



(19) 대한민국특허청(KR)
(12) 등록특허공보(B1)

(45) 공고일자 2017년08월24일
(11) 등록번호 10-1771352
(24) 등록일자 2017년08월18일

(51) 국제특허분류(Int. Cl.)
G06F 17/30 (2006.01) G06Q 50/00 (2006.01)
(52) CPC특허분류
G06F 17/30958 (2013.01)
G06F 17/30991 (2013.01)
(21) 출원번호 10-2016-0038036
(22) 출원일자 2016년03월30일
심사청구일자 2016년03월30일
(56) 선행기술조사문헌
JP2004272369 A*
KR1020150125511 A*
*는 심사관에 의하여 인용된 문헌

(73) 특허권자
경희대학교 산학협력단
경기도 용인시 기흥구 덕영대로 1732 (서천동, 경희대학교 국제캠퍼스내)
(72) 발명자
이영구
대전광역시 서구 청사서로 11, 106동 302호 (월평동, 무지개아파트)
서호진
경기도 수원시 영통구 청명로 132, 325동 1503호 (영통동, 벽산삼익아파트)
(74) 대리인
(뒷면에 계속)
인비전 특허법인

전체 청구항 수 : 총 14 항

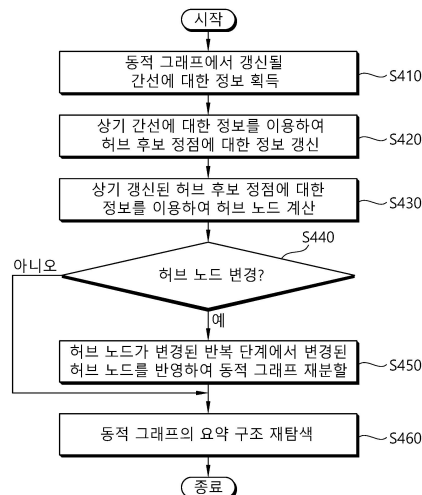
심사관 : 홍경아

(54) 발명의 명칭 동적 환경에서의 그래프 요약 방법 및 장치

(57) 요약

동적 환경에서의 그래프 요약 방법 및 장치를 제공한다. 그래프 요약 장치에 의한 그래프 요약 방법은 동적 그래프에서 갱신될 간선에 대한 정보를 획득하는 단계, 상기 동적 그래프의 허브 후보 정점에 대한 정보를 이용하여 상기 간선으로 인해 상기 동적 그래프에서 적어도 하나의 허브 노드가 변경되는지를 확인하는 단계 및 상기 적어도 하나의 허브 노드가 변경되는 경우, 해당 허브 노드를 기준으로 상기 동적 그래프를 재분할하는 단계를 포함하고, 상기 허브 노드는 상기 동적 그래프에 포함된 정점들 중에서 기 설정된 수 이상의 간선을 가지는 정점일 수 있다.

대표도 - 도4



(52) CPC특허분류

G06Q 50/01 (2013.01)

(72) 발명자

김현욱

경기도 용인시 기흥구 동탄기흥로 864, 105동 150
2호 (공세동, 성원상떼레이크뷰)

박기성

경기도 부천시 원미구 원미로 123-4, E동 4호 (원
미동, 백송연립)

이 발명을 지원한 국가연구개발사업

과제고유번호 2015R1A2A2A01008209

부처명 미래창조과학부

연구관리전문기관 한국연구재단

연구사업명 중견연구자지원사업(핵심)

연구과제명 요약 및 압축 그래프 기반의 빅 그래프 처리 플랫폼

기 여 율 1/1

주관기관 경희대학교(국제캠퍼스)

연구기간 2015.05.01 ~ 2016.04.30

명세서

청구범위

청구항 1

그래프 요약 장치에 의한 그래프 요약 방법에 있어서,

동적 그래프에서 허브 후보 정점에 대한 정보를 탐색하고 탐색된 정보를 유지하는 단계;

상기 동적 그래프에서 갱신될 간선에 대한 정보를 획득하는 단계;

상기 동적 그래프의 허브 후보 정점에 대한 정보를 이용하여 상기 간선으로 인해 상기 동적 그래프에서 적어도 하나의 허브 노드가 변경되는지를 확인하는 단계; 및

상기 적어도 하나의 허브 노드가 변경되는 경우, 해당 허브 노드를 기준으로 상기 동적 그래프를 재분할하는 단계

를 포함하고,

상기 허브 노드는,

상기 동적 그래프에 포함된 정점들 중에서 기 설정된 수 이상의 간선을 가지는 정점이고,

상기 허브 후보 정점은,

상기 갱신될 간선으로 인해 허브 노드의 집합에 속할 가능성이 있는 정점인 것을 특징으로 하는 그래프 요약 방법.

청구항 2

제1항에 있어서,

상기 동적 그래프는,

기 요약된 그래프인 것을 특징으로 하는 그래프 요약 방법.

청구항 3

제1항에 있어서,

상기 확인하는 단계는,

상기 동적 그래프에서 기 설정된 개수 이상의 간선이 갱신되는 경우 수행되는 것을 특징으로 하는 그래프 요약 방법.

청구항 4

제1항에 있어서,

상기 확인하는 단계는,

상기 동적 그래프에서 갱신될 간선에 대한 정보를 이용하여 상기 허브 후보 정점에 대한 정보를 갱신하는 단계; 및

상기 갱신된 허브 후보 정점에 대한 정보를 이용하여 상기 허브 노드를 계산하는 단계를 포함하는 것을 특징으로 하는 그래프 요약 방법.

청구항 5

제1항에 있어서,

상기 재분할하는 단계는,

특정 반복 단계에서 상기 허브 노드가 변경되는 경우, 상기 특정 반복 단계에서 가장 많은 정점을 포함하는 부분 그래프를 대상으로 수행되고,

상기 반복 단계는,

상기 가장 많은 정점을 포함하는 부분 그래프에서 상기 가장 많은 정점을 포함하는 부분 그래프의 부분 그래프와 허브 노드를 탐색하는 단계인 것을 특징으로 하는 그래프 요약 방법.

청구항 6

제5항에 있어서,

상기 재분할하는 단계 이후에,

상기 가장 많은 정점을 포함하는 부분 그래프의 부분 그래프에 상기 갱신될 간선에 대한 정보를 적용하여 상기 동적 그래프의 요약 구조를 재탐색하는 단계를 더 포함하는 것을 특징으로 하는 그래프 요약 방법.

청구항 7

제6항에 있어서,

상기 재탐색하는 단계는,

상기 동적 그래프의 요약 구조 중 상기 갱신될 간선에 연관되는 요약 구조를 탐색하는 단계; 및

상기 간선의 갱신으로 인해 상기 탐색된 요약 구조에 변함이 있는지를 확인하는 단계를 포함하는 것을 특징으로 하는 그래프 요약 방법.

