

최소 에지 크로싱을 위한 대사 경로 드로잉 알고리즘

송은하^o 김민경 이상호

이화여자대학교

ehsong@ewhain.net, {minkykim, shlee}@ewha.ac.kr

Metabolic Pathway Drawing Algorithm for Minimum Edge-crossing

Eun-Ha Song^o Min Kyung Kim and Sang-Ho Lee

Dept. of Computer Science and Engineering, Ewha Womans University

요약

기존의 레이아웃 알고리즘은 경로의 가독성과 속도 등에 주안점을 두어 개발되었다. 따라서 이러한 시스템의 경우 노드수가 증가할수록 에지 크로싱이 기하급수적으로 증가하는 문제가 있는데, genome scale에서의 대사 경로를 연구하기 위해서는 대사 경로 그래프 레이아웃 상에 나타나는 에지 크로싱을 줄이는 것이 시각화의 매우 중요한 부분이다. 대사 경로는 효소에 의한 화합물 간의 변화를 보여주는 네트워크로서, 대부분 척도 없는 네트워크 구조를 갖는다는 것이 알려져 있다. 이러한 대사 경로의 구조적 특징을 고려하여 에지 크로싱을 최소화하는 대사 경로 레이아웃 방법을 제안하고, 그 결과 노드수의 증가에 따른 에지 크로싱의 급격한 증가현상이 37~40%의 감소된 결과를 나타냈으며, 노드수가 증가하더라도 에지 크로싱이 오히려 감소하는 경우도 관찰되었다.

1. 서론

생물정보학(bioinformatics)은 다루는 데이터의 종류에 따라 서열, 구조, 발현, 네트워크 등으로 분류되고, 네트워크의 경우 전산학의 그래프 알고리즘이 적용되는 주제이다. 특히, 대사 경로를 비주얼하게 볼 수 있는 그래프로 작성, 연구하여 생명 활동을 총체적으로 이해할 수 있는 연구 분야인 대사체학(metabolomics)의 중요성이 부각되어 있다[1].

세포내에서는 생명 활동의 유지를 위해 세포 내외에 존재하는 단백질 및 여러 물질들 간의 복잡한 네트워크를 이루고 있다. 이러한 네트워크를 이해하는 것은 네트워크 내에 존재하는 각각의 단백질과 화합물들의 역할을 이해하는 데에 중요할 뿐만 아니라, 이와 관련된 질병의 원인을 알아내고 이에 적합한 약을 개발하는 데에도 기본이 되는 정보라 할 수 있다.

생물체 내에서 나타나는 이러한 네트워크 중 하나인 대사 경로는 어떤 화합물이 어떤 효소(단백질)에 의해 다른 화합물로 변화하는 과정을 보여주는 네트워크이다. 대사경로는 본질적인 복잡성 때문에 대사 경로 내의 흐름을 한 눈에 쉽게 이해하고 파악할 수 있는 대사 경로 가시화 도구가 필요하다. 특히 대사 경로의 구조적인 특징을 파악하여 이를 반영한 대사 경로 드로잉 알고리즘 설계가 중요하다.

현재 개발된 네트워크 뷰어로는 InterViewer, cytoscape, 등들이 있고 이들 중 대사 경로 네트워크를 고려하여 고안된 것으로는 M. Y. Becker와 I.Rojas가 제안한 대사 경로 레이아웃 알고리즘을 들 수 있다[1]. 이 알고리즘은 가독성을 증가시키기 위한 방안을 고려하였으며, 대사 경로 그래프의 환형·계층적인 특징에 기반하여 설계되었다.

본 논문에서는 대사 경로의 중요한 구조적 특징 중 하나로 알려진 척도 없는 네트워크에 기반하여[2] 이 특징을 고려한 대사 경로 그래프 레이아웃 알고리즘을 설계하고, 본래의 의도대로 알고리즘의 성능이 향상되었는가를 가독성과 에지 크로싱의 관점에서 평가하였다.

