(19) 대한민국특허청(KR)
(12) 공개특허공보(A)

(11) 공개번호 10-2015-0079370
(43) 공개일자 2015년07월08일

(51) 국제특허분류(Int. Cl.)
G06F 17/00 (2006.01) *G06F 17/30* (2006.01)
(21) 출원번호 10-2014-0062980
(22) 출원일자 2014년05월26일
심사청구일자 2014년05월26일
(30) 우선권주장
1020130165499 2013년12월27일 대한민국(KR)

(71) 출원인
경희대학교 산학협력단
경기도 용인시 기흥구 덕영대로 1732 (서천동, 경희대학교 국제캠퍼스내)
(72) 발명자
이영구
대전 서구 청사서로 11, 106동 302호 (월평동, 무지개아파트)
윤영선
경기 수원시 영통구 덕영대로1681번길 17, 302호 (영통동)
박기성
경기 부천시 원미구 원미로 123-4, E동 4호 (원미동, 백송연립)
(74) 대리인
서재승

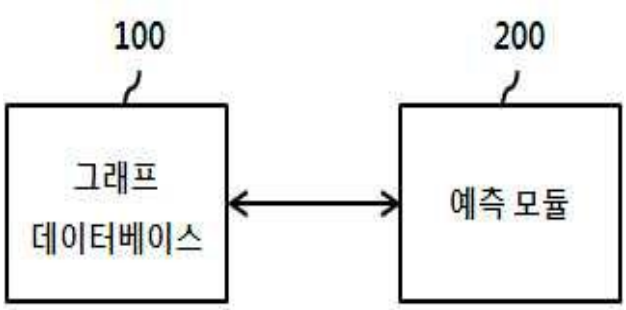
전체 청구항 수 : 총 15 항

(54) 발명의 명칭 **대용량 그래프 데이터베이스에서 링크를 예측하는 방법**

(57) 요 약

본 발명은 대용량 그래프 데이터베이스에서 링크를 예측하는 방법에 관한 것으로, 순환 그래프를 검색하는 방법에 관한 것으로, 보다 구체적으로 대용량 그래프 데이터베이스를 다수의 그래프 그룹으로 분산하여 전체 대용량 그래프 데이터베이스에서 노드 사이의 링크를 정확하게 예측할 수 있으며, 프리겔 프레임워크를 이용하여 모든 노드를 단일 시작점으로 하여 동시에 노드 사이의 링크를 예측할 수 있는 방법을 제공하는 것이다.

대 표 도 - 도2



이 발명을 지원한 국가연구개발사업
과제고유번호 17110-01260
부처명 미래부
연구관리전문기관 정보통신산업진흥원
연구사업명 2013 대학 IT 연구센터육성 지원사업
연구과제명 실시간 모바일 클라우드 서비스 플랫폼 연구개발
기 여 율 1/1
주관기관 경희대학교산학협력단
연구기간 2013.06.01 ~ 2013.12.31

명 세 서

청구범위

청구항 1

대용량 그래프 데이터베이스에서 링크를 예측하는 방법에 있어서,

상기 대용량 그래프 데이터베이스는 다수개의 그래프 그룹으로 분리되어 분산되어 있으며,

상기 그래프 그룹을 구성하는 각 버텍스를 기준 버텍스로 하여 상기 기준 버텍스에 인접한 인접 버텍스로 상기 기준 버텍스의 에지 차수를 송신하는 단계;

상기 인접 버텍스에서 계산된 상기 기준 버텍스와 상기 인접 버텍스 사이의 인접 유사도를 수신하는 단계;

상기 기준 버텍스에서 상기 인접 버텍스로부터 비인접 유사도 및 시작 버텍스 식별자를 구비하는 비인접 메시지를 수신하는 경우, 상기 비인접 유사도와 상기 인접 유사도를 곱하여 상기 비인접 유사도의 시작 버텍스와 상기 기준 버텍스 사이의 확장 유사도를 계산하는 단계; 및

상기 확장 유사도의 크기에 기초하여 상기 기준 버텍스에 링크로 연결할 시작 버텍스를 예측하는 단계를 포함하는 것을 특징으로 하는 링크 예측 방법.

청구항 2

제 1 항에 있어서, 상기 링크 예측 방법에서

상기 인접 버텍스로부터 비인접 메시지를 더 이상 수신하지 않을 때까지 상기 비인접 유사도의 시작 버텍스와 상기 기준 버텍스 사이의 확장 유사도를 계산하는 것을 특징으로 하는 링크 예측 방법.

청구항 3

제 2 항에 있어서, 동일한 시작 버텍스와 상기 기준 버텍스 사이에 다수의 확장 유사도가 계산되는 경우, 상기 동일한 시작 버텍스와 상기 기준 버텍스 사이의 다수 확장 유사도 중에서 가장 높은 확장 유사도를 상기 시작 버텍스와 상기 기준 버텍스 사이의 확장 유사도로 결정하는 것을 특징으로 하는 링크 예측 방법.

청구항 4

제 2 항에 있어서,

상기 인접 유사도는 상기 인접 버텍스에서 상기 기준 버텍스의 차수와 상기 인접 버텍스의 차수로부터 계산되는 것을 특징으로 하는 링크 예측 방법.