청구항 8

동적 그래프에서 갱신될 간선에 대한 정보를 입력받는 입력 인터페이스; 및

상기 동적 그래프에서 허브 후보 정점에 대한 정보를 탐색하여 탐색된 정보를 유지하고, 상기 허브 후보 정점에 대한 정보를 이용하여 상기 간선으로 인해 상기 동적 그래프에서 적어도 하나의 허브 노드가 변경되는지를 확인하고, 상기 적어도 하나의 허브 노드가 변경되는 경우 해당 허브 노드를 기준으로 상기 동적 그래프를 재분할하는 프로세서

를 포함하고,

상기 허브 노드는,

상기 동적 그래프에 포함된 정점들 중에서 기 설정된 수 이상의 간선을 가지는 정점이고,

상기 허브 후보 정점은,

상기 갱신될 간선으로 인해 허브 노드의 집합에 속할 가능성이 있는 정점인 것을 특징으로 하는 그래프 요약 장치.

청구항 9

제8항에 있어서,

상기 동적 그래프는,

이전 주기에 상기 프로세서에 의해 요약된 그래프인 것을 특징으로 하는 그래프 요약 장치.

청구항 10

제8항에 있어서,

상기 프로세서는,

상기 동적 그래프에서 기 설정된 개수 이상의 간선이 갱신되는 경우 상기 간선으로 인해 상기 동적 그래프에서 적어도 하나의 허브 노드가 변경되는지를 확인하는 것을 특징으로 하는 그래프 요약 장치.

청구항 11

제8항에 있어서,

상기 프로세서는,

상기 동적 그래프에서 갱신될 간선에 대한 정보를 이용하여 상기 허브 후보 정점에 대한 정보를 갱신하고, 상기 갱신된 허브 후보 정점에 대한 정보를 이용하여 상기 허브 노드를 계산하는 것을 특징으로 하는 그래프 요약 장치.

청구항 12

제8항에 있어서,

상기 프로세서는,

특정 반복 단계에서 상기 허브 노드가 변경되는 경우, 상기 특정 반복 단계에서 가장 많은 정점을 포함하는 부분 그래프를 대상으로 그래프 분할을 수행하되,

여기서 상기 반복 단계는,

상기 가장 많은 정점을 포함하는 부분 그래프에서 상기 가장 많은 정점을 포함하는 부분 그래프의 부분 그래프와 허브 노드를 탐색하는 단계인 것을 특징으로 하는 그래프 요약 장치.

청구항 13

제12항에 있어서,

상기 프로세서는,

상기 가장 많은 정점을 포함하는 부분 그래프의 부분 그래프에 상기 갱신될 간선에 대한 정보를 적용하여 상기 동적 그래프의 요약 구조를 재탐색하는 것을 특징으로 하는 그래프 요약 장치.

청구항 14

제13항에 있어서,

상기 프로세서는,

상기 동적 그래프의 요약 구조 중 상기 갱신될 간선에 연관되는 요약 구조를 탐색하고 상기 간선의 갱신으로 인해 상기 탐색된 요약 구조에 변함이 있는지를 확인하여 상기 요약 구조에 변함이 있는 경우 상기 동적 그래프의 요약 구조를 재탐색하는 것을 특징으로 하는 그래프 요약 장치.

발명의 설명

기술 분야

[0001] 본 발명은 동적 환경에서 효율적으로 그래프를 요약하는 방법 및 이를 이용하는 장치에 관한 것이다.

배경 기술

[0002] 최근 정보화 기기의 발달로 인해 소셜 네트워크(Social Network)를 이용하는 사용자가 증가하고 있다. 그런데, DBLP(Digital Bibliography & Library Project) 네트워크, 소셜 네트워크 등과 같은 실세계의 네트워크를 반영하는 동적 환경의 그래프는 시간이 지남에 따라 노드(node) 및 에지(edge)의 정보가 다양하게 변화한다. 나아가, 소셜 네트워크 등에서의 노드와 에지의 양이 방대해짐에 따라 네트워크 전체가 모두 표현된 원래 그래프(original graph) 이외에도, 대용량 그래프 데이터 분석을 위한 요약된 그래프(summarized graph)가 필요하게 되었다. 그래프 요약은 대용량 그래프를 밀집도가 높은 부분그래프로 압축 표현하는 기법이다. 따라서 그래프 요약은 그래프 압축(graph compression)이라 불릴 수도 있다.

[0003] 그래프 요약 기법 중 하나로서, 한국공개특허공보 제10-2015-0125511호 (공개일: 2015년 11월 9일) "그래프 압축 처리 방법 및 장치"에는 그래프 상에 존재하는 노드의 차수를 이용하여 허브 노드를 추출하고, 허브 노드 간의 연결을 고려하여 차수가 높은 순으로 허브 노드를 제거하고, 제거한 허브 노드로부터 분해되는 스포크

(spoke) 노드를 식별하고, 제거한 허브 노드의 차수를 고려하여 그래프 상에 상기 스포크 노드를 배열하고, 상기 스포크 노드의 배열에 따른 그래프의 압축을 처리하는 것이 기재되어 있다. 이와 같이 기존의 그래프 요약 기법은 차수가 높은 허브 노드를 기준으로 그래프를 반복적으로 다수의 부분그래프들로 분할하고, 분할된 부분 그래프들을 각각 요약한다. 그러나, 사용자 간의 관계는 실시간으로 변화하기 때문에 소셜 네트워크도 동적으로 변경되고, 그에 따라 이를 나타내는 그래프 또한 동적으로 변경된다. 하지만 기존의 그래프 요약 기법은 이러한 동적 환경을 고려하지 않아 그래프가 변경될 때 마다 전체 그래프 즉, 요약되지 않은 원래의 그래프를 대상으로 다시 요약 구조를 탐색해야 했기 때문에, 변경된 정보를 반영하는데 매우 오랜 수행시간이 필요하다.

발명의 내용

해결하려는 과제

- [0004] 본 발명의 기술적 과제는 동적으로 변화하는 그래프를 보다 빠르고 효율적으로 요약할 수 있는 동적 환경에서의 그래프 요약 방법을 제공함에 있다.
- [0005] 본 발명의 다른 기술적 과제는 동적으로 변화하는 그래프를 보다 빠르고 효율적으로 요약할 수 있는 그래프 요약 장치를 제공함에 있다.