2. 기존연구

2.1 환형·계층적인 특징에 기반을 둔 알고리즘

환형·계층적인 특징에 기반을 둔 알고리즘은 주어진 대사 경로 그래프의 위상을 환형 컴포넌트가 없는 경우, 그래프 자체가 환형 컴포넌트인 경우, 그리고 이 두 가지 경우에 모두 속하지 않는 경우의 세 가지로 나누어 각각에 적합한 레이아웃 알고리즘을 적용해 준다.

이 알고리즘은 화합물 그래프를 이용하여 대사 경로 그래프를 표현하고 있어 하이퍼에지가 빈번히 나타난다. 이러한 하이퍼에지는 레이아웃 결과에 잦은 에지 크로싱을 발생시킴으로써 가독성을 떨어뜨린다는 문제점이 있다[1].

3. 대사 경로의 구조적 특징 및 드로잉 알고리즘

3.1 척도 없는(scale-free) 네트워크

척도 없는 네트워크란 다음 두 성질을 만족하는 네트워크를 말한다[3, 4]. 첫째, 주어진 네트워크에 계속하여 새로운 노드가 추가되고 둘째, 추가되는 노드들은 특정한 소수의 노드에 링크되려는 성질을 가진다. 그러므로 척도 없는 네트워크의 위상을 살펴보면 [그림 1]과 같이 매우 높은 연결성(connectivity)을 가지는 소수의 노드들과 이 노드들에 연결되어 있는 낮은 연결성을 가지는 노드들로 이루어져 있는 것을 볼 수 있다.

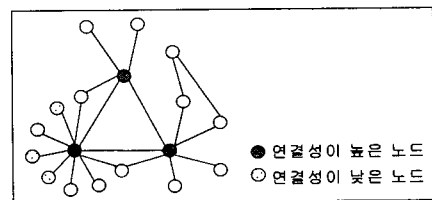


그림 1 척도없는 네트워크의 예

이와 같은 척도 없는 네트워크에 대해서 노드의 차수 (degree)값 k 와 각 노드가 k 개의 링크를 가지게 되는 확률값 $p(k)$ 의 관계를 좌표평면(plot)상에 표현해 보면 [그림 2]의 왼쪽 그림과 같이 기울기가 음수인 직선으로 나타난다. 따라서 주어진 네트워크에 대해 k 값과 $p(k)$ 값의 관계를 좌표평면 상에 그려보면, 이 네트워크가 척도 없는 네트워크인지 아닌지를 판단 할 수 있다[3, 4, 5]. [그림 2]의 오른쪽 그림은 WIT 데이터베이스의 43개의 생물체에 대한 대사 경로 네트워크의 k 값과 $p(k)$ 값의 관계를 그린 것이다. [그림 2]의 오른쪽에 나타난 음의 기울기를 가지는 직선으로 이루어 보았을 때, 대부분의 대사 경로는 척도 없는 네트워크의 클래스에 속한다는 것을 알 수 있다[5, 6]. 따라서 대사 경로 그래프를 레이아웃 해 줄 때, 연결성이 높은 단백질(노드)을 그 단백질(노드)에 연결된 컴포넌트들의 중심에 위치 시켜줌으로써 가독성을 높일 수 있다 [7].

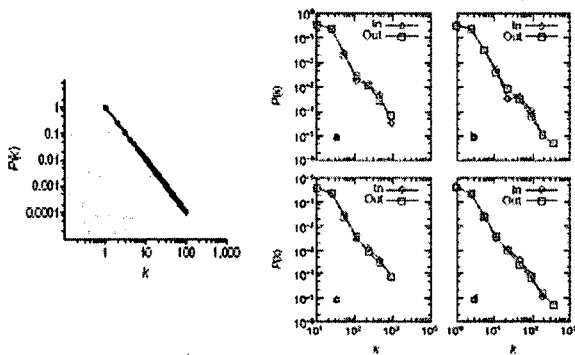


그림 2 척도 없는 네트워크의 k 값과 $p(k)$ 값의 관계

3.2 척도 없는 네트워크 구조를 고려한 대사 경로 드로잉 알고리즘

논문[2]에서 제안한 알고리즘은 입력으로 들어온 대사 경로 그래프의 구조적인 특성을 고려하여 적절한 레이아웃 모듈을 적용한다(표 2 참고).