청구항 5

제 4 항에 있어서,

상기 인접 버텍스(v_n)와 상기 기준 버텍스(v_s) 사이의 상기 인접 유사도(S(n))는 아래의 수학식(1)에 의해 계산되며,

[수학식 1]

$$S(n)=\frac{1}{\deg(v_n)+\deg(v_s)-1}$$

여기서 $\deg(v_i)$는 버텍스(v_i)의 에지 차수인 것을 특징으로 하는 링크 예측 방법.

청구항 6

제 4 항에 있어서, 상기 비인접 유사도는

상기 시작 버텍스로부터 상기 인접 버텍스까지 순차적으로 인접하고 있는 버텍스 사이의 인접 유사도를 곱하여 계산되는 것을 특징으로 하는 링크 예측 방법.

청구항 7

제 1 항 내지 제 6 항 중 어느 한 항에 있어서,

상기 시작 버텍스와 상기 기준 버텍스 사이의 확장 유사도에 기초하여 상기 기준 버텍스의 인접 버텍스를 제외하고 가장 큰 확장 유사도를 가지는 시작 버텍스를 상기 기준 버텍스에 링크로 연결할 시작 버텍스로 예측하는 것을 특징으로 하는 링크 예측 방법.

청구항 8

대용량 그래프 데이터베이스에서 링크를 예측하는 장치에 있어서,

다수개의 그래프 그룹으로 분리되어 있는 상기 대용량 그래프 데이터베이스 중 1개의 그래프 그룹이 분산 저장되어 있는 그래프 데이터베이스; 및

상기 그래프 그룹을 구성하는 각 버텍스를 기준 버텍스로 설정하여 상기 기준 버텍스로부터 인접한 인접 버텍스로부터 수신한 인접 유사도와 비인접 유사도에 기초하여 상기 비인접 유사도의 시작 버텍스와 상기 기준 버텍스 사이의 확장 유사도를 계산하고, 상기 확장 유사도에 기초하여 링크로 연결할 시작 버텍스를 예측하는 예측 모듈을 포함하는 것을 특징으로 하는 링크 예측 장치.

청구항 9

제 8 항에 있어서, 상기 예측 모듈은

상기 기준 버텍스에 인접한 인접 버텍스로 상기 기준 버텍스의 에지 차수를 송신하고 상기 인접 버텍스로부터 상기 인접 버텍스의 에지 차수를 수신하는 에지 차수 관리부;

상기 인접 버텍스의 에지 차수와 상기 기준 버텍스의 에지 차수에 기초하여 상기 인접 버텍스와 상기 기준 버텍스 사이의 인접 유사도를 계산하는 인접 유사도 계산부;

상기 인접 버텍스로부터 수신한 비인접 유사도 및 시작 버텍스 식별자에 기초하여 상기 비인접 유사도와 상기 인접 유사도를 곱하여 상기 비인접 유사도의 시작 버텍스와 상기 기준 버텍스 사이의 확장 유사도를 계산하는 확장 유사도 계산부; 및

상기 확장 유사도의 크기에 기초하여 상기 기준 버텍스에 링크로 연결할 시작 버텍스를 예측하는 예측부를 포함하는 것을 특징으로 하는 링크 예측 장치.

청구항 10

제 9 항에 있어서, 상기 확장 유사도 계산부는

상기 인접 버텍스로부터 비인접 메시지를 더 이상 수신하지 않을 때까지 상기 비인접 유사도의 시작 버텍스와 상기 기준 버텍스 사이의 확장 유사도를 계산하는 것을 특징으로 하는 링크 예측 장치.

청구항 11

제 10 항에 있어서, 상기 확장 유사도 계산부는

동일한 시작 버텍스와 상기 기준 버텍스 사이에 다수의 확장 유사도가 계산되는 경우, 상기 동일한 시작 버텍스와 상기 기준 버텍스 사이에 다수의 확장 유사도 중에서 가장 높은 확장 유사도를 상기 시작 버텍스와 상기 기준 버텍스 사이의 확장 유사도로 결정하는 것을 특징으로 하는 링크 예측 장치.

청구항 12

제 10 항에 있어서,

상기 인접 유사도는 상기 인접 버텍스에서 상기 기준 버텍스의 차수와 상기 인접 버텍스의 차수로부터 계산되는

것을 특징으로 하는 링크 예측 장치.

청구항 13

제 12 항에 있어서,

상기 인접 버텍스(v_n)와 상기 기준 버텍스(v_s) 사이의 상기 인접 유사도(S(n))는 아래의 수학식(2)에 의해 계산되며,

[수학식 2]

$$S(n)=\frac{1}{\deg(v_n)+\deg(v_s)-1}$$

여기서 $\deg(v_i)$는 버텍스(v_i)의 에지 차수인 것을 특징으로 하는 링크 예측 장치.

청구항 14

제 12 항에 있어서, 상기 비인접 유사도는

상기 시작 버텍스로부터 상기 인접 버텍스까지 순차적으로 인접하고 있는 버텍스 사이의 인접 유사도를 곱하여 계산되는 것을 특징으로 하는 링크 예측 장치.

청구항 15

제 8 항 내지 제 14 항 중 어느 한 항에 있어서,

상기 시작 버텍스와 상기 기준 버텍스 사이의 확장 유사도에 기초하여 상기 기준 버텍스의 인접 버텍스를 제외하고 가장 큰 확장 유사도를 가지는 시작 버텍스를 상기 기준 버텍스에 링크로 연결할 시작 버텍스로 예측하는 것을 특징으로 하는 링크 예측 장치.