과제의 해결 수단

- [0006] 본 발명의 일 양태에 따르면, 그래프 요약 장치에 의한 그래프 요약 방법은 동적 그래프에서 갱신될 간선에 대한 정보를 획득하는 단계, 상기 동적 그래프의 허브 후보 정점에 대한 정보를 이용하여 상기 간선으로 인해 상기 동적 그래프에서 적어도 하나의 허브 노드가 변경되는지를 확인하는 단계 및 상기 적어도 하나의 허브 노드가 변경되는 경우, 해당 허브 노드를 기준으로 상기 동적 그래프를 재분할하는 단계를 포함하고, 상기 허브 노드는 상기 동적 그래프에 포함된 정점들 중에서 기 설정된 수 이상의 간선을 가지는 정점일 수 있다.
- [0007] 일측에 따르면, 상기 동적 그래프는 기 요약된 그래프일 수 있다.
- [0008] 다른 측면에 따르면, 상기 확인하는 단계는 상기 동적 그래프에서 기 설정된 개수 이상의 간선이 갱신되는 경우 수행될 수 있다.
- [0009] 또 다른 측면에 따르면, 상기 확인하는 단계는 상기 동적 그래프에서 갱신될 간선에 대한 정보를 이용하여 상기 허브 후보 정점에 대한 정보를 갱신하는 단계 및 상기 갱신된 허브 후보 정점에 대한 정보를 이용하여 상기 허브 노드를 계산하는 단계를 포함할 수 있다.
- [0010] 또 다른 측면에 따르면, 상기 재분할하는 단계는 특정 반복 단계에서 상기 허브 노드가 변경되는 경우, 상기 특정 반복 단계에서 가장 많은 정점을 포함하는 부분 그래프를 대상으로 수행되고, 상기 반복 단계는 상기 가장 많은 정점을 포함하는 부분 그래프에서 상기 가장 많은 정점을 포함하는 부분 그래프의 부분 그래프와 허브 노드를 탐색하는 단계일 수 있다.
- [0011] 또 다른 측면에 따르면, 상기 재분할하는 단계 이후에 상기 가장 많은 정점을 포함하는 부분 그래프의 부분 그래프에 상기 갱신된 간선에 대한 정보를 적용하여 상기 동적 그래프의 요약 구조를 재탐색하는 단계를 더 포함할 수 있다.
- [0012] 또 다른 측면에 따르면, 상기 재탐색하는 단계는 상기 동적 그래프의 요약 구조 중 상기 갱신된 간선에 연관되는 요약 구조를 탐색하는 단계 및 상기 간선의 갱신으로 인해 상기 탐색된 요약 구조에 변함이 있는지를 확인하는 단계를 포함할 수 있다.
- [0013] 본 발명의 다른 양태에 따르면, 그래프 요약 장치는 동적 그래프에서 갱신될 간선에 대한 정보를 입력받는 입력 인터페이스 및 상기 동적 그래프의 허브 후보 정점에 대한 정보를 이용하여 상기 간선으로 인해 상기 동적 그래프에서 적어도 하나의 허브 노드가 변경되는지를 확인하고, 상기 적어도 하나의 허브 노드가 변경되는 경우 해당 허브 노드를 기준으로 상기 동적 그래프를 재분할하는 프로세서를 포함하고, 상기 허브 노드는 상기 동적 그래프에 포함된 정점들 중에서 기 설정된 수 이상의 간선을 가지는 정점일 수 있다.

발명의 효과

- [0014] 본 발명에 따르면 종래의 그래프 요약 방법보다 그래프의 요약 속도 및 효율이 증대된다.

도면의 간단한 설명

- [0015] 도 1는 그래프의 대표적인 요약 구조를 나타내는 예시도이다.
- 도 2는 그래프 요약에서 허브 노드(Hub node), GCC(Giant Connected Component) 및 스포크(Spoke)의 개념을 설명하기 위한 예시도이다.
- 도 3은 동적 그래프에서 시간 주기가 변화됨에 따라 허브 집합, GCC 및 스포크가 변경됨을 나타내는 예시도이다.
- 도 4는 본 발명의 일 실시예에 따른 그래프 요약 방법을 나타내는 흐름도이다.
- 도 5는 본 발명의 일 실시예에 따른 동적 그래프 요약 알고리즘을 나타내는 예시도이다.
- 도 6은 본 발명의 일 실시예에 따른 요약 구조 갱신 알고리즘을 나타내는 예시도이다.
- 도 7은 일 실시예에 따른 그래프 요약 장치를 나타내는 블록도이다.

발명을 실시하기 위한 구체적인 내용

- [0016] 아래에서는 첨부한 도면을 참고로 하여 본 발명의 실시예에 대하여 본 발명이 속하는 기술 분야에서 통상의 지식을 가진 자가 용이하게 실시할 수 있도록 상세히 설명한다. 그러나 본 발명은 여러 가지 상이한 형태로 구현될 수 있으며 여기에서 설명하는 실시예에 한정되지 않는다. 그리고 도면에서 본 발명을 명확하게 설명하기 위해서 설명과 관계없는 부분은 생략하였으며, 명세서 전체를 통하여 유사한 부분에 대해서는 유사한 도면 부호를 붙였다.
- [0017] 명세서 전체에서, 어떤 부분이 어떤 구성요소를 "포함"한다고 할 때, 이는 특별히 반대되는 기재가 없는 한 다른 구성요소를 제외하는 것이 아니라 다른 구성요소를 더 포함할 수 있는 것을 의미한다. 또한, 명세서에 기재된 "~부" 등의 용어는 적어도 하나의 기능이나 동작을 처리하는 단위를 의미하며, 이는 하드웨어나 소프트웨어 또는 하드웨어 및 소프트웨어의 결합으로 구현될 수 있다.
- [0018] 도 1는 그래프의 대표적인 요약 구조를 나타내는 예시도이고, 도 2는 그래프 요약에서 허브 노드(Hub node), GCC(Giant Connected Component) 및 스포크(Spoke)의 개념을 설명하기 위한 예시도이다.
- [0019] 그래프 $G=(V, E)$ 는 정점(vertex)의 집합 V 과 간선(edge)의 집합 E 로 구성된다. 그래프 G 에서 각각의 간선은 두 개의 정점을 연결한다.
- [0020] 그래프 요약이란 그래프 G 로부터 연결성, 정도(degree) 비대칭성과 같은 그래프의 구조적인 특징들을 나타내는 요약 구조 $S=\{s_1, s_2, \dots, s_m\}$ 을 찾는 것이다. 그래프의 대표적인 요약 구조로는 도 1에 도시된 것과 같이 스타(star) 구조, 클리크(clique) 구조, 이분 그래프(bipartite graph) 구조 및 체인(chain) 구조가 있다.
- [0021] 그래프 요약 기법 중 하나로 VoG 기법은 그래프를 반복적으로 분할하여 요약할 그래프들을 탐색하는 기법이다. 구체적으로, VoG 기법은 그래프 내에서 허브 노드의 집합(이하, 허브 집합이라 함)을 탐색하고, 허브 집합 내의 정점의 모든 간선을 제거하여 그래프를 다수의 부분 그래프($sg_1, sg_2, \dots, sg_i$)로 분할한다. 여기서, 허브 집합(hub set)은 그래프 G 내의 정점들의 집합 V 에서 정도(degree)가 높은 상위 k 개의 정점들의 집합을 의미할 수 있다. 또는, 허브 집합은 그래프 G 내의 정점들 중에서 기 설정된 수 이상의 간선을 가지는 정점들의 집합으로 정의될 수 있다.
- [0022] 한편 GCC(Giant Connected Component)는 그래프 G 에 포함되는 부분 그래프들 중 가장 많은 수의 정점을 포함하는 부분그래프로 정의될 수 있다. 그래프 G 에 포함되는 부분 그래프들 중 GCC를 제외한 나머지 부분 그래프들은 스포크(Spoke)라 정의될 수 있다. VoG 기법은 스포크 집합과 허브 집합에 해당하는 정점들로부터 대표적인 요약 구조를 탐색 및 압축률을 계산하고, GCC를 대상으로 GCC내의 정점의 수가 k 개보다 작아질 때까지 이와 같은 방법을 반복 수행한다.
- [0023] GCC(가장 많은 정점을 포함하는 부분 그래프)에서 스포크(또는 스포크 집합)과 허브 노드(또는 허브 집합)을 탐색하는 과정은 반복 단계(iteration level)라 정의될 수 있다. i 번째 반복 단계의 GCC는 GCC^i 로 정의될 수 있고, i 번째 반복 단계의 스포크 집합은 $SP^i=\{sp_1, sp_2, \dots, sp_l\}$ 로 정의될 수 있다.

