구조를 고려한 레이아웃 알고리즘을 새로 제안한 것이 아니라 단순히 '3차원 force-directed 레이아웃 알고리즘'과 같은 기존의 알고리즘을 전체 대사 경로 그래프에 적용하여 네트워크를 표현하고 있다.

따라서 대사 경로를 분석하고자 하는 시각화 시스템에 있어 그래프 이론에 초점을 맞춘 유연성 있는 새로운 3차원 대사 경로 레이아웃 방법이 필요하다.

3. 2-계층 레이아웃(2-Layered layout)을 이용한 2.5차원 대사 경로 레이아웃 모듈

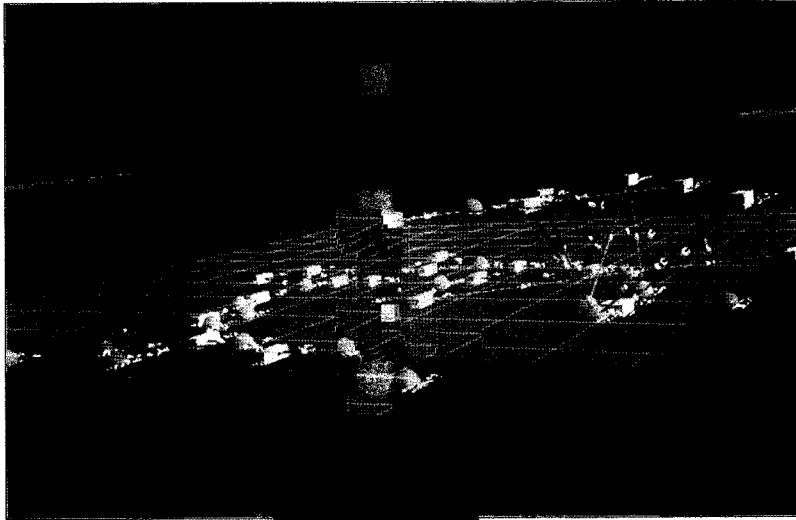


그림 2 3차원 상에서 척도없는 네트워크 기반 2-계층 레이아웃 모델을 적용한 대사 경로 그래프의 레이아웃:
그림 1에서 선택한 연결성 높은 노드와 그 이웃노드들을 3차원의 2-계층 중 상위 계층에 레이아웃한다.

표 2 환형 컴포넌트 기반 레이아웃 모델

[단계 1] 대사 경로 그래프에 존재하는 모든 환형 컴포넌트를 검색한다.
[단계 2] 검색된 환형 컴포넌트 중 상위 계층에 레이아웃할 컴포넌트를 선택한다.
[단계 3] 선택된 환형 컴포넌트를 상위 계층에 레이아웃한다.

아웃해 줌으로써 효과적으로 분석이 가능하다.

환형 컴포넌트 기반 레이아웃 과정은 표 2와 같다. 먼저 대사 경로 그래프에 존재하는 모든 환형 컴포넌트를 검색한다. 좋은 레이아웃 결과를 얻기 위해서는, 가장 큰 환형 컴포넌트를 찾는 것이 중요하다. 그러나 이 문제를 푸는 것은 NP-complete 문제로 알려져 있다. 그러므로 본 구현에서는 휴리스틱한 방법을 이용하여 가장 크다고 여겨지는 환형 컴포넌트를 찾도록 하였다. 먼저 임의의 환형 컴포넌트를 찾아 이 컴포넌트 상의 에지 하나를 제거하고, 에지 하나를 제거한 전체 그래프 상에서 다시 환형 컴포넌트를 찾아 이 환형 컴포넌트 상의 에지를 또 제거하여 환형 컴포넌트를 찾는 과정을 반복한다. 이렇게 찾아낸 환형 컴포넌트들 중 가장 큰 것을 그래프 상의 가장 큰 환형 컴포넌트로 반환한다. 다음으로 가장 큰 환형 컴포넌트를 제외한 부분을 연결성분으로 나누어 각각의 연결성분 안에서 다시 가장 큰 환형 컴포넌트를 찾는다. 이와 같은 방법으로 찾은 대사 경로내의 모든 환형 컴포넌트들 중에 상위 계층에 레이아웃하고자 하는 컴포넌트를 한 개 이상 임의로 선택한 후 레이아웃 결과를 얻는다. 그림 3과 그림 4는 2차원과 2.5차원에서의 레이아웃 결과이다.

3.3 연결성 높은 부그래프(highly connected subgraph) 기반 레이아웃 모델

연결성이 높은 부그래프(highly connected subgraph 또는 cliquish)란 부그래프 내에 포함되는 노드 간에는 연결성이 높고 그래프의 나머지 노드와는 연결성이 낮은 그래프를 말한다. 즉, 서로 상호작용이 높은 군집(cluster)을 의미한다. 이것은 단백질 상호작용 데이터를 기본으로 하여 단백질 복합체 및 대사경로 정보를 추출할 수 있음은 물론 동일 군집에 속하는 단백질들의 기능 추정이 가능하다[10,11]. 특히, 분자 네트워크 분석에 있어 중요한 대사경로 상에 존재하는 연결성 높은 부그래프를 찾아 각각의 부그래프를 계층적으로 레이아웃함으로써 대사 경로의 구조를 단순화하여 대사 경로 내의 단백질 간의 관계들을 효율적으로 분석할 수 있다.

표 3은 연결성 높은 부그래프 기반의 레이아웃 수행 과정이다. 먼저 원하는 연결성 높은 부그래프에 대한 밀도값(density)을 입력한다. 일반적으로 그래프 밀도값 Q 는 0~1사이의 값을 가지며, 0인 경우 연결성이 전혀 없는 그래프, 1인 경우 완전 그래프, 이외의 경우($0 < Q < 1$)는 연결성이 있는 그래프로 나눌 수 있다. 따라서 사용자가 원하는 그래프 밀도값을 입력한 후 입력한 그래프 밀도값에 대한 연결성 높은 부그래프를 결과로 얻을 수 있다. 입력으로 받은 그래프 밀도값에 대한 부그래프 찾는 알고리즘은 다음과 같다. 먼저 그래프 상에 존재하는 임의의 에지 2개를 선택하여 4개의 노드에 대한 그래프 밀도값 즉, 노드 사이에 몇 개의 에지가 존재하는지를 계산한다. 계산된 그래프 밀도값이 입력받은 밀도값보다 큰 경우 연결성 높은 부그래프로 결정한다. 찾은 부그래