발명의 설명

기 술 분 야

[0001] 본 발명은 대용량 그래프 데이터베이스에서 링크를 예측하는 방법에 관한 것으로, 보다 구체적으로 대용량 그래프 데이터베이스를 다수의 그래프 그룹으로 분산하여 전체 대용량 그래프 데이터베이스에서 노드 사이의 링크를 정확하게 예측할 수 있으며, 프리겔 프레임워크를 이용하여 모든 노드를 단일 시작점으로 하여 동시에 노드 사이의 링크를 예측할 수 있는 방법을 제공하는 것이다.

배 경 기 술

[0002] 페타(peta) 바이트 이상의 규모의 대용량 데이터 분석은 다양한 분야에서 연구되고 있다. 특히, 소셜 네트워크, XML 등과 같은 구조적인 정보를 갖는 대용량의 그래프들을 분석하는 기술이 활발히 연구되고 있다. 예를 들어, 소셜 네트워크에서 사용자 식별자는 노드 버텍스로, 사용자 사이의 관계는 노드 버텍스를 연결하는 노드 에지로 표현될 수 있다.

[0003] 최근 소셜 네트워크 서비스의 등장으로 인해 정보 전달의 속도가 빨라지고, 사용자 정보나 사용자 사이의 관계 데이터 등 방대한 양의 데이터가 생성되고 있다. 소셜 네트워크에서 발생한 빅 데이터는 의견 분석, 관계망 분석, 사회 현상 예측 등 다양한 애플리케이션에 응용할 수 있기 때문에, 대량의 소셜 네트워크 데이터를 분석하기 위한 기술의 필요성이 증가하고 있다. 소셜 네트워크 분석에서는 사용자 간의 관계 패턴을 분석하거나, 관계의 형성, 변화 등의 흐름을 분석한다. 특히, 관계의 흐름을 예측하기 위해 그래프 상에서의 링크 예측 기법이 널리 사용된다. 링크 예측 기법은 친구 추천 서비스, 상관 관계분석, 테러리스트 네트워크 모니터링 등 다양한 분야에서 사용되고 있다.

[0004] 링크 예측 기법이란 특정 시각(t1)의 소셜 네트워크로부터 주어진 시각(t2)에 발생할 링크를 예측하는 기법이다. 기존의 링크 예측 기법들은 데이터 용량이 너무 방대하여 전체 소셜 네트워크 데이터를 고려하지 않고 전체 소셜 네트워크 데이터 중 일정 범위 이내에 위치하는 소셜 네트워크 데이터만을 이용하기 때문에 정확

도가 떨어지며, 최단 경로를 이용한 종래 링크 예측 방법의 경우 소셜 네트워크 전체를 활용하기 위해 메모리상에서 모든 노드간의 최단 경로 연산을 수행하여 대용량 데이터에 적합하지 않다.

발명의 내용

해결하려는 과제

[0005] 본 발명은 위에서 언급한 종래 링크 예측 방법이 가지는 문제점을 해결하기 위한 것으로, 본 발명이 이루고자 하는 목적은 대용량 그래프 데이터베이스를 다수의 그래프 그룹으로 분산하여 대용량 그래프 데이터베이스에서 링크를 예측하는 방법을 제공하는 것이다.

[0006] 한편, 본 발명이 이루고자 하는 다른 목적은 대용량 그래프 데이터베이스를 다수의 그래프 그룹으로 분산하여 전체 대용량 그래프 데이터베이스에서 노드 사이의 링크를 정확하게 예측할 수 있는 방법을 제공하는 것이다.

[0007] 본 발명이 이루고자 하는 또 다른 목적은 프리겔 프레임워크를 이용하여 모든 노드를 단일 시작점으로 하여 동시에 노드 사이의 링크를 예측할 수 있는 방법을 제공하는 것이다.

과제의 해결 수단

[0008] 본 발명의 목적을 달성하기 위하여, 본 발명의 일 실시예에 따른 링크 예측 방법은 대용량 그래프 데이터베이스가 다수개의 그래프 그룹으로 분리되어 분산되어 있으며,

[0009] 그래프 그룹을 구성하는 각 버텍스를 기준 버텍스로 하여 기준 버텍스에 인접한 인접 버텍스로 기준 버텍스의 에지 차수를 송신하는 단계와, 인접 버텍스에서 계산된 기준 버텍스와 인접 버텍스 사이의 인접 유사도를 수신하는 단계와, 기준 버텍스에서 인접 버텍스로부터 비인접 유사도 및 시작 버텍스 식별자를 구비하는 비인접 메시지를 수신하는 경우 비인접 유사도와 인접 유사도를 곱하여 비인접 유사도의 시작 버텍스와 기준 버텍스 사이의 확장 유사도를 계산하는 단계와, 확장 유사도의 크기에 기초하여 기준 버텍스에 링크로 연결할 시작 버텍스를 예측하는 단계를 포함하는 것을 특징으로 한다.

[0010] 본 발명에 따른 링크 예측 방법에서 인접 버텍스로부터 비인접 메시지를 더 이상 수신하지 않을 때까지 비인접 유사도의 시작 버텍스와 기준 버텍스 사이의 확장 유사도를 계산하는 것을 특징으로 한다.

