

# 그래프 구조 정보를 이용한 효율적인 그래프 메디안 탐색 기법

박기성, 윤영선, 김태연, 이영구<sup>1)</sup>

경희대학교 컴퓨터공학과

e-mail:{kspark, yysun21, cutup, yklee}@khu.ac.kr

## An Efficient Technique using Graph Topology Information for Finding Graph Median

Kisung Park, Youngsun Yun, Taeyeon Kim Young-Koo Lee  
Dept of Computer Engineering, Kyung Hee University

### 요 약

최근 정보 기술의 발달로 XML, 화학 복합물, 소셜 네트워크 등과 같은 구조적 정보를 갖는 빅 데이터들이 대량으로 축적되고 있다. 이러한 구조적 정보를 갖는 그래프 데이터에서 메디안을 찾기 위한 연구가 진행되고 있다. 기존에는 그래프 메디안을 효율적으로 계산하기 위해 하한값을 이용한 그래프 메디안 탐색 기법이 제안되었다. 그러나 탐색을 시작하는 버텍스를 선정하는 데에 따라 가지치기 효과가 다르게 발생하는 문제점이 있다. 본 논문에서는 버텍스의 그래프 구조 정보를 이용한 효율적인 메디안 탐색 기법을 제안한다. 제안하는 탐색 기법은 버텍스의 차수와 에지 가중치를 이용하여 그래프 메디안 예측 값을 정의하고, 그래프 메디안과 유사한 버텍스들부터 우선적으로 탐색한다. 실험을 통하여 제안하는 기법이 기존의 방법보다 최대 10%까지 수행시간을 단축함을 보인다.

### 1. 서론

빅 데이터는 페타 바이트 이상의 규모를 갖는 데이터를 의미하며, 다양한 형태로 표현된다. 특히, 정보 기술의 발달로 화학 복합물, XML, 소셜 네트워크 등과 같은 구조적 정보를 갖는 데이터들이 대량으로 축적되고 있다. 이러한 구조적 정보를 갖는 데이터에서 메디안을 찾기 위한 연구가 진행되고 있다.

그래프 메디안[1]은 그래프 내의 모든 버텍스들 간의 최단 경로의 합이 최소가 되는 버텍스이다. 그래프 메디안을 계산하기 위하여 모든 쌍의 최단 경로(all-pair shortest path) 기법을 사용한다. 따라서 버텍스의 수가 많은 대용량 그래프에서는 메디안을 계산하기 위해 많은 시간이 요구된다.

[2]는 그래프 메디안의 하한 값을 정의하고, 브랜치-앤드-바운드(branch and bound) 탐색을 통해 성능을 개선하였다. 그러나 최단 경로의 합을 계산하는 버텍스를 무작위로 선정하여, 현재까지 계산한 그래프 메디안 값이 크게 측정된 경우 가지치기(pruning) 효과가 다르게 발생하여 성능이 저하된다.

그래프의 구조 정보는 그래프 메디안과 관련성이 있다. 높은 차수를 갖는 버텍스에서는 다른 버텍스들로 도달하는 경로의 길이가 작아져 최단 경로의 합이 작아지는 경향을 보인다. 또한 에지의 가중치가 작은 버텍스에서는 다

른 버텍스와의 최단 경로에 에지가 포함될 가능성이 커져 최단 경로의 합이 작아지는 경향을 보인다.

본 논문에서는 그래프 구조 정보를 이용한 효율적인 메디안 기법을 제안한다. 제안하는 방법은 그래프 메디안과 최단 경로의 합이 유사한 버텍스들부터 우선적으로 탐색하여 빠르게 그래프 메디안을 탐색한다. 실험을 통하여 기존의 기법보다 최대 10%까지 수행시간을 단축함을 보인다.

### 2. 관련 연구

그래프 메디안은 모든 쌍의 최단 경로 기법을 사용하여 계산한다. [1]에서는 그래프에서 메디안을 정의하고, 너비 우선 탐색으로 최단 경로의 합을 계산하며 메디안을 탐색하는 기법을 제안하였다. [2]에서는 그래프 메디안의 하한 값을 정의하여, 그래프 메디안의 계산을 위한 많은 양의 탐색공간을 제거하였다. 단일 쌍의 최단 경로 기법을 모든 버텍스에 대해서 반복수행하고, 각각의 버텍스에 대해서 최단 경로의 합을 계산한다. 그래프 메디안의 하한 값을 현재까지 방문한 최단 경로 합의 하한 값과 아직 방문하지 않은 버텍스들의 최소 가중치의 합으로 계산하였다. 그래프를 탐색하며 계산한 하한 값이 그래프 메디안보다 큰 경우, 그래프 메디안이 될 수 없으므로 해당 버텍스에 대해서 탐색을 중단한다. 그러나 초기에 그래프 메디안이 계산되는 값에 따라서 탐색공간이 줄어드는 효과가 달라지

1) 교신 저자

는 문제점이 발생한다.