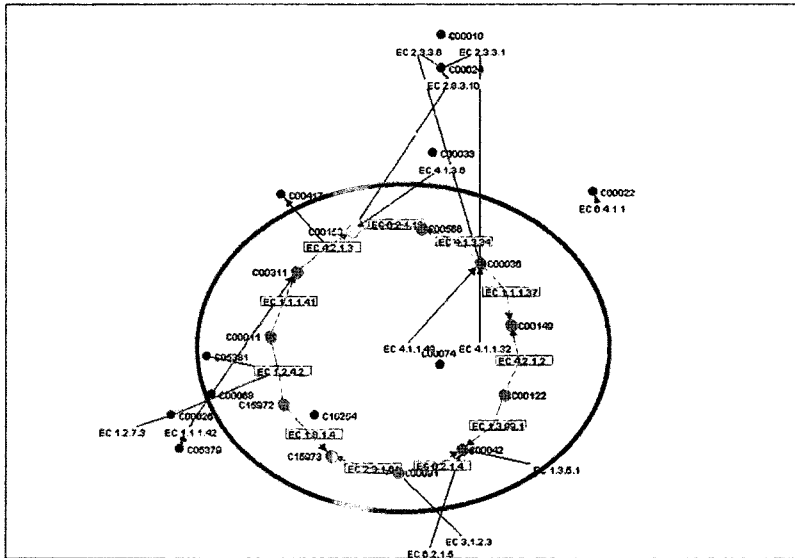


그림 3 2차원 대사 경로 그래프에 존재하는 환형 컴포넌트:

2차원 대사 경로 레이아웃 결과로부터 대사 경로 그래프에 존재하는 환형 컴포넌트를 찾을 수 있다.

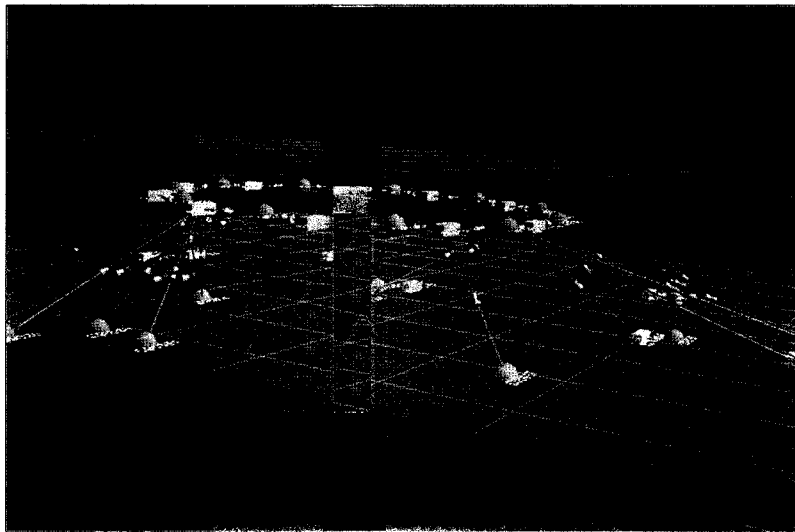


그림 4 3차원 상에서 환형 컴포넌트 기반 2-계층 레이아웃 모듈을 적용한 대사 경로 그래프의 레이아웃 결과: 그림 3에서 대사 경로에 존재하는 하나의 환형 컴포넌트를 상위 계층에 레이아웃한다.

표 3 연결성 높은 부그래프 기반 레이아웃 모듈

- [단계 1] 그래프의 밀도값을 입력한다.
- [단계 2] 입력한 밀도값에 만족하는 부그래프를 검색한다.
- [단계 3] 검색된 부그래프 중 3차원에서의 2-계층 중 상위 계층에 레이아웃하고자하는 부그래프를 선택한다.
- [단계 4] 선택된 부그래프를 상위 계층에 레이아웃한다.

프레 노드를 한 개씩 추가하면서 그래프 밀도값을 계산하고 연결성 높은 부그래프인지를 체크한다. 입력조건에

맞는 부그래프들 중 원하는 부그래프를 선택하고 이에 속하는 노드들을 상위 계층에 레이아웃시킨다. 그림 5와 그림 6은 2차원과 2.5차원에서의 연결성 높은 부그래프 기반 모듈을 이용한 대사 경로 레이아웃 결과를 각각 보여준다.

4. 실험 및 실험 결과

4.1 사용 언어와 실행 환경

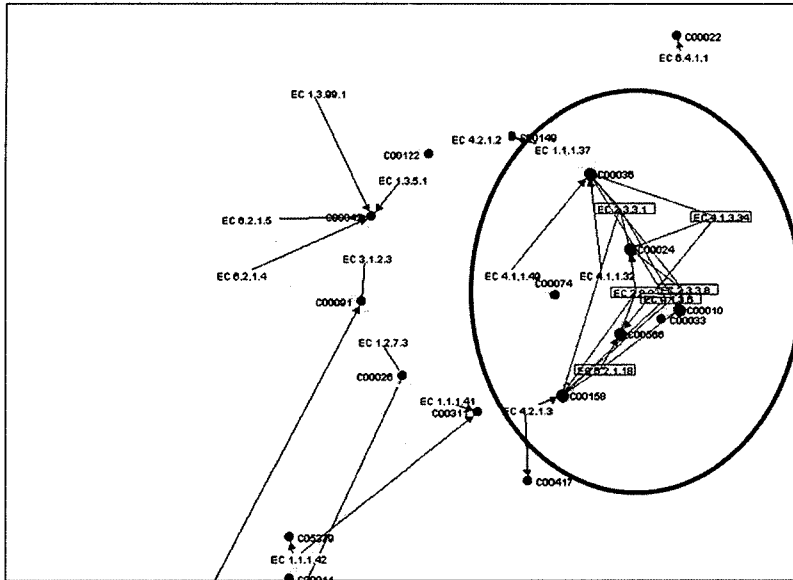


그림 5 2차원 대사 경로에 존재하는 연결성 높은 부그래프:

먼저 찾고자 하는 부그래프의 밀도값을 입력받고 그 값에 대한 연결성 높은 부그래프를 찾는다. 그 중 선택된 하나의 부그래프를 나타내고 있다.

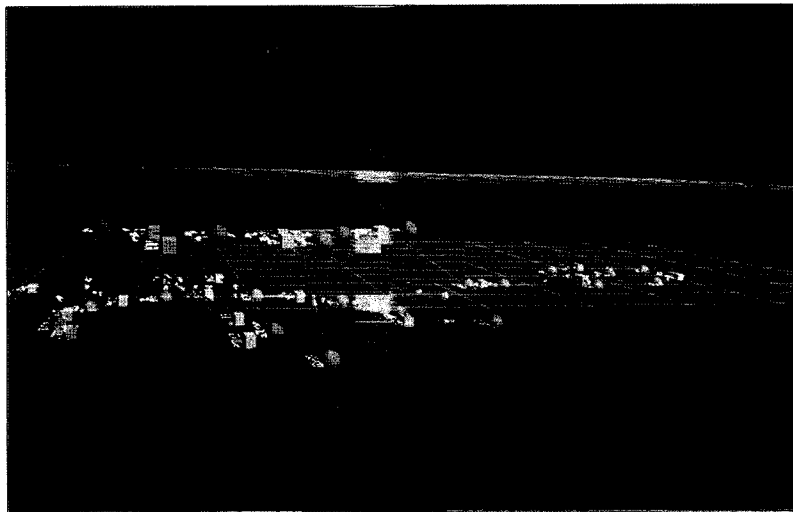


그림 6 3차원 상에서 연결성 높은 부그래프 기반 2-계층 레이아웃 모듈을 적용한 대사 경로 그래프의 레이아웃 결과: 그림 5에서 선택한 연결성 높은 부그래프를 3차원상의 2-계층 중 상위 계층에 레이아웃한다.

