

# 그래프 가시화 기술 조사

전효림\*, 이정훈\*, 한옥신\*\*

\*포항공과대학교 창의IT융합공학과

\*\*포항공과대학교 창의IT융합공학과/컴퓨터공학과

e-mail:{hrjeon, jhlee}@dmlab.postech.ac.kr, wshan@postech.ac.kr

## Survey on Graph Visualization Techniques

Hyo-Lim Jeon\*, Jeong-Hoon Lee\*, Wook-Shin Han\*\*\*

\*Dept of Creative IT Engineering, POSTECH

\*\*Dept of Creative IT Engineering/Dept of Computer Science and Engineering, POSTECH

### 요 약

그래프는 실세계의 객체와 정보를 표현하는 데이터구조로서 널리 사용되고 있다. 최근 들어 그래프 데이터의 활용도가 높아지고 다루는 그래프의 크기가 조 단위 규모로 증가함에 따라, 그래프 데이터에 관한 효과적인 가시화 기술에 대한 연구의 중요성도 증가하고 있다. 본 논문에서는 향후 그래프 가시화 툴 개발에 활용할 목적으로 그래프 데이터의 가시화를 위해 사용되는 기술을 조사하고 정리하였다.

## 1. 서론

그래프 데이터는 정점과 간선으로 구성된 기본적인 데이터 구조이며, 실세계의 객체와 정보를 표현하는 데 널리 사용되고 있다. 웹 그래프, 프로그램 코드, 시멘틱 웹, 인간의 뇌 구조 등이 모두 그래프로 표현될 수 있는 대표적인 객체들이다. 최근 소셜 네트워크 연구, 금융사기 방지 연구, 사이버 침입 감지 연구, 뇌 질환 연구 등에 그래프 데이터가 사용됨에 따라, 그래프의 활용도가 높아지고 있고 다루는 그래프의 규모도 수억에서 수조 규모로 증가하고 있다. 이에 따라 그래프 데이터를 효과적으로 가시화하는 연구 역시 중요성이 높아지고 있다. 그래프 데이터의 효과적인 가시화를 위해서는 그래프가 가진 의미를 효과적으로 파악할 수 있도록 가시성을 높이는 그래프 배치 기술과 그래프의 가시성을 평가할 수 있는 측정자에 관한 연구가 필수적이다. 또한 최근의 변화를 고려하여 대규모의 그래프 데이터를 효과적으로 가시화하는 빅 그래프 데이터 가시화 기술 연구가 필요하다. 본 논문에서는 3가지 측면에서 기존의 연구에서 사용하고 있는 기술들을 정리하고 대표적인 그래프 가시화 도구들을 소개한다.

## 2. 그래프 가시화 기술

### 2-1. 그래프 배치 기술

그래프 배치 기술은 가시성을 높일 수 있도록 효과적으로 그래프를 배치하는 기술이다. 대표적인 배치 기술로는 힘-기반 배치, 스펙트럼 배치, 트리 배치, 계층화된 그래프 배치, 아크 다이어그램, 원형 배치, 우성(Dominance) 그래프 배치 등이 있다. 힘-기반 배치는 스프링으로 연결