- [0024] 도 2에는 일 예로 11개의 정점들(1번 정점 내지 11번 정점)을 포함하는 그래프가 도시되어 있다. 그래프 요약 을 위해 k 를 2로 설정하여 요약 구조를 탐색하는 경우, 첫 번째 반복 단계에서는 11번 정점과 1번 정점의 정도 (간선)가 '6'으로 전체 정점들 중에서 가장 높기 때문에 허브 노드로서 11번 정점과 1번 정점이 선정된다. 이 후, 각 허브 노드와 연결된 간선들을 모두 제거하여 그래프 G 를 제1 부분 그래프(sg_1), 제2 부분 그래프(sg_2), 제3 부분 그래프(sg_3) 및 제4 부분 그래프(sg_4)로 분할한다. 이 경우, 제4 부분 그래프(sg_4)가 가장 많은 정점 들을 포함하므로 GCC가 되며, 그 외의 부분 그래프(제1 내지 제3 부분 그래프)는 스포크가 된다. 이후, 제1 내 지 제3 부분 그래프를 포함하는 스포크 집합(spoke set)에서 각 스포크에 대응하는 요약 구조가 존재하는지를 탐색하고, MDL(Minimum Description Length) 원리에 의해 압축률이 가장 좋은 구조로 요약한다. 두 번째 반복 단계에서는 제4 부분 그래프(sg_4)를 대상으로 상술한 과정(허브 노드 선정, 그래프 분할 등)을 반복한다.
- [0025] 도 3은 동적 그래프에서 시간 주기가 변화됨에 따라 허브 집합, GCC 및 스포크가 변경됨을 나타내는 예시도이다.
- [0026] 동적 그래프 G_{T_i} 는 시간 주기 T_i 에 출현한(또는 요약된) 그래프로 정의될 수 있다. 동적 그래프 G_{T_i} 는 이전 주기 T_{i-1} 의 동적 그래프 $G_{T_{i-1}}$ 및 이전 주기 T_{i-1} 부터 갱신(추가, 삭제 및/또는 변경)된 간선의 집합 ΔE_{T_i} 로 구성된다. 여기서, 상기 시간 주기 T_i 는 기 설정된 수의 간선이 변경되었을 때 다음 주기 T_{i+1} 로 변 경된다.
- [0027] 동적 그래프에서 그래프를 요약하는 방법은 시간 주기 $T=\{T_1, T_2, \dots, T_n\}$ 에 대한 동적 그래프 $G_T=\{G_{T_1}, G_{T_2}, \dots, G_{T_q}\}$ 가 존재할 때, 각각의 G_{T_i} 에 대한 요약 구조 $S_{T_i}=\{s_1, s_2, \dots, s_m\}$ 를 찾는 것이다.
- [0028] T_i 주기의 동적 그래프 G_{T_i} 를 효율적으로 요약하기 위한 가장 간단한 방법은 T_{i-1} 주기의 동적 그래프 $G_{T_{i-1}}$ 에서 탐색한 요약 구조 $S_{T_{i-1}}$ 와 변경된 간선 집합 ΔE_{T_i} 를 고려하여 S_{T_i} 를 점진적으로(Incremental) 계산 하는 것이다. 그러나 $G_{T_{i-1}}$ 와 $S_{T_{i-1}}$ 에서 변경된 간선 집합 ΔE_{T_i} 를 적용하여 S_{T_i} 를 점진적으로 계산할 때, 모든 요약 구조를 정확히 탐색하기 위해서는 다음의 두 가지 상황을 고려해야 한다.
- [0029] 첫째, ΔE_{T_i} 로 인해 허브 집합이 변경되어 그래프가 분할되는 기준이 변경되는 경우가 발생한다. 그래프가 이 전 주기와 다른 형태로 분할된다면, 이전주기에 탐색한 요약 구조들 중 변경되는 요약 구조들이 발생한다.
- [0030] 둘째, ΔE_{T_i} 로 인해 스포크의 분할과 병합이 발생한다. 예를 들어, 간선이 추가되는 경우, 두 개의 스포크(두 개의 부분 그래프)가 상기 추가되는 간선에 의해 서로 연결되면 상기 두 개의 스포크는 하나의 스포크로 병합된 다. 하나의 스포크 내에서 간선이 삭제되는 경우 이는 두 개의 스포크로 분할될 수 있다.
- [0031] 일 예로 도 3을 참조하면, k 를 2로 설정하였을 때 T_1 주기에는 1번 정점과 10번 정점이 차수(간선의 수)가 각각 '4', '5'로 가장 높아 허브 노드로 선택되었으나, T_2 주기에는 4번 정점에 5개의 간선이 추가됨에 따라 허브 노 드는 4번 정점 및 10번 정점으로 변경된다. 도 3에서 T_2 주기에 추가되는 간선은 점선으로 도시되어 있다.
- [0032] 이러한 허브 노드의 변경으로 인해 그래프를 분할하는 조건이 변경되면서 T_2 주기에는 GCC 및 스포크가 변경된다. 구체적으로 T_1 주기에는 14번 정점, 15번 정점, 16번 정점 및 17번 정점을 포함하는 부분 그래프가 G_{T_1} 내의 부분 그래프들 중 가장 많은 정점을 포함하므로 GCC로 선정되었으나, T_2 주기에는 1번 정점, 2번 정점, 3번 정점, 5번 정점, 6번 정점, 7번 정점, 8번 정점 및 9번 정점을 포함하는 부분 그래프가 G_{T_2} 내의 부분 그래 프들 중 가장 많은 정점을 포함하므로 GCC로 선정된다. 또한 T_2 주기에 간선 (v_{12}, v_{13}) 과 (v_{13}, v_{14}) 가 추가됨에 따라 14번 정점, 15번 정점, 16번 정점 및 17번 정점을 포함하는 스포크, 12번 정점을 포함하는 스포크 및 13번