본 실험은 마이크로소프트 윈도우 XP 환경에서 수행하였으며, Java, 3차원 상에 대사 경로 그래프를 표현하기 위해 JAVA3D와 Java 기반 그래프 라이브러리인 YFiles를 사용하여 구현하였다[12].

4.2 인터페이스

그림 7은 2.5차원 대사 경로 드로잉을 위한 인터페이스의 초기 화면을 나타낸다. 레이아웃하고자 하는 대사

경로에 대한 데이터 파일을 열어 알고리즘을 수행하면 2차원 레이아웃 결과가 2차원과 3차원 공간상에 동시에 레이아웃된다. 2차원 상에서의 레이아웃 결과는 Editor 버튼을 누르면 확인할 수 있다.

2차원 레이아웃 결과에 대해 대사 경로에 존재하는 연결성 높은 노드, 환경 컴포넌트, 연결성 높은 부그래프를 찾을 수 있다. 원하는 주요 컴포넌트를 검색하면

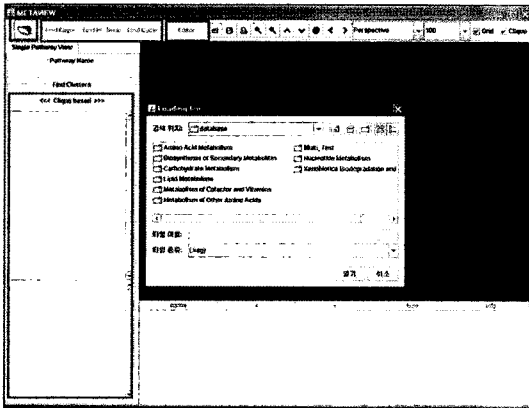


그림 7 2.5차원 대사 경로 드로잉 인터페이스

왼쪽 아래 창에 찾은 컴포넌트 리스트를 확인할 수 있다. 3차원 상의 상위 계층에 레이아웃 하고자하는 컴포넌트들을 선택하면 바로 2-계층 레이아웃 알고리즘이 적용된다. 2차원 레이아웃 상에서는 상위 계층에 레이아웃된 컴포넌트를 확인할 수 있고, 그 결과 얻어진 2.5차

원 레이아웃 결과는 오른쪽 화면에서 얻을 수 있다.

2.5차원 대사 경로 드로잉을 위한 모듈은 <http://blog.daum.net/milky73>에서 다운로드하여 사용해 볼 수 있다.

4.3 기존 방법과의 비교 분석

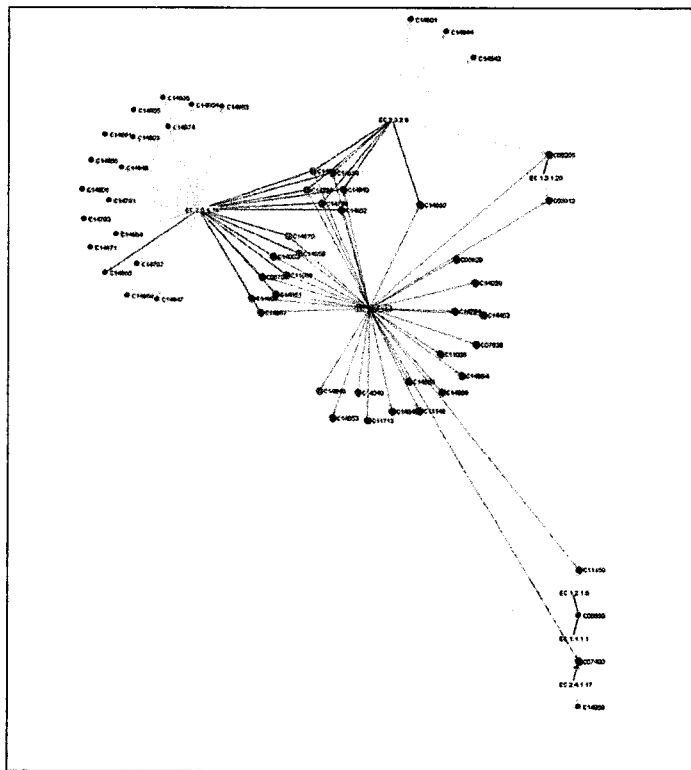
위와 같은 환경을 기반으로 하여 KEGG에서 제공하는 8개 카테고리에 속하는 대사 경로들을 실험데이터로 이용하였다.

약 80개의 대사 경로 그래프를 대상으로 기존에 제안한 방법과 본 논문에서 제안한 3차원 2-계층 레이아웃 방법을 적용하여 대사 경로 레이아웃 결과를 얻었다.

연결성 높은 노드 기반 모듈, 환형 컴포넌트 기반 모듈, 연결성 높은 부그래프 기반 모듈 적용하여 얻은 레이아웃 결과와 기존 2차원 대사 경로 레이아웃 알고리즘을 수행한 후 얻어낸 레이아웃 결과에 대한 에지 교차 발생수를 비교 분석하였다.

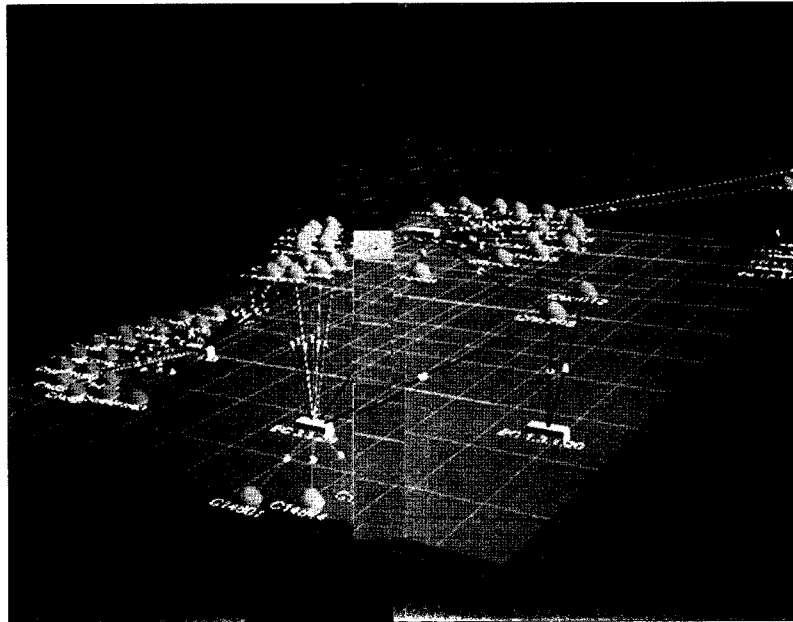
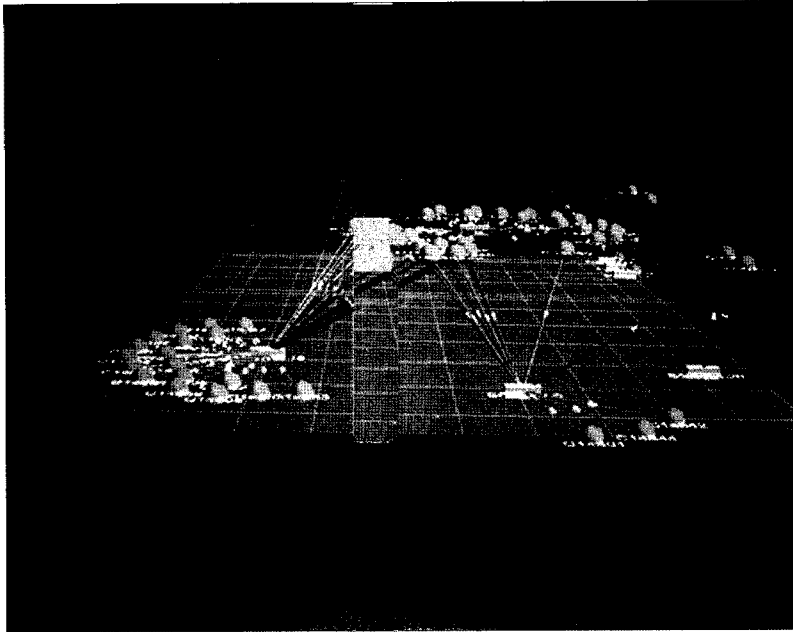
(1) 척도 없는 네트워크 기반 모듈 적용 결과