표 2 구조적 특징을 고려한 대사 경로 드로잉 알고리즘

알고리즘: 구조적 특징을 고려한 대사 경로 드로잉 알고리즘 입력: 대사 경로 그래프 출력: 각 노드의 좌표 과정: 1. 연결성이 높은 노드의 존재 여부 검사 1.1 연결성이 높은 노드가 존재할 경우 '노드의 연결성을 고려한 레이아웃' 적용 1.2 연결성이 높은 노드가 존재하지 않을 경우 1.2.1 환형 컴포넌트의 존재 여부 검사 1.2.1.1 존재할 경우 '환형·계층적인 특징을 고려한 레이아웃' 적용 1.2.1.2 존재하지 않을 경우 '계층적 구성요소의 레이아웃' 적용

입력으로 들어온 대사 경로 그래프에 대해 첫 번째로 고려되는 특성은 노드의 연결성이다. 즉, 주어진 대사 경로 그래프가 척도 없는 네트워크 구조를 갖는지를 확인한다. 연결성 높은 노드가 존재하는 경우, 먼저 조건에 맞는 노드들을 선택한 후에, 선택된 노드들을 제외한 부분을 연결성분(connected component)으로 나누어 각 연결성분에 적절한 레이아웃 알고리즘을 적용한다. 연결성 높은 노드가 존재하지 않는 경우에는 단계적으로 환형, 계층적 구성요소의 존재유무를 확인하고 이에 적절한 레이아웃 알고리즘을 적용함으로써 전체 레이아웃 결과를 얻는다.

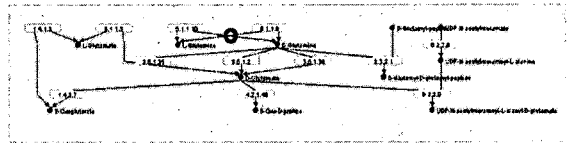
4. 성능평가

4.1 실험 환경 및 테스트 데이터

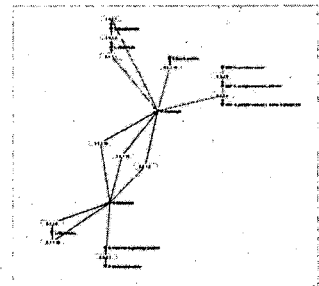
본 실험은 마이크로소프트 윈도우 XP 환경에서 수행하였으며, Java 프로그래밍 언어와 Java 기반 그래프 라이브러인 YFiles를 사용하여 구현한 테스트 알고리즘을 사용하였다[8]. 이러한 환경을 기반으로 하여 KEGG에서 제공하는 *Carbohydrate Metabolism*에 속하는 대사 경로를 대상으로, 척도 없는 네트워크 구조를 고려한 대사 경로 레이아웃 알고리즘과 고려하지 않은 레이아웃 알고리즘을 수행한 후 나타난 레이아웃 결과를 비교함으로써 척도 없는 네트워크 구조를 고려한 드로잉 알고리즘의 성능을 평가하였다.

4.2 실험결과

그림 3은 D- Glutamine and D-Glutamate 대사 경로에 대한 레이아웃 결과를 보여준다. (a), (b) 모두 에지 크로싱은 거의 일어나지 않지만, 척도 없는 네트워크 구조를 고려한 경우가 연결성 높은 노드들을 적절하게 배치하여 전체 대사 경로 그래프에 대한 가독성을 높였음을 확인할 수 있다.