[0011] 여기서 동일한 시작 버텍스와 기준 버텍스 사이에 다수의 확장 유사도가 계산되는 경우, 동일한 시작 버텍스와 기준 버텍스 사이의 다수 확장 유사도 중에서 가장 높은 확장 유사도를 시작 버텍스와 기준 버텍스 사이의 확장 유사도로 결정하는 것을 특징으로 한다.

[0012] 여기서 인접 유사도는 인접 버텍스에서 기준 버텍스의 차수와 인접 버텍스의 차수로부터 계산되는 것을 특징으로 한다.

[0013] 인접 버텍스(v_n)와 기준 버텍스(v_s) 사이의 인접 유사도(S(n))는 아래의 수학식(1)에 의해 계산되며,

[0014] [수학식 1]

$$S(n)=\frac{1}{\deg(v_n)+\deg(v_s)-1}$$

[0015]

[0016] 여기서 $\deg(v_i)$는 버텍스(v_i)의 에지 차수인 것을 특징으로 한다.

[0017] 여기서 비인접 유사도는 시작 버텍스로부터 인접 버텍스까지 순차적으로 인접하고 있는 버텍스 사이의 인접 유사도를 곱하여 계산되는 것을 특징으로 한다.

[0018] 본 발명에 따른 링크 예측 방법에서 시작 버텍스와 기준 버텍스 사이의 확장 유사도에 기초하여 기준 버텍스의 인접 버텍스를 제외하고 가장 큰 확장 유사도를 가지는 시작 버텍스를 기준 버텍스에 링크로 연결할 시작 버텍스로 예측하는 것을 특징으로 한다.

[0019] 한편, 본 발명에 따른 링크 예측 장치는 다수개의 그래프 그룹으로 분리되어 있는 대용량 그래프 데이터베이스 중 1개의 그래프 그룹이 분산 저장되어 있는 그래프 데이터베이스와, 그래프 그룹을 구성하는 각 버텍스를 기준 버텍스로 설정하여 기준 버텍스로부터 인접한 인접 버텍스로부터 수신한 인접 유사도와 비인접 유사도에 기초하

여 비인접 유사도의 시작 버텍스와 기준 버텍스 사이의 확장 유사도를 계산하고 확장 유사도에 기초하여 링크로 연결할 시작 버텍스를 예측하는 예측 모듈을 포함하는 것을 특징으로 한다.

[0020] 여기서 예측 모듈은 기준 버텍스에 인접한 인접 버텍스로 기준 버텍스의 에지 차수를 송신하고 인접 버텍스로부터 인접 버텍스의 에지 차수를 수신하는 에지 차수 관리부와, 인접 버텍스의 에지 차수와 기준 버텍스의 에지 차수에 기초하여 인접 버텍스와 기준 버텍스 사이의 인접 유사도를 계산하는 인접 유사도 계산부와, 인접 버텍스로부터 수신한 비인접 유사도 및 시작 버텍스 식별자에 기초하여 비인접 유사도와 인접 유사도를 곱하여 비인접 유사도의 시작 버텍스와 기준 버텍스 사이의 확장 유사도를 계산하는 확장 유사도 계산부와, 확장 유사도의 크기에 기초하여 기준 버텍스에 링크로 연결할 시작 버텍스를 예측하는 예측부를 포함하는 것을 특징으로 한다.

[0021] 여기서 확장 유사도 계산부는 인접 버텍스로부터 비인접 메시지를 더 이상 수신하지 않을 때까지 비인접 유사도의 시작 버텍스와 기준 버텍스 사이의 확장 유사도를 계산하는 것을 특징으로 한다.

[0022] 확장 유사도 계산부는 동일한 시작 버텍스와 기준 버텍스 사이에 다수의 확장 유사도가 계산되는 경우, 동일한 시작 버텍스와 기준 버텍스 사이에 다수의 확장 유사도 중에서 가장 높은 확장 유사도를 시작 버텍스와 기준 버텍스 사이의 확장 유사도로 결정하는 것을 특징으로 한다.

발명의 효과

[0023] 본 발명에 따른 링크 예측 방법은 대용량 그래프 데이터베이스를 다수의 그래프 그룹으로 분산하고 전체 대용량 그래프 데이터베이스에서 노드 사이의 링크를 예측함으로써, 정확하게 노드 사이의 링크를 예측할 수 있다.

[0024] 또한 본 발명에 따른 링크 예측 방법은 프리겔 프레임워크를 이용하여 모든 노드를 단일 시작점으로 하여 동시에 노드 사이의 링크를 예측함으로써, 적은 연산량으로 빠르게 노드 사이의 링크를 예측할 수 있다.

도면의 간단한 설명

[0025] 도 1은 본 발명의 일 실시예에 따른, 다수의 그래프 그룹으로 이루어진 대용량 그래프 데이터베이스의 일 예를 도시하고 있다.

도 2는 본 발명의 일 실시예에 따른 대용량 그래프 데이터베이스에서 링크를 예측하는 장치를 설명하기 위한 기능 블록도이다.

도 3은 본 발명의 일 실시예에 따른 예측 모듈을 보다 구체적으로 설명하기 위한 기능 블록도이다.

도 4는 본 발명의 일 실시예에 따른 대용량 그래프 데이터베이스에서 링크를 예측하는 방법을 설명하기 위한 흐름도이다.

도 5는 본 발명에 따른 링크 예측 방법에서 확장 유사도를 갱신하는 방법을 설명하기 위한 흐름도이다.

도 6은 본 발명의 일 예를 설명하기 위한 그래프 그룹의 예를 도시하고 있다.