그림 8의 (a)는 Xenobiotics Biodegradation and Metabolism에 속하는 대사 경로 'Metabolism of xenobiotics by cytochrome p450'에 대한 2차원 레이아웃 결



(a) 척도 없는 네트워크 기반의 2차원 레이아웃 결과: 연결성 높은 노드가 존재하는 'Metabolism of xenobiotics by cytochrome p450'의 2차원 레이아웃 결과이다.

그림 8 'Metabolism of xenobiotics by cytochrome p450'에 대한 레이아웃 결과 (계속)



(b) 척도 없는 네트워크 기반의 2.5차원 레이아웃: (a)의 대사 경로 그래프에 존재하는 연결성 높은 노드 중 선택된 연결성 높은 노드와 그 이웃노드들을 2-계층 레이아웃을 적용하여 3차원 상에서의 상위 계층에 레이아웃한 후 여러 각도에서 관찰한 결과이다.

그림 8 'Metabolism of xenobiotics by cytochrome p450'에 대한 레이아웃 결과

파이다. 기존에 제안한 척도 없는 네트워크를 고려한 2차원 레이아웃 알고리즘을 적용한 경우 고려하지 않은 경우에 비해 에지 교차수가 64개에서 24개로 줄어든다. 같은 대사 경로에 3차원 2-계층 레이아웃 모듈을 적용한

경우 3개의 연결성 높은 노드가 검색되며 각각의 연결성 높은 노드와 그 이웃노드들을 상위 계층에 레이아웃한 결과 에지 교차수가 0, 1, 1개로 줄일 수 있었다(표 4 참고). 그림 8의 (b)는 선택된 연결성 높은 노드에 대해 3

표 4 척도 없는 네트워크 기반 2.5차원 레이아웃 결과 상의 에지 교차수

대사 경로 그래프		노드수	에지 교차수						연결성 높은 노드수
			2차원 레이아웃		2.5차원 레이아웃				
			척도 없는 네트워크를 고려하지 않은 경우	척도 없는 네트워크를 고려한 경우	척도 없는 네트워크 기반 모듈 적용				
					한 개의 연결성 높은 노드를 상위계층에 레이아웃하는 경우	두 개 이상의 연결성 높은 노드를 상위계층에 레이아웃하는 경우	하위계층 에지 교차수	상위계층 에지 교차수	
Carbohydrate Metabolism	Glycolysis_Gluconeogenesis	72	47	33	20	0	10	0	5
					15				
					21				
					16				
Glyoxylate and dicarboxylate metabolism	108	47	36	7	0	2	0	5	
				7					
				7					
				7					
Lipid Metabolism	Glycerophospholipid metabolism	91	421	121	58	0	13	0	10
					54				
					51				
					56				
					89				
					111				
					93				
					55				
83									
Amino Acid Metabolism	Glycine, serine and threonine metabolism	116	39	10	7	0	0	0	2
					7				
					24				
					19				
Valine, leucine and isoleucine degradation	75	86	32	22	0	3	1	7	
				30					
				19					
				13					
Metabolism of Other Amino Acids	Glutathione metabolism	47	49	19	3	0	1	4	3
					1				
Metabolism of Cofactors and Vitamins	One carbon pool by folate	33	39	19	2	0	1	2	4
					13				
					15				
Biosynthesis of Secondary Metabolites	Caffeine metabolism	27	15	12	0	0	-	-	2
					0				
Xenobiotics Biodegradation and Metabolism	Benzoate degradation via CoA ligation	98	40	29	0	0	1	1	3
					21				
Metabolism of xenobiotics by cytochrome P450		64	24	16	0	0	-	-	3
					1				
					1				

차원 2-계층 레이아웃을 적용한 레이아웃 결과이다.

표 4에서는 각각의 대사 경로를 척도 없는 네트워크에 기반하여 2.5차원 레이아웃을 적용한 경우, 레이아웃 결과에 나타난 에지 교차수가 줄어들었음을 확인할 수 있다.

Lipid Metabolism에 속하는 대사 경로 'Glycerophos-

pholipid metabolism'의 경우 10개의 연결성 높은 노드가 검색되었고 각각의 연결성 높은 노드를 상위 계층에 레이아웃한 결과 51~111개의 에지 교차가 발생하였다. 2차원 레이아웃 적용한 경우에 비해 에지 교차수가 줄어들었지만 여전히 많은 에지 교차가 존재한다. 따라서

이러한 경우 검색된 연결성 노드들 중 여러 개를 상위 계층에 레이아웃해 줌으로써 에지 교차수를 현저히 줄이는 결과를 얻을 수 있다. 표 4에서 각 대사 경로에 존재하는 연결성 높은 노드들 중 빗금 친 노드들을 상위 계층에 레이아웃하는 경우 현저히 감소한 에지 교차수를 확인할 수 있다. 'Glycerophospholipid metabolism' 대사 경로는 10개의 연결성 높은 노드 중 빗금 친 3개의 연결성 높은 노드와 그 이웃노드를 상위 계층에 레이아웃한 경우 에지 교차수를 13개로 줄일 수 있었다.