정점을 포함하는 스포크가 서로 병합된다.

[0033] 도 4는 본 발명의 일 실시예에 따른 그래프 요약 방법을 나타내는 흐름도이다.

[0034] T_{i-1} 주기의 j 번째 반복 단계에서 ΔE_{T_i} 에 의해 허브 집합(또는 허브 노드)이 변경되는지 여부는 모든 반복 단계에서 그래프 $G_{T_{i-1}}$ 내의 정점을 대상으로 허브 집합을 다시 계산함으로써 검사할 수 있다. 그러나 이와 같은 방법은 그래프를 구성하는 정점의 수가 많고 반복 단계가 많아질수록 더욱 오랜 수행시간을 요구하게 된다. 따라서, 본 발명에 따른 그래프 요약 방법은 각 반복 단계마다 허브 후보 정점에 대한 정보를 탐색하고 탐색된 정보를 유지할 수 있다. 상기 허브 후보 정점에 대한 정보는 허브 후보 정점 집합의 형태로 유지될 수 있다.

[0035] 허브 후보 정점은 T_i 주기에 변경된 간선 집합 ΔE_{T_i} 와 허브 집합 $H=\{h_1, h_2, \dots, h_k\}$ 가 존재할 때, 정점 집합 V 에서 허브 노드를 제외한 정점들 중에서 정도가 $\deg(\arg\min_{h \in H} \deg(h)) - u$ 보다 큰 정점으로 정의될 수 있다. 여기서 $\deg(v)$ 는 정점 v 에 대한 정도(정점 v 가 가지는 간선의 수)를 나타내고, u 는 해당 주기에 갱신(또는 추가)되는 간선의 수를 나타낸다. 즉, 허브 후보 정점은 간선의 변경으로 인해 허브 집합에 속할 가능성이 있는 정점을 의미한다. 따라서 각 반복 단계마다 허브 후보 정점 집합을 탐색하여 유지하면, 모든 정점에 대한 탐색 없이 허브 후보 정점이 허브 노드가 되는지만 검색하면 된다.

[0036] 한편, 본 발명에 따른 그래프 요약 방법은 다음의 정리 1을 기반으로 그래프를 요약한다.

[0037] 정리 1: 반복 단계 j 에서 ΔE_{T_i} 로 인해 H^j 가 H^i 로 변경되었을 때, $H^1, H^2, \dots, H^{j-1}$ 은 변경되지 않는다.

[0038] 증명: 매 반복 단계마다 가장 차수가 높은 정점들(가장 많은 간선을 가지는 정점들)을 허브 집합으로 선택하였기 때문에 모든 $p(0 < p < j-1)$ 에 대해 아래 수학적 1을 만족한다. 따라서 H^j 가 변경되어도 $H^1, H^2, \dots, H^{j-1}$ 은 변경되지 않는다.

수학적 1

$$\max_{v \in H^j} \deg(v) \leq \min_{v' \in H^{j-p}} \deg(v')$$

[0039]

[0040] 따라서, 본 발명에 따른 그래프 요약 방법은 정리 1에 의해 현재 주기에서 허브 노드가 갱신되는 경우, 이전 주기의 해당 반복 단계(허브 노드가 변경되는 반복 단계)에서부터 요약 구조를 재탐색하더라도 매 주기마다 그래프를 재요약하여 요약 구조를 재계산하는 방법과 동일한 결과 즉, 동일한 정확도를 가진다.

[0041] 일 실시예로 도 4를 참조하면, 본 발명에 따른 그래프 요약 장치는 동적 그래프에서 갱신될 간선에 대한 정보를 획득하면(S410), 이전 주기의 동적 그래프 $G_{T_{i-1}}$ 와 해당 요약 구조 $S_{T_{i-1}}$ 로부터 현재 주기의 요약 구조 S_{T_i} 를 탐색하기 위하여 $H^1_{T_{i-1}}$ 부터 $H^l_{T_{i-1}}$ 까지 탐색할 수 있다. 여기서, l 은 총 반복 단계의 수를 나타낸다. 구체적으로,

로, 그래프 요약 장치는 이전 주기에 요약된 동적 그래프에서 간선이 갱신되는 경우, 동적 그래프 $G_{T_{i-1}}$ 의 허브 후보 정점에 대한 정보를 이용하여 간선의 갱신으로 인해 해당 동적 그래프에서 적어도 하나의 허브 노드가 변경되는지를 확인할 수 있다. 이를 위하여 그래프 요약 장치는 이전 주기의 동적 그래프에서 갱신될 간선에 대한 정보를 이용하여 상기 허브 후보 정점에 대한 정보를 갱신하고(S420), 갱신된 허브 후보 정점에 대한 정보를 이용하여 현재 주기에서 각 반복 단계의 허브 노드를 계산할 수 있다(S430). 이러한 과정은 동적 그래프에서 기 설정된 개수 이상의 간선이 갱신되는 경우 즉, 기 설정된 개수 이상의 간선에 대한 정보가 획득되는 경우 수행될 수 있다.

[0042] 그래프 요약 장치는 상기 계산 결과를 기반으로 이전 주기의 동적 그래프에서 허브 노드가 변경되는지를 판단할 수 있다(S440). 만약 적어도 하나의 허브 노드가 변경되는 것으로 판단되는 경우, 그래프 요약 장치는 해당 반복 단계에서 해당 허브 노드(또는 허브 집합)를 기준으로 상기 동적 그래프를 재분할할 수 있다(S450). 일 예

로 그래프 요약 장치는 이전 주기 T_{i-1} 의 j 번째 반복 단계에서 $H_{T_i}^j$ 가 변경되는 경우, 정리 1을 기반으로 이전 주기의 j 번째 반복 단계의 그래프 $G_{T_i}^j$ (또는 GCC)를 대상으로 슬래시번(SlashBurn) 알고리즘을 수행하여 해당 그래프를 재분할할 수 있다. 슬래시번 알고리즘은 그래프 내에서 허브 집합을 탐색하고, 탐색된 허브 집합을 기준으로 GCC를 탐색하여 분할하는 알고리즘이다. 슬래시번 알고리즘은 분할된 그래프(부분 그래프)에 포함되는 정점의 수가 k 보다 작아질 때까지 수행될 수 있다.