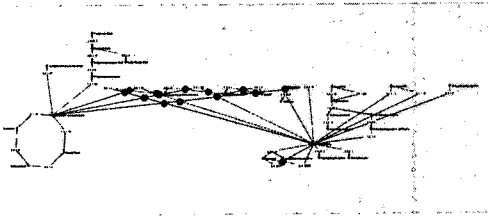
(a) 척도 없는 네트워크 구조를 고려하지 않은 경우



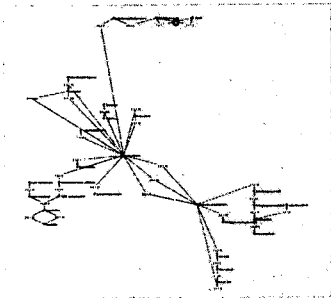
(b) 척도 없는 네트워크 구조를 고려한 경우

그림 3 D- Glutamine and D-Glutamate 대사 경로

그림 4는 좀 더 복잡한 구조를 갖는 β - Alanine 대사 경로에 대한 레이아웃 결과이다. 실험결과, 척도 없는 네트워크 구조를 고려한 경우에 높은 연결성을 가지는 노드를 중심으로 계층적 컴포넌트들이 적절하게 배치되어 있으며 에지 크로싱도 현저히 줄어들었음을 알 수 있다.



(a) 척도 없는 네트워크 구조를 고려하지 않은 경우



(b) 척도 없는 네트워크 구조를 고려한 경우

그림 4 β-Alanine 대사 경로

표 2는 Carbohydrate Metabolism에 속하는 5개의 대사 경로에 대한 실험결과를 정리한 것이다.

대사 경로 그래프의 노드수가 급격히 증가하는데 있어 척도 없는 네트워크 구조를 고려하지 않은 경우와 비교할 때 37~40%의 에지 크로싱 감소효과를 얻을 수 있었다. 따라서 척도 없는 네트워크 구조를 고려한 경우에 에지 크로싱이 줄어들어 가독성이 높일 수 있었음을 알 수 있다. 또한 사이클이 존재하지 않는 선형적 구조를 갖는 대사 경로의 경우, 에지 크로싱이 현저히 줄어들었고, 향상된 레이아웃 결과를 얻을 수 있었다.

표 2 각 대사경로에 대한 실험결과

대사경로	노드수	에지 크로싱 수	
		척도 없는 네트워크 구조 고려하지 않은 경우	척도 없는 네트워크 구조 고려한 경우
D-Glutamine and D-Glutamate	23	1	0
β-Alanine	60	13	1
Alanine aspartate	71	37	17
Pyroate	94	170	106
Aspartate/Glutamate	122	160	96

5. 결론 및 향후 연구과제

본 논문에서는 척도 없는 네트워크 구조를 고려한 대사 경로 드로잉 알고리즘의 효율성을 분석하였다. KEGG의 Carbohydrate Metabolism에 속하는 대사 경로 그래프를 대상으로 실험한 결과, 척도 없는 네트워크 구조를 고려한 경우 에지 크로싱이 현저히 줄어들었고 전체 대사 경로의 구조와 흐름을 이해하는 데 도움이 되는 레이아웃을 얻을 수 있었다.

향후 과제로는 척도 없는 네트워크 구조를 고려함에 있어 에지 크로싱의 수 뿐만 아니라 네트워크에 존재하는 연결성 높은

부그래프(clique) 등을 고려한 레이아웃 알고리즘을 적용해 볼 계획이다. 또한, 대사경로의 흐름의 가독성을 높이는 방안에 대한 고려도 필요할 것으로 생각된다.

6. 참고 문헌

- [1] M. Y. Becker and I. Rojas, "A Graph Layout Algorithm for Drawing Metabolic Pathways," *BIOINFORMATICS*, Vol. 17, No. 5, pp.461-467, 2001.
- [2] 이소희, 송은하, 이상호, 박현석, "구조적 특성에 기반한 대사 경로 드로잉 알고리즘의 설계 및 구현", 정보과학회 춘계학술대회, 2004.
- [3] A. Barabasi and R.Albert, "Emergence of Scaling in Random Networks", *SCIENCE*, Vol. 286, pp.509-512, 1999.
- [4] 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.
- [5] 정하중, 강병남, "복잡계의 이해-네트워크의 구조적 성질 및 그 응용," *물리학과 첨단기술* 10권 23, 2001.
- [6] 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.
- [7] P. Holme, M. Huss and H. Jeong, "Subnetwork Hierarchies of Biochemical Pathways," *BIOINFORMATICS*, Vol. 19, No. 4, pp.532-538, 2003.
- [8] 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.