(2) 환형 컴포넌트 기반 모듈 적용 결과

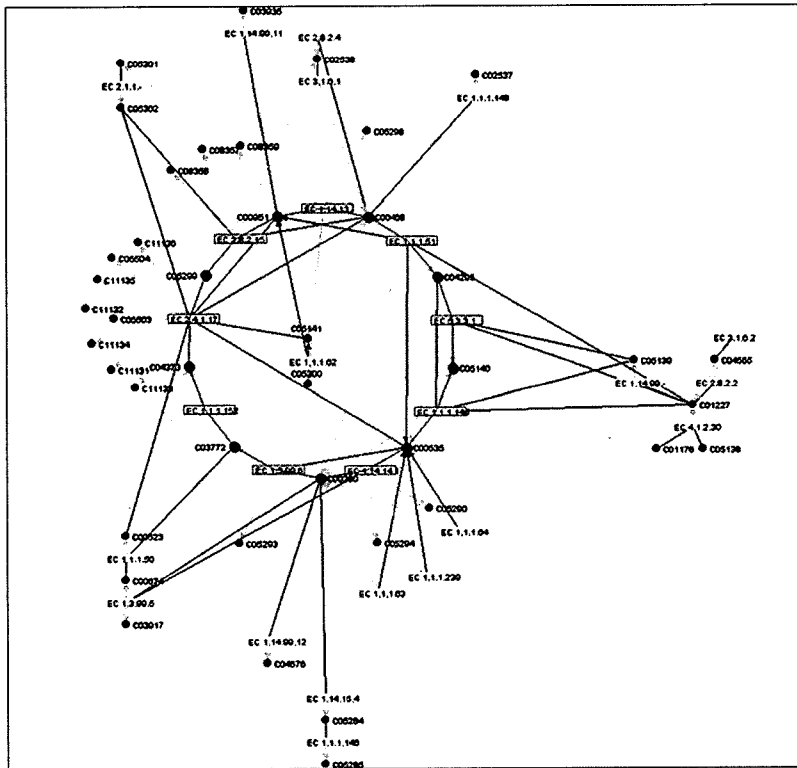
환형 컴포넌트 기반 모듈을 적용한 경우에도 레이아웃 결과 발생한 에지 교차수가 감소된 것을 확인할 수 있었다.

그림 9는 Lipid metabolism에 속하는 'Androgen and estrogen metabolism'에 대한 레이아웃 결과이다. 2차원 레이아웃에서 척도 없는 네트워크 구조를 고려하지 않은 경우와 비교하여 척도 없는 네트워크 구조를 고려한 경우 에지 교차수가 73에서 63으로 줄어들음을 알 수 있다(표 5 참고). 하지만 그림 9(a)에 나타난 것처럼 여전히 많은 에지 교차가 일어나는 것을 확인할 수 있다. 이러한 경우 대사 경로에 존재하는 하나의 환형 컴

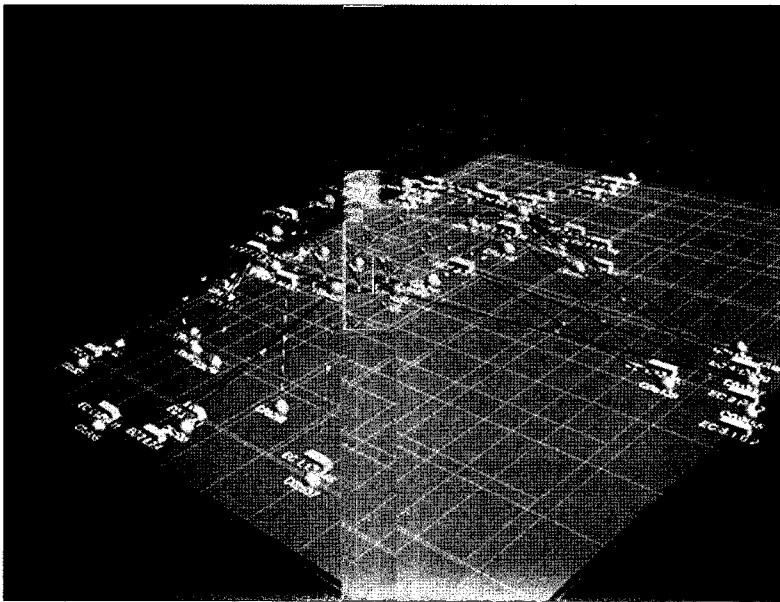
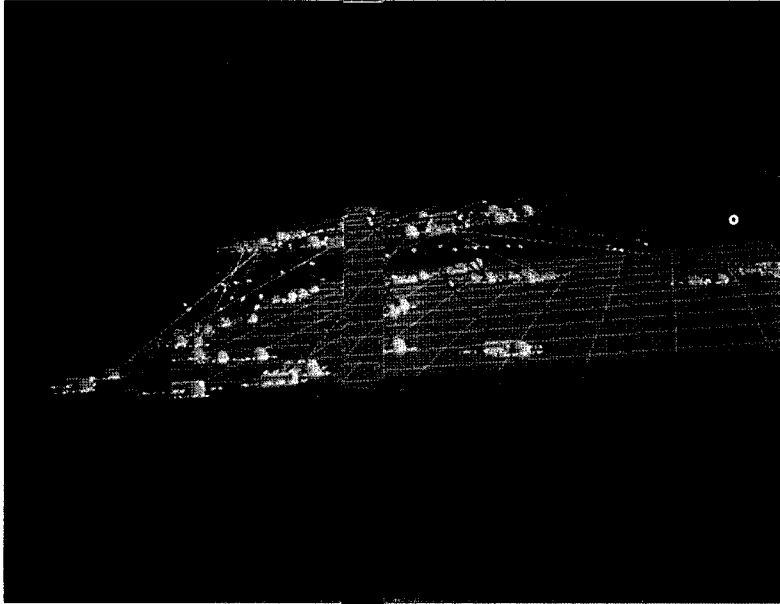
포넌트를 상위 계층에 올려줌으로써 에지 교차수가 감소된 레이아웃 결과를 얻을 수 있었다. 그림 9의 (b)는 환형 컴포넌트 기반의 2.5차원 레이아웃을 적용한 후 다른 각도에서 관찰한 예이다.

특히 여러 개의 환형 컴포넌트가 존재하는 대사 경로의 경우 존재하는 모든 환형 컴포넌트를 상위 계층에 레이아웃해 줌으로써 에지 교차수가 현저히 감소된 결과를 얻을 수 있었다. 그림 10의 (a)와 (b)는 각각 대사 경로 'Methionine metabolism'에 대한 2차원과 2.5차원 레이아웃 결과이다. 대사 경로 상에 존재하는 2개의 환형 컴포넌트를 각각 한 개씩 상위 계층에 레이아웃한 경우 0, 17개의 에지 교차가 발생하였다. 이에 반해 2개의 환형 컴포넌트 모듈을 상위 계층에 레이아웃함으로써 에지 교차가 전혀 일어나지 않는 결과를 얻을 수 있었다(그림 10의 (b)).

표 5는 환형 컴포넌트 기반 2.5차원 레이아웃을 적용한 경우 각 대사 경로의 레이아웃 결과에 나타난 에지 교차수를 보여준다. 여러 개의 환형 컴포넌트가 존재하는 경우 모든 환형 컴포넌트를 상위 계층에 올린 경우 모두 좋은 결과를 얻었다.