[0043] 이후, 그래프 요약 장치는 해당 GCC(가장 많은 정점을 포함하는 부분 그래프) 내의 부분 그래프(스포크)에 상기 갱신될 간선에 대한 정보를 적용하여 동적 그래프의 요약 구조를 재탐색할 수 있다(S460). 이를 위하여 그래프 요약 장치는 동적 그래프의 요약 구조 중 상기 갱신될 간선에 연관되는 요약 구조를 탐색하고, 상기 간선의 갱신으로 인해 상기 탐색된 요약 구조에 변함이 있는지를 확인할 수 있다. 예를 들어, 그래프 요약 장치는 스포크 집합 $SP_{T_i}^1$ 부터 $SP_{T_i}^j$ 내의 요약 구조들에 ΔE_{T_i} 를 적용하여 점진적으로 갱신할 수 있다. 이때, 간선의 추가로 인해 스포크가 분할되거나 간선의 삭제로 인해 둘 이상의 스포크가 병합되는 경우가 발생할 수 있다. 이러한 스포크의 병합과 분할을 효율적으로 계산하기 위해 점진적인 연결 컴포넌트 레이블링(Incremental Connected Component Labeling) 기술을 사용할 수 있다. 점진적인 연결 컴포넌트 레이블링 기술은 그래프 내에서 서로 연결되지 않은 부분그래프들을 서로소(disjoint set)로 구성하고, 간선이 갱신(추가, 삭제 등)될 때 서로소를 효율적으로 갱신하여 전체 그래프의 탐색 없이 연결되지 않은 부분그래프들을 효율적으로 탐색하는 기술이다.

[0044] 도 5는 본 발명의 일 실시예에 따른 동적 그래프 요약 알고리즘을 나타내는 예시도이다.

[0045] 도 5를 참조하면, 동적 그래프 요약 알고리즘은 허브 후보 정점에 대한 정보를 이용하여 각 반복 단계마다 허브 집합을 계산한다. 그리고, T_i 주기와 T_{i-1} 주기의 허브 집합이 변경되었는지 확인하여 허브 집합 정점이 변경된 경우, 허브 집합 H 가 변경되었는지 검사한다(line1-7). 만일 허브 집합 H 가 j 번째 반복 단계에서 변경된 경우, 요약 구조를 다시 계산하기 위해 SlashBurn을 호출한다(line8-9). 마지막으로, 이전 주기의 스포크 집합 $SP_{T_{i-1}}$ 에 ΔE_{T_i} 를 적용하여 점진적으로 T_i 주기의 요약 구조 S_{T_i} 를 계산한다(line10-11).

[0046] 도 6은 본 발명의 일 실시예에 따른 요약 구조 갱신 알고리즘을 나타내는 예시도이다.

[0047] 도 6에는 스포크 집합 $SP_{T_{i-1}}$ 에 ΔE_{T_i} 를 적용하여 점진적으로 요약 구조 S_{T_i} 를 갱신하는 알고리즘의 일 예가 도시되어 있다. 도 6의 요약 구조 갱신 알고리즘은 점진적인 연결 컴포넌트 레이블링 기술을 이용하여 ΔE_{T_i} 를 적용한 스포크 집합 SP_{T_i} 를 계산한다(line 3). 그리고 SP_{T_i} 에서 요약 구조 집합 S'_{T_i} 를 탐색한다(line 4). 이와 같은 과정을 j 번째 반복 단계까지 반복 수행하여 모든 요약 구조 집합 S_{T_i} 를 계산하고 반환한다.

[0048] 도 7은 일 실시예에 따른 그래프 요약 장치를 나타내는 블록도이다.

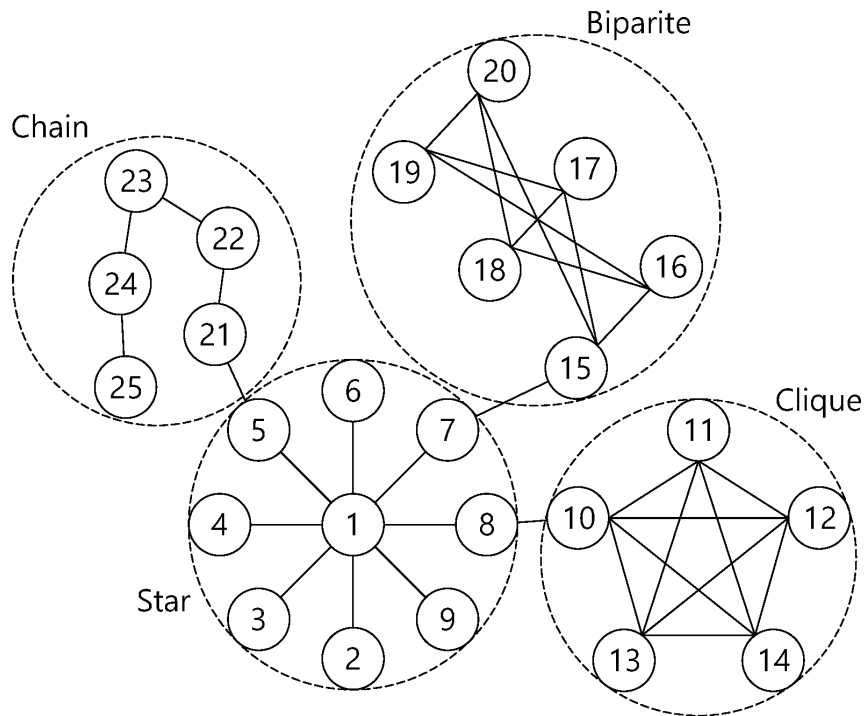
[0049] 도 7을 참조하면, 본 발명에 따른 그래프 요약 장치(700)는 프로세서(710), 저장 장치(720), 출력 인터페이스(730), 입력 인터페이스(740), 통신모듈(750) 중 적어도 하나를 포함할 수 있다. 여기서, 프로세서(710)는 ASIC(application-specific integrated circuit), 다른 칩셋, 논리 회로 및/또는 데이터 처리 장치를 포함할 수 있다. 저장 장치(720)는 디스크 드라이브, ROM(read-only memory), RAM(random access memory), 플래쉬 메모리, 메모리 카드, 저장 매체를 포함할 수 있다. 출력 인터페이스(730)는 프로세서(710)의 처리 결과를 출력하기 위한 인터페이스를 제공한다. 입력 인터페이스(740)는 프로세서(710)의 처리에 필요한 정보의 입력을 위한 인터페이스를 제공한다. 통신 모듈(750)은 유선망 또는 무선망을 통해 데이터를 송수신하기 위하여 RF 안테나 또는 모뎀으로 구성될 수 있으며, 무선 신호를 처리하기 위한 베이스밴드 회로를 포함할 수도 있다.