도 7은 도 6에 도시되어 있는 기준 버텍스와 인접 버텍스 사이의 인접 유사도를 계산한 일 예를 도시하고 있다.

도 8은 도 6에 도시되어 있는 기준 버텍스와 비인접 버텍스 사이의 확장 유사도를 계산한 일 예를 도시하고 있다.

도 9는 서로 다른 인접 버텍스로부터 동일 시작 버텍스 식별자를 구비하는 경우 확장 유사도를 계산하는 일 예를 설명하기 위한 도면이다.

발명을 실시하기 위한 구체적인 내용

[0026] 이하 첨부한 도면을 참고로 본 발명의 일 실시예에 따른 대용량 그래프 데이터베이스에서 링크를 예측하는 방법 및 그 장치에 대해 보다 구체적으로 설명한다.

[0027] 도 1은 본 발명의 일 실시예에 따른, 다수의 그래프 그룹으로 이루어진 대용량 그래프 데이터베이스의 일 예를 도시하고 있다.

[0028] 도 1을 참고로 살펴보면, 원으로 표현된 다수의 노드 버텍스와 노드 버텍스 사이를 연결하는 에지로 구성되어 있는 대용량 그래프 데이터베이스가 존재한다. 여기서 노드 버텍스는 소셜 네트워크의 경우 사용자 이름, 아이

디 등과 같은 사용자 식별자이며, 노드 에지는 사용자의 관계를 연결한다.

[0029] 대용량 그래프 데이터베이스는 다수의 그래프 그룹(GG1, GG2, GG3)으로 구분되어 그래프 그룹 단위로 링크를 예측하는 장치에 분산 저장되어 있는데, 링크를 예측하는 장치는 각 그래프 그룹 단위로 각 그래프 그룹을 구성하는 버텍스 사이의 링크 또는 서로 다른 그래프 그룹에 존재하는 버텍스 사이의 링크를 예측한다.

[0030] 여기서 그래프 그룹에서 다른 그래프 그룹과의 경계에 위치하는 경계 버텍스는 다른 그래프 그룹에서 경계 버텍스에 연결되어 있는 인접 버텍스의 식별자 정보를 가지는데, 예를 들어 그래프 그룹(GG1)의 경계 버텍스A는 그래프 그룹(GG1)의 경계에 위치하는 그래프 그룹(GG3)에서 경계 버텍스A에 연결되어 있는 인접 버텍스B의 식별자 정보를 가지고 있다. 이러한 인접 버텍스B의 식별자 정보는 그래프 데이터베이스에 저장될 수 있다.

[0031] 도 2는 본 발명의 일 실시예에 따른 대용량 그래프 데이터베이스에서 링크를 예측하는 장치를 설명하기 위한 기능 블록도이다.

[0032] 도 2를 참고로 보다 구체적으로 살펴보면, 그래프 데이터베이스(100)는 다수개의 그래프 그룹으로 분리되어 있는 대용량 그래프 데이터베이스 중 적어도 1개 이상의 그래프 그룹이 분산 저장되어 있다.

[0033] 예측 모듈(200)는 그래프 그룹을 구성하는 각 버텍스를 기준 버텍스로 설정하여 기준 버텍스에 인접한 인접 버텍스로부터 수신한 인접 유사도와 비인접 유사도에 기초하여 비인접 유사도의 시작 버텍스와 기준 버텍스 사이의 확장 유사도를 계산하고, 확장 유사도에 기초하여 링크로 연결할 시작 버텍스를 예측한다.

[0034] 도 3은 본 발명의 일 실시예에 따른 예측 모듈을 보다 구체적으로 설명하기 위한 기능 블록도이다.

[0035] 도 3을 참고로 살펴보면, 에지 차수 관리부(210)는 기준 버텍스에 인접한 인접 버텍스로 기준 버텍스의 에지 차수를 송신하고 인접 버텍스로부터 인접 버텍스의 에지 차수를 수신한다.

[0036] 인접 유사도 계산부(230)는 인접 버텍스로부터 수신한 인접 버텍스의 에지 차수와 기준 버텍스의 에지 차수에 기초하여 인접 버텍스와 기준 버텍스 사이의 인접 유사도를 계산한다. 확장 유사도 계산부(270)는 인접 버텍스로부터 수신한 비인접 유사도 및 시작 버텍스 식별자에 기초하여 비인접 유사도와 인접 유사도를 곱하여 비인접 유사도의 시작 버텍스와 기준 버텍스 사이의 확장 유사도를 계산한다. 인접 유사도 계산부(230)에서 계산한 인접 버텍스와 기준 버텍스 사이의 인접 유사도 및 확장 유사도 계산부(270)에서 계산한 비인접 유사도의 시작 버텍스와 기준 버텍스 사이의 확장 유사도는 저장부(250)에 저장된다.

[0037] 예측부(290)는 확장 유사도의 크기에 기초하여 기준 버텍스에 링크로 연결할 시작 버텍스를 예측한다. 여기서 기준 버텍스는 인접 버텍스로부터 인접 유사도 또는 확장 유사도를 수신하여 기준 버텍스에 연결할 링크의 시작 버텍스를 판단하는 버텍스를 의미하는데, 본 발명에서는 그래프 그룹에 존재하는 모든 버텍스 각각을 기준으로 동시에 연결할 링크의 시작 버텍스를 판단하므로, 그래프 그룹에 존재하는 모든 버텍스는 기준 버텍스로 동작하며, 기준 버텍스에 인접하고 있는 버텍스는 상대적으로 인접 버텍스로 동작한다.