(a) 환형 컴포넌트가 존재하는 'Androgen and estrogen metabolism'의 2차원 레이아웃
그림 9 'Androgen and estrogen metabolism'에 대한 레이아웃 결과 (계속)



(b) 한 개의 환형 컴포넌트가 존재하는 대사 경로의 2.5차원 레이아웃: (a)의 2차원 대사 경로 그래프 상에 존재하는 한 개의 환형 컴포넌트를 선택한 후, 2-계층 레이아웃을 적용하여 3차원 상에서의 상위 계층에 레이아웃한다. 2.5차원 레이아웃 결과는 여러 각도에서 관찰가능하다.

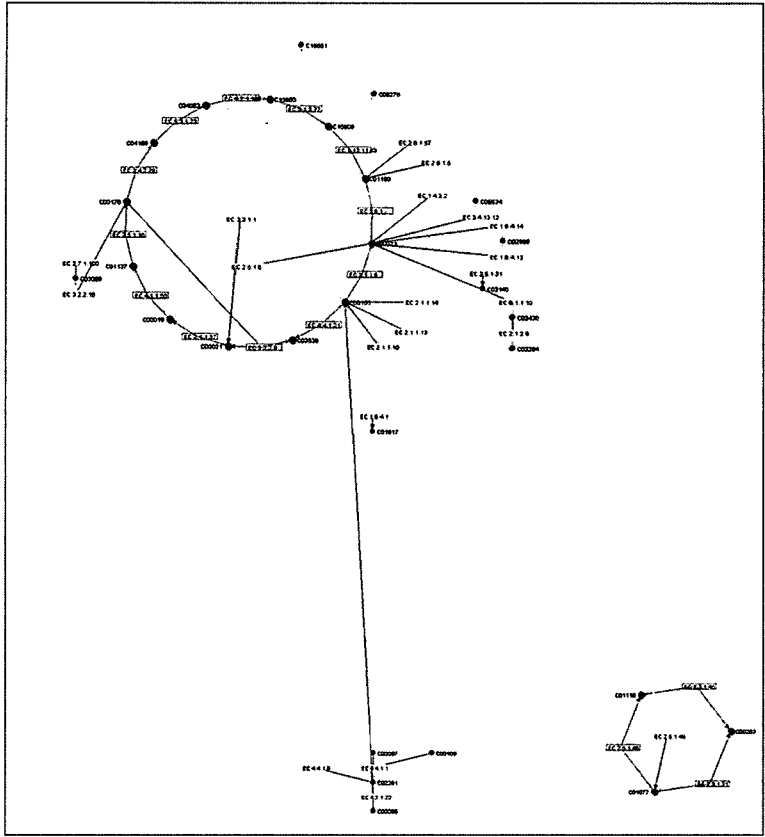
그림 9 'Androgen and estrogen metabolism'에 대한 레이아웃 결과

(3) 연결성 높은 부그래프 기반 모듈 적용 결과

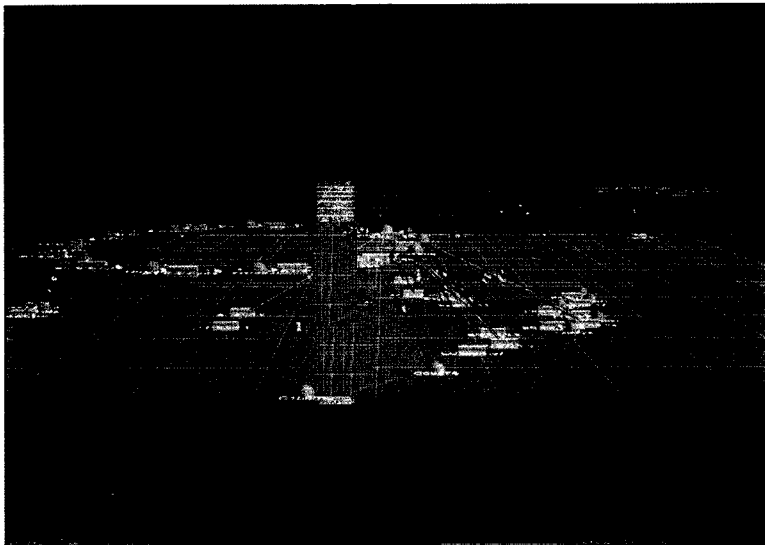
대부분의 대사 경로 상에 존재하는 연결성 높은 부그래프들은 그래프 밀도값 Q 가 0.8이하의 값에서 검색된다. 따라서 각 대사 경로에 대해 그래프 밀도값 $Q=0.8$ 인 연결성 높은 부그래프를 기반으로 하여 2.5차원 레이

아웃 모듈을 적용하였다.

실험 결과 연결성 높은 부그래프 기반 모듈을 적용한 경우에도 2차원 레이아웃에 나타난 에지 교차수를 많이 감소시킬 수 있었다. 그림 11은 Carbohydrate Metabolism의 'Pyruvate metabolism'에 대한 2차원 레이아웃



(a) 2개의 환형 컴포넌트가 존재하는 'Methionine metabolism'의 2차원 레이아웃



(b) 2개의 환형 컴포넌트가 존재하는 대사경로의 2.5차원 레이아웃: (a)의 2개의 환형 컴포넌트가 존재하는 대사 경로에 대해 모든 환형 컴포넌트를 3차원 상의 상위 계층에 레이아웃하여 예지 교차수를 줄인다.