### 3. 그래프 구조 정보를 이용한 효율적인 그래프 메디안 기법

기존의 하한 값을 이용한 그래프 메디안 기법은 최단 경로의 합을 계산하면서, 하한 값이 이전에 계산한 그래프 메디안보다 큰 경우에 더 이상 탐색하지 않는다. 이러한 기법은 그래프 내의 버텍스들을 탐색하는 순서에 따라서 탐색공간이 크게 달라진다. 따라서 초기에 선택하는 버텍스가 그래프 메디안 수치와 유사할 경우, 많은 양의 탐색 공간을 줄일 수 있다.

버텍스의 차수와 버텍스와 연결된 에지들의 가중치의 합과 같은 그래프 구조 정보를 이용하면 그래프 메디안 수치를 예측할 수 있다. 버텍스의 차수가 높을수록 직접적으로 연결된 버텍스들이 많아진다. 따라서 다른 버텍스로 방문하기 위한 경로의 길이가 작아져 최단 경로가 될 가능성이 높아진다. 또한 버텍스를 중심으로 특정 반경 내에 있는 에지들의 가중치 합이 작은 경우, 최단 경로가 될 가능성이 더욱 높아지게 된다.

그림 1은 그래프 구조 정보와 그래프 메디안과의 관계를 나타낸 실험 결과이다. 시작 버텍스를 중심으로 반경을 1, 2, 3으로 선정하여 그래프 메디안과 점수를 측정하였다. 반경은 시작 버텍스로부터 연결된 에지의 수를 의미한다. 가로축은 버텍스의 차수와 에지 가중치를 이용하여 계산한 점수를 나타낸다. 버텍스의 차수가 높을수록 낮은 점수를, 에지들의 가중치의 합이 작을수록 낮은 점수를 부여한다. 세로축은 점수를 기준으로 오름차순으로 정렬한 버텍스를 의미한다. 각 버텍스에서 측정된 점수가 작을수록 그래프 메디안과 유사한 경향을 보인다. 따라서 그래프 메디안 수치와 유사한 최단 경로의 합을 갖는 버텍스를 시작 버텍스로 선정하여 성능을 개선할 수 있다.

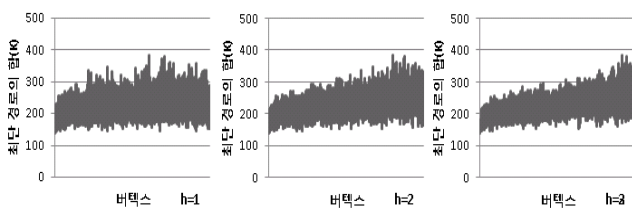


그림 1. 그래프 구조 정보와 그래프 메디안과의 관계 분석 실험

버텍스의 차수와 에지들의 가중치와 같은 그래프 구조 정보를 이용한 그래프 메디안의 척도를 식(1)과 같이 정의한다.

$$vScore(v, h) = (-1) * (degree(v) + \sum_{k=1}^h \sum_{p \in P, hop(p)=k} weights(p)) \quad (1)$$

식(1)은 반경을  $h$ 로 하여 시작 버텍스  $v$ 의 그래프 메디

안을 예측한 척도이다.  $v$ 의 차수와  $v$ 와 반경  $h$ 내의 에지 가중치들의 합을 통하여 점수를 계산한다.  $P$ 는 버텍스  $v$ 를 시작 지점으로 갖는 모든 경로들을 말하며,  $hop(p)$ 는 경로  $p$  내의 연결된 에지의 수를 의미한다. 그리고  $weights(p)$ 는 경로  $p$  내의 연결된 에지 가중치의 합을 의미한다. 계산한 값을 음수로 표기하여  $v$ 의 차수와 에지 가중치의 합이 클수록 작은 값을 갖게 설계한다.