[0038] 도 4는 본 발명의 일 실시예에 따른 대용량 그래프 데이터베이스에서 링크를 예측하는 방법을 설명하기 위한 흐름도이다.

[0039] 도 4를 참고로 보다 구체적으로 살펴보면, 그래프 그룹의 각 버텍스는 자신을 기준 버텍스로 하여 기준 버텍스에 에지로 직접 연결되어 있는 인접 버텍스로 기준 버텍스의 에지 차수에 대한 정보를 송신하고 기준 버텍스에 인접하고 있는 인접 버텍스로 인접 버텍스의 에지 차수에 대한 정보를 수신한다(S110). 기준 버텍스는 인접 버텍스로부터 인접 버텍스에서 계산된 기준 버텍스와 인접 버텍스 사이의 인접 유사도를 수신하는데(S120), 인접 버텍스는 기준 버텍스의 차수와 인접 버텍스의 차수로부터 인접 유사도를 계산한다. 바람직하게, 인접 버텍스는 아래의 수학식(1)과 같이 인접 버텍스(v_n)와 기준 버텍스(v_s) 사이의 인접 유사도(S(n))를 계산하는데,

[0040] [수학식 1]

$$S(n)=\frac{1}{\deg(v_n)+\deg(v_s)-1}$$

[0041]

[0042] 여기서 $\deg(v_n)$은 버텍스(v_n)의 에지 차수이며, $\deg(v_s)$는 버텍스(v_s)의 에지 차수인 것을 특징으로 한다.

[0043] 인접 버텍스로부터 비인접 유사도와 시작 버텍스 식별자를 구비하는 비인접 메시지를 수신하는지 판단하여(S130), 비인접 메시지를 수신하는 경우 기준 버텍스는 비인접 유사도와 인접 유사도를 곱하여 비인접 유사도의 시작 버텍스와 기준 버텍스 사이의 확장 유사도를 계산한다(S140). 여기서 비인접 유사도는 시작 버텍스로부터 인접 버텍스까지 순차적으로 인접하고 있는 버텍스 사이의 인접 유사도를 곱하여 계산된다. 인접 버텍스로부터 비인접 메시지를 더 이상 수신하지 않을 때까지 비인접 유사도의 시작 버텍스와 기준 버텍스 사이의 확장 유사도를 계산하는데, 이를 통해 모든 그래프 그룹에 존재하는 버텍스를 시작 버텍스로 하여 시작 버텍스와 기준 버텍스 사이의 확장 유사도를 계산한다.

[0044] 인접 버텍스로부터 더 이상 비인접 메시지를 수신하지 않는 경우, 현재 시점에서 계산한 확장 유사도에 기초하여 다음 시점에 기준 버텍스에 링크로 연결될 시작 버텍스를 예측한다(S150). 바람직하게, 모든 그래프 그룹에 존재하는 버텍스를 각 시작 버텍스로 하여 시작 버텍스와 기준 버텍스 사이의 확장 유사도를 계산하며, 가장 높은 확장 유사도를 가지는 또는 임계 확장 유사도 이상의 확장 유사도를 가지는 시작 버텍스를 기준 버텍스와 링크로 연결할 시작 버텍스로 예측한다. 더욱 바람직하게, 현재 시점에서 기연결되어 있는 시작 버텍스를 제외하고 가장 높은 확장 유사도를 가지는 또는 임계 확장 유사도 이상의 확장 유사도를 가지는 시작 버텍스를 기준 버텍스와 링크로 연결할 시작 버텍스로 예측한다.

[0045] 도 5는 본 발명에 따른 링크 예측 방법에서 확장 유사도를 갱신하는 방법을 설명하기 위한 흐름도이다.

[0046] 도 5를 참고로 보다 구체적으로 살펴보면, 서로 다른 인접 버텍스로부터 동일한 시작 버텍스 식별자를 구비하는 비인접 메시지를 다수 수신하였는지 판단하고(S141), 서로 다른 인접 버텍스로부터 다수의 비인접 메시지를 수신하는 경우 각 비인접 메시지에 구비되어 있는 시작 버텍스와 인접 메시지 사이의 비인접 유사도 및 기준 버텍스와 인접 메시지 사이의 인접 유사도에 기초하여 계산한 확장 유사도의 크기를 비교한다(S143). 동일한 시작 버텍스와 기준 버텍스 사이의 확장 유사도 중 가장 높은 확장 유사도를 선택하거나 기존에 저장된 동일한 시작 버텍스와 기준 버텍스 사이의 확장 유사도를 높은 확장 유사도로 갱신한다(S145). 즉, 동시에 다수의 인접 버텍스로부터 동일한 시작 버텍스 식별자를 구비하는 비인접 메시지를 수신하는 경우, 가장 높은 확장 유사도를 선택한다. 한편, 순차적으로 다수의 인접 버텍스로부터 동일한 시작 버텍스 식별자를 구비하는 비인접 메시지를 수신하는 경우, 기존에 저장되어 있는 확장 유사도보다 큰 값을 가지는 확장 유사도로 확장 유사도를 갱신한다.