그림 10 'Methionine metabolism'에 대한 레이아웃 결과

표 5 환형 컴포넌트 기반 2.5차원 레이아웃 결과상의 에지 교차수

대사 경로 그래프	노드 수	에지 교차 수				환형 컴포넌트 수			
		2차원 레이아웃		2.5차원 레이아웃					
		최도 많은 네트워크를 고려하지 않은 경우	최도 없는 네트워크를 고려한 경우	환형 컴포넌트 기반 모듈 적용					
				한 개의 컴포넌트를 상위 계층에 레이아웃하는 경우	두 개 이상의 컴포넌트를 상위 계층에 레이아웃하는 경우				
		하위계층 에지 교차수	상위계층 에지 교차수	하위계층 에지 교차수	상위계층 에지 교차수				
Carbohydrate Metabolism	Glycolysis, Gluconeogenesis	72	47	33	28 25 14	0 0 0	1 0 0	0 0 0	3
	Citrate cycle (TCA cycle)	45	19	14	0	3	-	-	1
	Pentose phosphate pathway	69	36	25	1	0	-	-	1
	Fructose and mannose metabolism	109	121	55	1	0	-	-	1
	Aminosugars metabolism	70	50	25	1	0	-	-	1
	Nucleotide sugars metabolism	69	27	10	5	0	0	1	2
	Propanoate metabolism	82	27	26	2	1	-	-	1
	Butanoate metabolism	92	24	14	4	1	-	-	2
	CS-branched dibasic acid metabolism	37	2	1	3	0	0	-	1
	Inositol phosphate metabolism	60	15	14	0	1	-	-	1
Lipid Metabolism	Bile acid biosynthesis	66	39	17	5 12	0 3	0 0	0 0	2
	Ether lipid metabolism	51	7	4	0	0	-	-	1
	Fatty acid elongation in mitochondria	34	79	75	0	5	-	-	1
	Glycerolipid metabolism	69	33	22	0 21	0 0	0 0	0 0	2
Amino Acid Metabolism	Alanine and aspartate metabolism	70	25	23	1 17	0 0	1 0	0 0	2
	Arginine and proline metabolism	97	36	30	11 14 21	0 5 3	0 0 0	0 0 0	3
	Cysteine metabolism	48	90	33	2	0	-	-	1
	Glutamate metabolism	62	33	14	0	0	-	-	1
	Histidine metabolism	74	5	3	0 3	0 0	0 0	0 0	2
	Lysine degradation	100	24	15	6	0	-	-	1
	Methionine metabolism	64	42	19	0 17	0 0	0 0	0 0	2
	Phenylalanine metabolism	84	43	27	0	0	-	-	1
	Tryptophan metabolism	135	122	112	1	1	-	-	1
	Urea cycle and metabolism of amino groups	91	70	39	0	1	-	-	1
Metabolism of Other Amino Acids	Valine, leucine and isoleucine biosynthesis	45	21	12	0	3	-	-	1
	beta-Alanine metabolism	60	14	7	6 1	0 0	1 0	0 0	2
Metabolism of Cofactors and Vitamins	Glutathione metabolism	47	49	19	1	1	-	-	1
	Biotin metabolism	23	12	11	0	1	-	-	1
	One carbon pool by folate	33	39	19	0	0	-	-	1
Biosynthesis of Secondary Metabolites	Vitamin B6 metabolism	52	42	25	0	8	-	-	1
	Alkaloid biosynthesis I	88	9	4	1	0	-	-	1
Xenobiotics Biodegradation and Metabolism	Caffeine metabolism	27	15	12	0	6	-	-	1
	Benzoate degradation via hydroxylation	98	34	31	1	4	-	-	1

과 2.5차원 레이아웃 결과이다.

표 6에서 각 대사 경로에 대해서 연결성 높은 부그래프 기반 모듈을 적용한 레이아웃 결과에 나타난 에지 교차수를 확인할 수 있다. 연결형 높은 부그래프 기반 모듈을 적용한 경우에도 여러 개의 부그래프가 존재하는 대사 경로 그래프에 대해 여러 개의 부그래프를 동시에 상위 계층에 레이아웃해 줌으로써 이전보다 더 좋은 레이아웃 결과를 얻을 수 있었다. Carbohydrate Metabolism의 'Starch and sucrose metabolism'의 경우 검색된 3개의 부그래프 중 빗금 친 2개의 부그래프를 상위 계층에 레이아웃하는 경우 하위계층과 상위계층의 에지 교차수를 각각 18, 14로 줄였다.

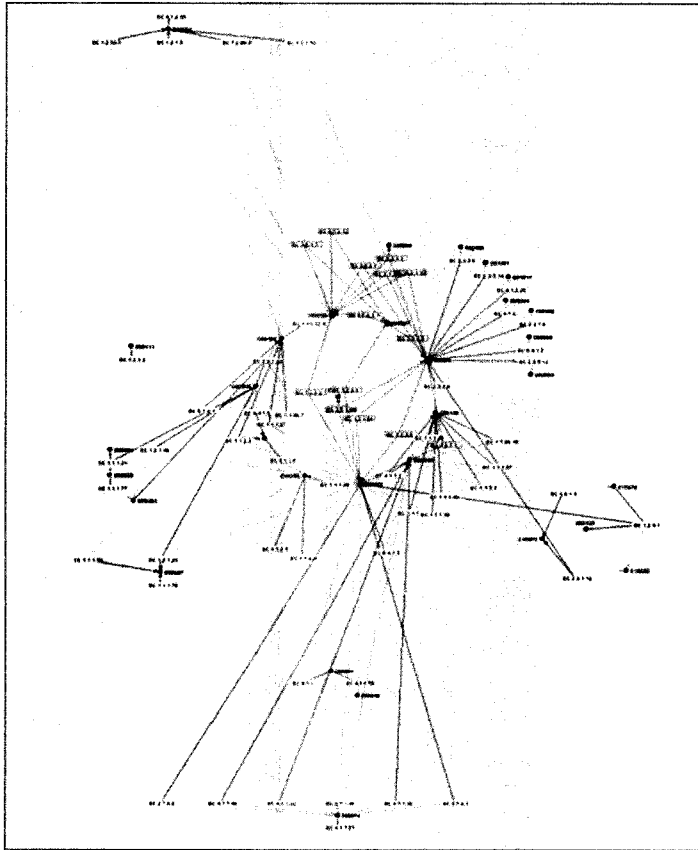
5. 결론 및 향후 연구과제

본 논문에서는 대사 경로 그래프 상에 존재하는 척도 없는 그래프, 환형 컴포넌트, 연결성 높은 부그래프와

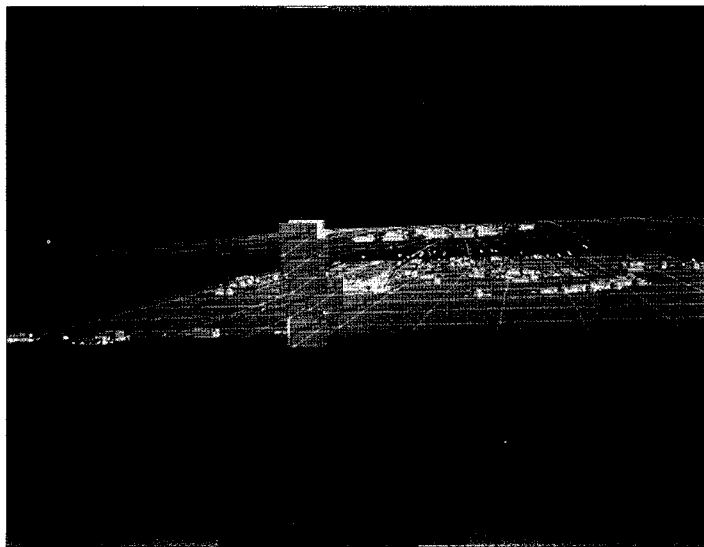
같은 대사 경로의 구조적인 특징을 기반으로 한 2.5차원 대사 경로 레이아웃 모듈을 설계 구현하였다.

KEGG에 속하는 80개의 대사 경로 그래프를 대상으로 실험한 결과, 2차원 대사 경로 레이아웃 알고리즘을 적용한 경우보다 에지 교차가 줄어든 결과를 얻을 수 있었다. 특히 여러 개의 컴포넌트들이 존재하는 경우 여러 개의 컴포넌트를 동시에 상위 계층에 레이아웃해 주는 경우 에지 교차수가 현저히 줄어든 결과를 얻을 수 있었다.

2-계층 레이아웃을 이용한 2.5차원 대사 경로 레이아웃 모듈은 3차원 레이아웃의 장점을 최대한 이용하여 대사 경로가 내포하고 있는 정보들을 좀 더 효율적으로 분석할 수 있는 작업공간을 제공한다. 또한 단계적으로 대사 경로 구조를 세분화, 계층화하여 세분화된 컴포넌트들의 생물학적 의미를 찾고 대사 경로 간의 상호관계를 파악하고 효율적이고 체계적으로 분석할 수 있는 기초가 된다.