[0050] 본 발명에 따른 그래프 요약 방법이 소프트웨어로 구현될 때, 본 발명에 따른 그래프 요약 방법 또는 알고리즘들은 상술한 프로세서(710), 저장 장치(720), 출력 인터페이스(730), 입력 인터페이스(740), 통신모듈(750)의 전부 또는 일부의 조합으로서 구현될 수 있다. 본 발명에 따른 동적 그래프 요약 알고리즘 및 요약 구조 갱신 알고리즘은 저장 장치(720)에 저장되고, 프로세서(710)에 의해 실행될 수 있다. 저장 장치(720)는 프로세서(710)의 내부 또는 외부에 있을 수 있고, 잘 알려진 다양한 수단으로 프로세서(710)와 연결될 수 있다.

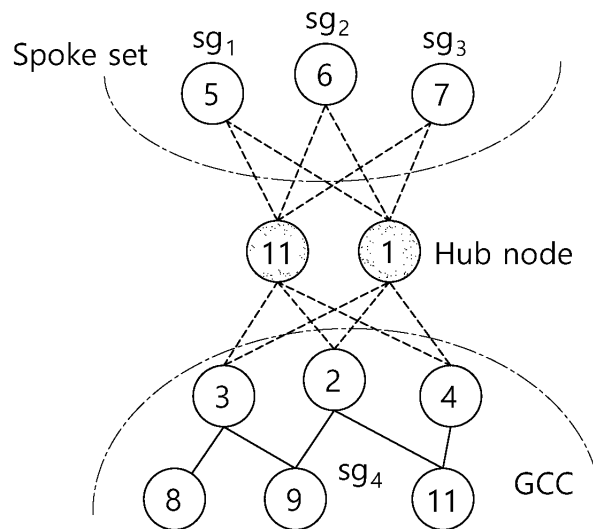
- [0051] 구체적으로, 입력 인터페이스(740)는 동적 그래프에서 갱신될 간선에 대한 정보를 입력 받는다. 여기서 상기 동적 그래프는 이전 주기에 프로세서(710)에 의해 요약된 그래프일 수 있으며, 데이터의 형태로 데이터베이스에 저장될 수 있다. 상기 데이터베이스는 그래프 요약 장치(700)에 포함되거나 별도의 장치로 구현될 수 있다.
- [0052] 프로세서(710)는 동적 그래프의 허브 후보 정점에 대한 정보를 이용하여 상기 갱신될 간선으로 인해 상기 동적 그래프에서 적어도 하나의 허브 노드가 변경되는지를 확인하고, 상기 적어도 하나의 허브 노드가 변경되는 경우 해당 허브 노드를 기준으로 상기 동적 그래프를 재분할하는 한다. 예를 들어, 프로세서(710)는 동적 그래프에서 기 설정된 개수 이상의 간선이 갱신되는 경우 상기 간선으로 인해 상기 동적 그래프에서 적어도 하나의 허브 노드가 변경되는지를 확인할 수 있다. 이를 위하여 프로세서(710)는 동적 그래프에서 갱신될 간선에 대한 정보를 이용하여 허브 후보 정점에 대한 정보를 갱신하고, 상기 갱신된 허브 후보 정점에 대한 정보를 이용하여 허브 노드를 계산할 수 있다.
- [0053] 또한, 프로세서(710)는 특정 반복 단계에서 허브 노드가 변경되는 경우, 특정 반복 단계에서 가장 많은 정점을 포함하는 부분 그래프(GCC)를 대상으로 그래프 분할을 수행할 수 있다. 그리고, 상기 GCC 내의 부분 그래프(스포크)에 상기 갱신될 간선에 대한 정보를 적용(또는 반영)하여 동적 그래프의 요약 구조를 재탐색할 수 있다. 이를 위하여 프로세서(710)는 이전 주기의 요약 구조 중 상기 갱신될 간선에 연관되는 요약 구조를 탐색하고, 상기 간선의 갱신으로 인해 상기 탐색된 요약 구조에 변함이 있는지를 확인할 수 있다. 그리 상기 요약 구조에 변함이 있는 경우 변화된 동적 그래프에 해당하는 요약 구조를 재탐색할 수 있다.
- [0054] 이와 같이 본 발명에 따른 그래프 요약 장치(700)는 새로 추가 및/또는 삭제된 간선으로 인해 이전 주기의 분할된 그래프에서 변경된 부분 그래프를 탐색하므로, 전체 그래프에 대한 재분할 없이 효율적인 그래프 요약을 수행할 수 있다.
- [0055] 본 발명에 따른 그래프 요약 방법은 기존의 기법에 비해 전체 간선 중 10%가 변경되었을 때 최대 57% 수행시간이 향상된다. 이는 전체 그래프를 매 번 입력 받아 분석 및 요약하는 기존 기법에 비해 본 발명에 따른 그래프 요약 방법 보다 효율적으로 그래프를 요약할 수 있음을 나타낸다.
- [0056] 본 발명에 따르면, 빈발히 변경되는 대용량 그래프에서 연결성, 정도 비대칭성 등과 같은 그래프가 갖는 고유한 구조적인 정보를 효율적으로 갱신하고 분석할 수 있으며 압축 표현할 수 있다. 이를 통해, 자신이 속한 소셜 네트워크를 여러 분류로 압축할 수 있고, 해당 소셜 네트워크에서 가장 영향력이 있는 인물을 찾아 낼 수도 있다. 기업에서는 요약된 소셜 네트워크를 분석하여 회사 상품의 시장성, 가치 등을 평가할 수 있으며, 집중적으로 마케팅 해야 하는 연령대 및 성별을 찾아 활용할 수 있다. 이러한 서비스를 제공함으로써 수익을 창출 해낼 수 있다.