된 동적 객체들이 물리학의 법칙에 따라 위치가 안정화되는 과정을 그래프 가시화 기술에 활용한 배치 기술이다. 이 방법은 알고리즘이 단순하고 구현하기 쉽지만, 성능이 느리고, 힘이 최소가 되는 그래프의 배치가 로컬 최솟값을 지향하여, 다수의 정점으로 구성된 그래프를 배치하기에는 적합하지 않다. 대표적인 연구로는, Force Atlas2[8]가 있다. 스펙트럼 배치는 그래프에 대응하는 인접 혹은 라플라시안(Laplacian) 행렬로부터 구한 고유벡터를 좌표축으로 이용하는 방법이다. 이 방법은 그래프의 정점 및 간선의 영향을 점진적으로 증가/감소시킴으로써, 부드럽게 그래프를 배치한다. 트리 배치는 트리 형태의 그래프를 가시화하는 데 적합한 방법이며, 각 정점의 자식 정점을 그 정점을 중심으로 하는 원 상에 배치한다. 계층적 그래프 배치는 정점을 수평 계층들에 위치시키고, 어떤 계층에 있는 정점에서 다음 계층에 있는 정점으로 간선을 연결한다. 또한, 각 계층상에 위치하는 정점들 간에는 간선 접점이 최소가 되도록 배치한다. 이 방법은 플래나(Planar) 그래프나 DAG(Directed Acyclic Graph)를 배치하기에는 적합하지만, 일반적인 그래프는 사이클이 존재하기 때문에 간선이 서로 겹치지 않도록 그래프를 배치하기 위해 많은 휴리스틱을 사용한다. 아크 다이어그램 방법은, 유클리디언 평면상의 라인을 따라서 정점을 배치하고 라인의 위쪽 혹은 아래쪽에 반원 형태의 간선들로 정점을 연결한다. 원형 배치는 원 상에 그래프의 정점들을 배치하는 방법으로, 간선들은 아크 다이어그램에서와 유사하게 원상에 반원 형태로 배치된다. 이 방법은 정점들 간의 거리가 모두 같으므로, 정점이 위치에 따라 특권을 부여 받지 않는 소셜 네트워크와 같은 그래프 배치에 사용된다. 마지막으로 우성 그래프 배치는 어떤 정점에서 접근 가능한

†: 교신저자

본 연구는 2018년도 정부(과학기술정보통신부)의 재원으로 한국연구재단의 지원을 받아 수행된 연구임 (NRF-2017R1A2B3007116).

(Reachable) 정점들은 그 정점보다 큰 좌푯 값을 가지도록 배치하는 방법으로, 정점간의 접근 가능성을 잘 보여주는 그래프 배치 기술이다.

## 2-2. 그래프 품질 측정 기술

그래프의 효과적인 가시화를 위해서는 정성적인 평가 뿐만 아니라, 그래프의 품질을 평가할 수 있는 정량적인 평가도 중요하며, 이를 위해서는 그래프의 가시성을 판단할 수 있는 척도가 필요하다.

1) 배치된 그래프에서 서로 겹치는 간선의 개수는 중요한 척도 중의 하나이다. 간선의 겹침이 많으면 가시성이 떨어져 그래프의 의미를 파악하기 힘들다. 플래나 그래프나 그래프 임베딩 기법(예를 들어, 아크 다이어그램)이 적용 가능한 그래프에서는 이러한 문제가 없지만, 사이클이 존재하는 일반적인 그래프에서는 간선 간의 겹침이 존재하는 것이 일반적이다. 따라서 겹침이 최소화되도록 그래프를 배치 및 조작하는 기술이 필요하다. 2) 그래프가 배치되는 영역의 크기가 작을수록 인접한 정점 간의 거리를 줄여서 영역이 큰 경우보다 일반적으로 그래프의 가시성이 높다. 3) 간선의 길이 역시 중요한 측정자 중의 하나이다. 일반적으로 전체 간선 길이의 합과 최장 간선의 길이를 줄이는 것이 가시성을 높일 수 있다. 4) 그 밖에도 동일 정점에서 출발하는 간선들이 이루는 각의 크기인 각해상도, 간선의 경사도 개수 등이 있다.

## 2-3. 빅 그래프 데이터 가시화 기술

소셜 네트워크 등의 대규모 그래프 데이터 분석 연구가 활발해짐에 따라, 대규모의 그래프를 가시화하는 연구 역시 중요해지고 있다. 대규모의 그래프를 가시화하는 연구는 중요한 데이터들만을 추출하여 가시화하는 샘플링 기반의 가시화 기법[9]과 그래프 데이터를 병합하여 보여주는 그래프 병합 기법[10][7], 그리고 정해진 중요 정점을 기준으로 부분 그래프를 가시화[6]하는 기법이 있다.