[0047] 도 6은 본 발명의 일 예를 설명하기 위한 그래프 그룹의 예를 도시하고 있으며, 도 7은 도 6에 도시되어 있는 기준 버텍스와 인접 버텍스 사이의 인접 유사도를 계산한 일 예를 도시하고 있으며, 도 8은 도 6에 도시되어 있는 기준 버텍스와 비인접 시작 버텍스 사이의 확장 유사도를 계산한 일 예를 도시하고 있다.

[0048] 도 6 내지 도 8을 참고로 보다 구체적으로 살펴보면, 도 6의 그래프 그룹을 구성하는 각 버텍스(V1, V2, V3, V4, V5, V6)은 각각 기준 버텍스로 동작하며 기준 버텍스에 에지로 연결되어 있는 인접한 버텍스는 인접 버텍스로 동작한다. 예를 들어 V1이 기준 버텍스로 동작하는 경우 V1에 에지로 연결되어 있는 V2, V3, V6은 인접 버텍스로 동작하며, V2가 기준 버텍스로 동작하는 경우 V2에 에지로 연결되어 있는 V4, V1은 인접 버텍스로 동작한다.

[0049] 처음 단계(1 superstep)에서 각 버텍스는 기준 버텍스로 동작하여 인접 버텍스로 기준 버텍스의 에지 차수에 대한 정보를 제공한다. 예를 들어, 도 7(a)에 도시되어 있는 바와 같이 기준 버텍스(V1)는 에지로 연결되어 있는 인접 버텍스(V2)로부터 인접 버텍스(V2)의 에지 차수에 대한 정보를 수신하고, 기준 버텍스(V1)도 인접 버텍스(V2)로 기준 버텍스(V1)의 에지 차수에 대한 정보를 송신한다. 한편, 기준 버텍스(V4)는 인접 버텍스(V2)로부터 인접 버텍스(V2)의 에지 차수에 대한 정보를 수신하고, 기준 버텍스(V4)는 인접 버텍스(V2)로 기준 버텍스

(V4)의 에지 차수에 대한 정보를 송신한다.

[0050] 다음 단계(2 superstep)에서 인접 버텍스(V2)는 기준 버텍스(V1)의 에지 차수(3)와 인접 버텍스(V2)의 에지 차수(2)를 이용하여 인접 버텍스(V2)와 기준 버텍스(V1) 사이의 인접 유사도를 계산하고, 계산한 인접 버텍스(V2)와 기준 버텍스(V1) 사이의 인접 유사도를 기준 버텍스(V1)와 인접 버텍스(V2)에 인접한 버텍스(V4)로 제공한다. 또한 인접 버텍스(V2)는 기준 버텍스(V4)의 에지 차수(3)와 인접 버텍스(V2)의 에지 차수(2)를 이용하여 인접 버텍스(V2)와 기준 버텍스(V4) 사이의 인접 유사도를 계산하고, 계산한 인접 버텍스(V2)와 기준 버텍스(V4) 사이의 인접 유사도를 기준 버텍스(V4)와 인접 버텍스(V2)에 인접한 버텍스(V1)로 제공한다. 도 7(b)는 기준 버텍스와 인접 버텍스 사이에서 계산된 인접 유사도를 도시하고 있다.

[0051] 즉, 버텍스(V4)는 인접 버텍스(V2)로부터 인접하지 않은 기준 버텍스(V1)의 식별자와 기준 버텍스(V1)와 인접 버텍스(V2) 사이의 비인접 유사도를 가지는 비인접 메시지를 수신한다. 버텍스(V4)의 입장에서 버텍스(V4)는 기준 버텍스이며 버텍스(V1)은 버텍스(V4)에 인접하지 않은 시작 버텍스이다.

[0052] 도 8(a)를 참고로 살펴보면, 다음 단계(3 superstep)에서 기준 버텍스(V4)는 버텍스(V1)와 인접 버텍스(V2) 사이의 비인접 유사도와 기준 버텍스(V4)와 인접 버텍스(V2) 사이의 인접 유사도를 곱하여 기준 버텍스(V4)와 시작 버텍스(V1) 사이의 확장 유사도를 계산한다. 도 8(b)는 다음 단계에서 계산된 버텍스 사이의 확장 유사도를 계산한 것이다.

[0053] 그 다음 단계(superstep4)에서 버텍스(V5)를 기준 버텍스로 하는 경우 버텍스(V5)는 인접 버텍스(V4)로부터 시작 버텍스(V1) 식별자와 시작 버텍스(V1)로부터 인접 버텍스(V4)까지의 비인접 유사도(버텍스(V1)과 버텍스(V3) 사이의 인접 유사도×버텍스(V3)과 버텍스(V4)사이의 인접 유사도)를 구비하는 비인접 메시지를 수신하고 버텍스(V5)와 인접 버텍스(V4) 사이의 인접 유사도 및 시작 버텍스(V1)로부터 인접 버텍스(V4)까지의 비인접 유사도를 곱하여 버텍스(V5)와 버텍스(V1) 사이의 확장 유사도를 계산한다. 이와 동일하게 인접 버텍스로부터 비인접 메시지를 더 이상 수신하지 않을 때까지 비인접 유사도의 시작 버텍스와 기준 버텍스 사이의 확장 유사도를 계산한다.