도면

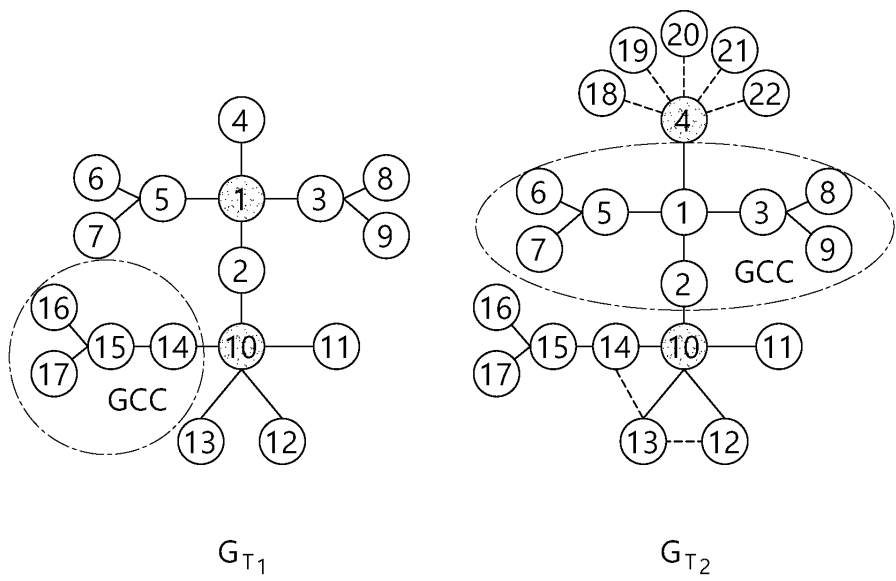
도면1



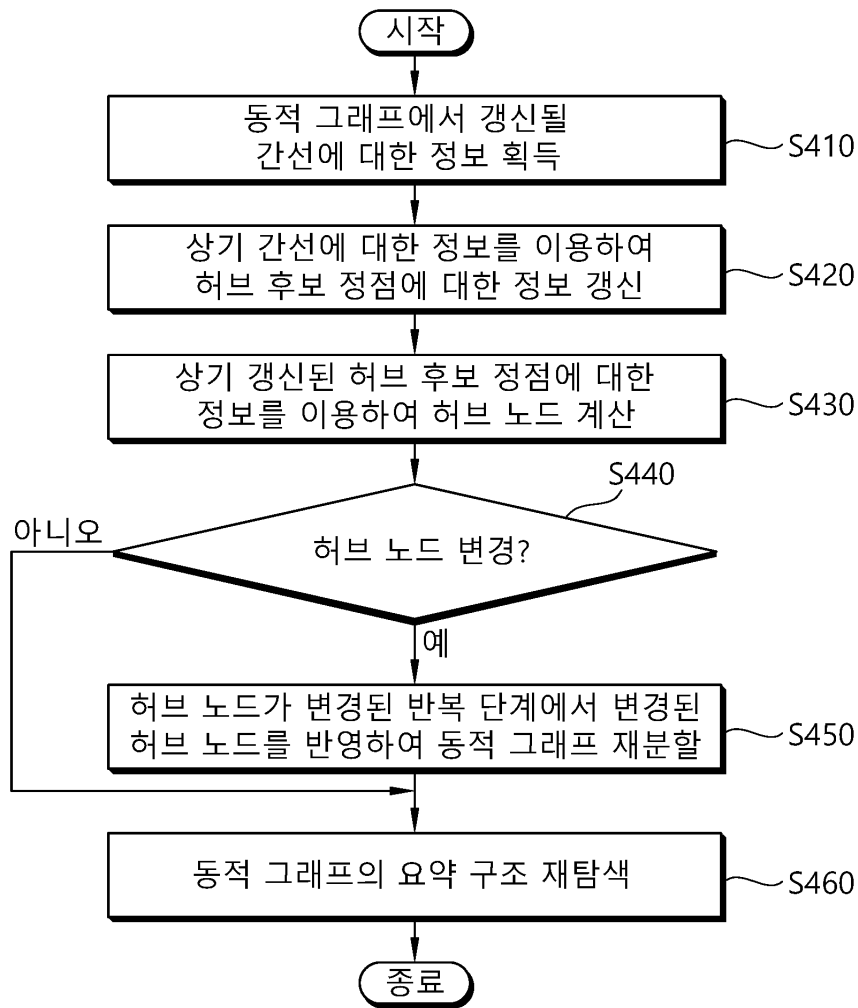
도면2



도면3



도면4



도면5

Algorithm 1. DynamicGS ($G_{T_{i-1}}, HC_{T_{i-1}}, \Delta E_{T_{i-1}}$)

Input : a graph $G_{t_{i-1}}$ ($i > 1$),
 hubset candidate $HC_{t_{i-1}}$
 a set of edge $\Delta E_{t_{i-1}}$

Output : a set of summarized structure $S_{t_i} = \{s_1, s_2, \dots, s_m\}$

- 1: $j = 0$
- 2: $j_{\max} \leftarrow$ get a max iteration level in $G_{t_{i-1}}$
- 3: calculate a set of hubset $H_{t_i}^j$ by using $HC_{t_i}^j$
- 4: **WHILE** ($j < j_{\max}$ **AND** $H_{t_i}^j = H_{t_{i-1}}^j$)
- 5: $j \leftarrow j + 1$
- 6: update a set of hubset candidate $HC_{t_i}^j$
 by using ΔE_{T_i} from G_{t_i}
- 7: calculate a set of hubset $H_{t_i}^j$ by using
 $HC_{t_i}^j$
- 8: **if** $j \neq j_{\max}$ **THEN**
- 9: $S_{t_i} \leftarrow$ SlashBurn($GCC_{t_i}^j$)
- 10: $S_{t_i} \leftarrow S_{t_i} \cup$ UpdateStructures($SP_{t_{i-1}}, \Delta E_{T_i}, j$)
- 11: **RETURN** S_{t_i}

도면6

Algorithm 2. UpdateStructures($SP_{t_{i-1}}, \Delta E_{T_i}, j$)

Input : a set of Spoke $SP_{t_{i-1}} = \{sp_1, sp_2, \dots, sp_m\}$,
 a set of edge ΔE_{T_i} ,
 an iteration level j

Output : a set of summarized structure $S_{t_i} = \{s_1, s_2, \dots, s_m\}$

- 1: $i' = 0, S_{t_i} \leftarrow \emptyset$
- 2: **WHILE** ($i' < j$)
- 3: $SP_{t_i}^{i'} \leftarrow$ connected_components($SP_{t_{i-1}}^{i'}, \Delta E_{T_i}$)
- 4: find a set of summarized structure $S_{t_i}^{i'}$
 in $SP_{t_i}^{i'}$
- 5: $S_{t_i} \leftarrow S_{t_i} \cup S_{t_i}^{i'}$
- 6: $i' \leftarrow i' + 1$
- 7: **RETURN** S_{t_i}

도면7