(a) 연결성 높은 부그래프가 존재하는 'Pyruvate metabolism'의 2차원 레이아웃



(b) 연결성 높은 부그래프가 존재하는 대사 경로에 대한 2.5차원 레이아웃: 입력한 그래프 밀도값에 따른 연결성 높은 부그래프를 찾고 그 중 선택된 부그래프에 대해 2-계층 레이아웃을 적용하여 3차원 상의 상위계층에 레이아웃한다.

그림 11 'Pyruvate metabolism'에 대한 레이아웃 결과

표 6 연결성 높은 부그래프 기반 2.5차원 레이아웃 결과상의 에지 교차수

대사 경로 그래프		노드수	에지 교차수						부그래프 개수 (Q = 0.8)
			2차원 레이아웃		2.5차원 레이아웃				
			전도 없는 네트워크를 고려하지 않은 경우	전도 없는 네트워크를 고려한 경우	연결성 높은 부그래프 기반 모듈 추출		두 개 이상의 부그래프를 위계층에 레이아웃하는 상위계층에 레이아웃하는 경우		
		하위계층 에지 교차	상위계층 에지 교차	하위계층 에지 교차수	상위계층 에지 교차수				
Carbohydrate Metabolism	Citrate cycle (TCA cycle)	45	19	14	13	1	1	13	2
	Penitose phosphate pathway	69	36	23	7	3	-	-	1
	Galactose metabolism	78	49	38	10	2	-	-	1
	Starch and sucrose metabolism	127	374	150	122	0	18	14	3
					45	3			
	Pyruvate metabolism	98	161	126	38	17	-	-	1
Propanoate metabolism	82	27	26	8	7	-	-	1	
Lipid Metabolism	Bile acid biosynthesis	66	39	17	7	2	-	-	1
	Glycerolipid metabolism	69	33	19	5	2	-	-	1
	Glycerophospholipid metabolism	91	421	121	31	3	42	3	2
Amino Acid Metabolism	Arginine and proline metabolism	97	36	30	14	0	-	-	1
	Cysteine metabolism	48	90	33	5	2	-	-	1
	Valine, leucine and isoleucine degrad	75	86	32	15	0	24	1	2
Metabolism of Other Amino Acids	Glutathione metabolism	47	49	28	1	1	-	-	1
Metabolism of Cofactors and Vitamins	One carbon pool by folate	33	39	19	0	4	-	-	1
Biosynthesis of Secondary Metabolites	Terpenoid biosynthesis	33	19	11	0	8	-	-	1

향후 과제로는 2.5차원 대사 경로 레이아웃 모듈을 기반으로 하여 좀 더 체계적이고 효율적인 3차원 대사 경로 레이아웃 모듈로 확장함으로써 많은 양의 대사 경로 정보 데이터들을 쉽게 이해하고 파악할 수 있도록 유연성 있고 유지관리가 용이한 통합시스템을 구축하고 활용하는 것이다. 또한 3차원 대사 경로 레이아웃 알고리즘 개발을 통해 개별 대사 경로의 시각화뿐만 아니라 여러 개의 대사 경로의 집합체인 대사 경로 네트워크를 동시에 효과적으로 시각화하는 것이 중요한 연구과제가 될 것이다.

참 고 문 헌

[1] Y. Deville, 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*, p.174, 2003.

[2] P. D. Karp, M. Riley, M. Saier, I. T. Paulsen, J. Collado-Vides, S. M. Paley, A. Pellegrini-Toole, C. Bonavides and S. Gama-Castro, "The EcoCyc Database," *Nucleic Acids Research*, vol.30, no.1, pp.56-58, 2002.

[3] M. Y. Becker and I. Rojas, "A Graph Layout Algorithm for Drawing Metabolic Pathways," *BIOINFORMATICS*, vol.17, no.5, pp.461-467, 2001.

[4] Y. Yang, E. S. Wurtele, C. Cruz-Neira, 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.

[5] I. Rojdestvenski, "Metabolic pathways in three dimensions," *BIOINFORMATICS*, vol.19, no.18, pp.2436-2441, 2003.

[6] A. Barabasi and R. Albert, "Emergence of Scaling in Random Networks," *SCIENCE*, vol.286, pp.509-512, 1999.

[7] K. -I. Goh, B. Kahng and D. Kim, "Universal Behavior of Load Distribution in Scale-Free Networks," *Physical Review Letters*, vol.87, no.27, 2001.

[8] H. Jeong, B. Tombor, R. Albert, Z. N. Oltvai and A. -L. Barabasi, "The Large-scale Organization of Metabolic Networks," *NATURE*, vol.407, pp.651-654, 2000.

[9] P. Holme, M. Huss and H. Jeong, "Subnetwork Hierarchies of Biochemical Pathways," *BIOINFORMATICS*, vol.19, no.4, pp.532-538, 2003.

[10] V. Spirin and L. A. Mirny, "Protein complexes and functional modules in molecular networks," *PNAS*, vol.100, no.21, pp.12123-12128, 2003.

[11] M. Koyutürk, A. Grama and W. Szpankowski, "An efficient algorithm for detecting frequent subgraphs in biological networks," *BIOINFORMATICS*, vol.20

suppl.1, pp.i200-i207, 2004.

- [12] R. Wiese, M. Eiglsperger and P. Schabert, "The Y-files Graph Library: Documentation and Code available at <http://www-pr.informatik.uni-tuebingen.de/yfiles/>," 2000.



송 은 하

1996년 가톨릭대학교 수학과 학사. 1999년 이화여자대학교 공과대학 컴퓨터학과 석사. 1999년~2001년 한국과학기술연구원(KIST) 연구원. 2002년~현재 이화여자대학교 컴퓨터공학과 박사과정. 관심분야는 생물정보학, 데이터마이닝, 그래프

드로잉 등



함 성 일

연세대학교 건축공학과 졸업(학사, 석사) 2005년~현재 연세대학교 건축공학과 박사수료. 2008년~현재 이화여자대학교 국제대학교 기술융합과정 겸임교수. 관심분야는 BIM,CAD,유비쿼터스 컴퓨팅, 인터랙티브 컴퓨팅&디자인,바이오 인포메틱

스 등



이 상 호

1979년 서울대학교 계산통계학과 이학사 1981년 한국과학기술원 전산학과 이학석사. 1987년 한국과학기술원 전산학과 공학박사. 1990년 미국 일리노이대학교 전산학과 방문교수. 현 이화여자대학교 컴퓨터공학과 교수



박 현 석

서울대학교 전자공학과 학사. 1994년 펜실베이니아대학 전산학과 석사. 1997년 캠브리지대학 전산학과 박사. 1997년~1998년 동경대학교 Research Associate. 1999년~2001년 세종대학교 조교수. 2004년~2006년 (주) 마크로젠 대표이사. 2002년~

현재 이화여자대학교 조교수, 부교수