샘플링 기반의 그래프 가시화 연구의 목표는 원본 데이터의 특성을 잘 반영할 수 있는 그래프 데이터 샘플링에 있다. 기존의 임의 표본 샘플링이나 층화 표본 샘플링이 원본 데이터의 특성을 잘 반영하지 못하는 문제가 있어, [9]는 원본 데이터의 회귀, 밀도, 그리고 클러스터링을 잘 반영하는 가시화 인지 샘플링 기법을 제시하였다. [10]은 중요 정점을 기반으로 그래프를 병합하는 기법으로 사용자에 의해 주어진 중요 정점을 중심으로 같은 레이블을 가지는 간선과 정점으로 구성된 서브 그래프를 병합하고, 단계별로 보여 준다. [7]은 대규모 그래프 데이터를 가시화하기 위해, 반복되는 패턴의 그래프를 특정 도형으로 가시화하는 기법을 제시하였다. 제안하는 방법은 패턴에 포함된 간선의 개수, 노드의 개수 등의 차이를 고려하여 도형의 크기를 달리하였다. [6]은 기계학습을 기반으로 사용자가 관심이 있을 것으로 판단되는 정점들을 추천해 주고, 선정된 정점을 기준으로 서브 그래프를 가시화한다.

## 3. 그래프 가시화 툴

현재 널리 사용되는 가시화 라이브러리 혹은 툴로는 그래프 비즈(Graphviz), 링크리어스(Linkurious), 자동 그래프 배치(Automatic Graph Layout), 사이토스케이프(Cytoscape), 게피(Gephi) 등이 있다.

그래프 비즈[3]는 오픈소스 그래프 가시화 소프트웨어로서, 계층화된 그래프 배치, 힘-기반 배치 기법인 스프링 모델, 원형 배치, 동심원 배치 등의 다양한 배치 방법과 그래프에 대한 설명을 텍스트로 작성할 수 있는 기능을 제공한다. 링크리어스[4]는 오그마(Ogma)라는 라이브러리를 사용해서 웹을 기반으로 그래프를 보여주는 계층, 노드제이엑스(NodeJS)를 기반으로 그래프 데이터베이스에 접근하여 데이터를 읽고 쓰는 로직 계층, 그리고 그래프 데이터베이스가 존재하는 데이터 계층으로 구성된다. 링크리어스는 백만 개 이상의 정점과 간선으로 구성된 그래프 데이터를 효과적으로 가시화할 수 있다. 또한, 웹 브라우저 상에서 사용자가 그래프를 수정할 수 있는 인터페이스를 제공한다. 마이크로소프트의 자동 그래프 배치[5]는 그래프 배치를 위한 엔진, 그리기 모듈 및 그래프의 렌더링 기능을 제공하는 뷰어(Viewer)로 구성된다. 뷰어는 그래프 렌더링 기능뿐만 아니라, 그래프의 객체 검색과 초점 기능을 제공한다. 사이토스케이프[1]는 분자 상호 작용 네트워크와 생물학적 경로를 시각화하는 데 널리 사용되는 가시화 툴이었으나, 최근에는 네트워크 분석 및 시각화에서도 사용되고 있다. 게피[2]는 힘-기반 배치 기법을 사용하며, 10만 개의 정점과 100만 개의 간선으로 구성된 그래프를 가시화할 수 있다. 또한 중심점 찾기, 최단 거리 찾기, 페이지 랭크 등 그래프 분석을 위한 기능들을 제공한다.

## 참고문헌

- [1] Cytoscape, [http://www.cytoscape.org/what\\_is\\_cytoscape.html](http://www.cytoscape.org/what_is_cytoscape.html).
- [2] Gephi, <https://gephi.org/features/>.
- [3] Graphviz, <https://www.graphviz.org/>.
- [4] Linkurious, <https://doc.linkurio.us/admin-manual/latest/>.
- [5] Microsoft Automatic Graph Layout, <https://www.microsoft.com/en-us/research/project/microsoft-automatic-graph-layout/>.
- [6] Akoglu, L., et al.: OPAvion: Mining and visualization in large graphs, ACM SIGMOD 2012: 717-720.
- [7] Dunne, C., et al.: Improving Network Visualization Readability with Fan, Connector, and Clique Glyphs. ACM SIGCHI 2013: 3247-3256.
- [8] Jacomy, M., et al.: ForceAtlas2: A Continuous Graph Layout Algorithm for Handy Network Visualization Designed for the Gephi Software. Plos one, vol 9, issue 6, e98679, pp. 1-12, (2014).
- [9] Park, Y., et al.: Visualization-aware Sampling for Very Large Databases, IEEE ICDE 2016: 755-766.
- [10] Sundara, S., et al.: Visualizing Large-scale RDF Data Using Subsets, Summaries, and Sampling in Oracle, IEEE ICDE 2010: 1048-1059.