알고리즘 1은 그래프의 구조 정보를 이용한 그래프 메디안 탐색 알고리즘이다. 알고리즘 1은 그래프  $G$ 와 반경  $h$ 를 입력받는다. 그리고 출력 값으로 그래프 메디안  $m$ 과 메디안 버텍스  $v_{median}$ 을 반환한다. 우선, 모든 버텍스에 대해서  $vScore$ 를 통해 그래프 메디안에 대한 예측 값을 측정하고(line1-2), 예측 값을 기준으로 오름차순으로 정렬한다(line 3). 하한 값을 이용한 그래프 메디안 기법을 통하여 메디안을 계산한다(line4-8).

<p><b>Algorithm</b> GTI-Median(<math>G, h</math>)</p> <p>Input :</p> <ul style="list-style-type: none"> <li>- <math>G</math> : undirected graph (<math>V, E</math>)</li> <li>- <math>h</math> : hop count</li> </ul> <p>Output :</p> <ul style="list-style-type: none"> <li>- <math>m</math> : value of graph median</li> <li>- <math>v_{median}</math> : vertex having graph median</li> </ul> <pre> 1: for each <math>v \in V</math> do 2:   enumerate <math>vScore(v, h)</math>; 3: Sort the vertices <math>V</math> by their <math>vScore</math> 4: for each <math>v \in V</math> do 5:   enumerate <math>s</math> as sum of shortest paths      from <math>v</math> to all vertices using lower bound 6:   if <math>s</math> is minimum sum of shortest path 7:     <math>v_{median} = v</math>; 8:     <math>m = s</math>; 9: return <math>v_{median}, m</math>;                 </pre>
---

알고리즘 1. 그래프 구조 정보를 이용한 그래프 메디안 기법

### 4. 성능 평가

실험을 위하여 가상 그래프 생성기[]를 통하여 에지 가중치가 있는 비방향성 연결 그래프를 생성하였다. 실험간 사용한 그래프 정보는 표1과 같다. 1부터 10까지 균등분포로 난수를 생성하여 에지의 가중치에 할당하였다.

<표 1> 실험간 사용한 데이터 셋

데이터 셋	버텍스 수	에지 수
R011	5,000	6,000
R012	5,000	10,000
R021	10,000	12,000
R022	10,000	20,000

그림 2는 제안하는 기법(GTI-Median)과 기존의 하한 값을 이용한 메디안 기법(LB)의 수행시간을 비교한 실험 결과이다. GTI-Median의 반경은 3으로 설정하였다. 제안하는 기법은 모든 데이터 셋에서 기존의 기법보다 약 5~10%정도 수행시간이 향상되었다.

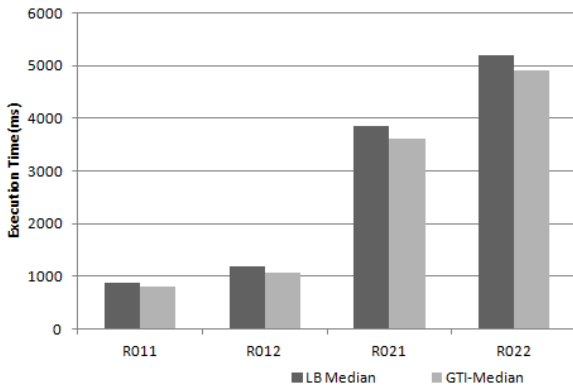


그림 2. 그래프 메디안 기법의 수행 속도 비교

## 5. 결론

그래프 구조 정보를 이용한 그래프 메디안 탐색 기법을 제안하였다. 하한 값을 이용한 그래프 메디안 탐색 기법에 적용하여 제안한 알고리즘이 기존의 알고리즘보다 최대 10%까지 성능이 향상됨을 보였다. 제안하는 방법은 벡터의 그래프 구조 정보를 이용하여 그래프 메디안 예측 값을 구하고 그래프 메디안과 유사한 벡터들부터 탐색을 우선적으로 시작하여 그래프 메디안을 계산하기 위한 많은 비용을 개선하였다.

\* 이 논문은 2013년도 정부(교육과학기술부)의 재원으로 한국연구재단의 지원을 받아 수행된 연구임(No. 2012R1A2A2A01047478).

## 참고문헌

- [1] Bruell S.C, Ghosh S, Karaata M.H. and Pemmaraju S.V, "Self-stabilizing algorithms for finding centers and medians of trees," SIAM J. Comput, 29(2), pp600 - 614, 2009.
- [2] 윤영선, 김태연, 홍지혜, 이영구, "하한 조건을 이용한 프리젤 기반의 효율적인 메디안 탐색 기법," 한국컴퓨터종합학술대회, pp277-279, 2013.
- [3] Hang T. L, *A java library of graph algorithms and optimization*, 2007