[0054] 이상 설명한 링크 예측 방법은 프리겔(Pregel)과 같은 분산 처리 프레임워크를 이용하여 수행될 수 있는데, 프리겔은 버텍스 단위로 기설계한 함수에 따라 버텍스의 값을 변경하거나 다른 버텍스에게 메시지를 보내는 슈퍼스텝이란 과정으로 동작한다. 프리겔은 활성화와 비활성화라는 두 가지의 버텍스 상태가 존재하는데 활성화 상태의 버텍스들은 사용자가 정의한 함수를 수행하며 연결된 버텍스에 메시지를 전달한다. 이때 더 이상 전송하는 메시지가 없을 때 비활성화 상태가 되며, 메시지를 전송받아 값이 변경되는 작업이 발생하지 않으면 이때 또한 비활성화 상태로 전이가 발생한다. 이러한 프리겔은 모든 버텍스가 비활성화 상태이거나 버텍스 간 메시지 전달이 발생하지 않을 때 종료하게 된다.

[0055] 도 9는 버텍스(V4)가 서로 다른 인접 버텍스(V5, V3)으로부터 동일한 시작 버텍스(V1) 식별자를 구비하는 비인접 메시지를 순차적으로 수신할 때마다 시작 버텍스(V1)와 버텍스(V4) 사이의 확장 유사도를 계산하는데, V1-V3-V4 사이의 확장 유사도는 0.0625이고 V1-V6-V5-V1 사이의 확장 유사도는 0.020625이므로, 더 큰 확장 유사도 0.0625를 유지한다.

[0056] 한편, 상술한 본 발명의 실시 예들은 컴퓨터에서 실행될 수 있는 프로그램으로 작성 가능하고, 컴퓨터로 읽을 수 있는 기록 매체를 이용하여 상기 프로그램을 동작시키는 범용 디지털 컴퓨터에서 구현될 수 있다.

[0057] 상기 컴퓨터로 읽을 수 있는 기록 매체는 전기 또는 자기식 저장 매체(예를 들어, 롬, 플로피 디스크, 하드디스크 등), 광학적 판독 매체(예를 들면, 시디롬, 디브이디 등) 및 캐리어 웨이브(예를 들면, 인터넷을 통한 전송)와 같은 저장 매체를 포함한다.

[0058] 본 발명은 도면에 도시된 실시예를 참고로 설명되었으나 이는 예시적인 것에 불과하며, 본 기술 분야의 통상의 지식을 가진 자라면 이로부터 다양한 변형 및 균등한 타 실시예가 가능하다는 점을 이해할 것이다. 따라서, 본 발명의 진정한 기술적 보호 범위는 첨부된 등록청구범위의 기술적 사상에 의해 정해져야 할 것이다.

부호의 설명

[0059] 100: 그래프 데이터베이스 200: 예측 모듈

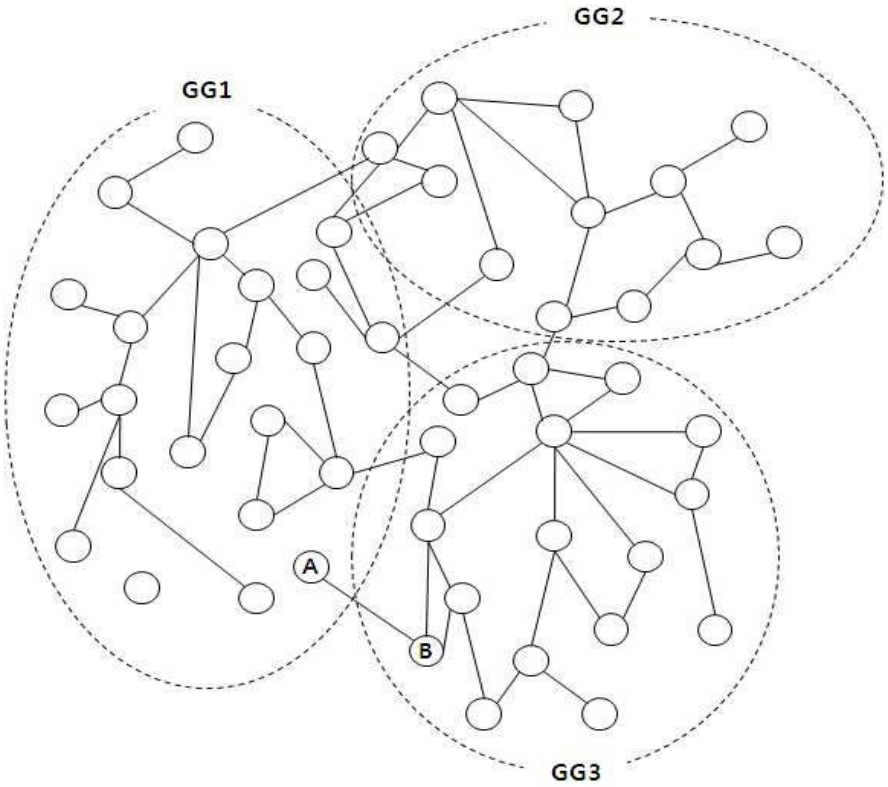
210: 에지차수 관리부 230: 인접 유사도 계산부

250: 저장부 270: 확장 유사도 계산부

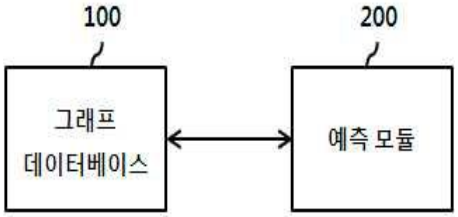
290: 예측부

도면

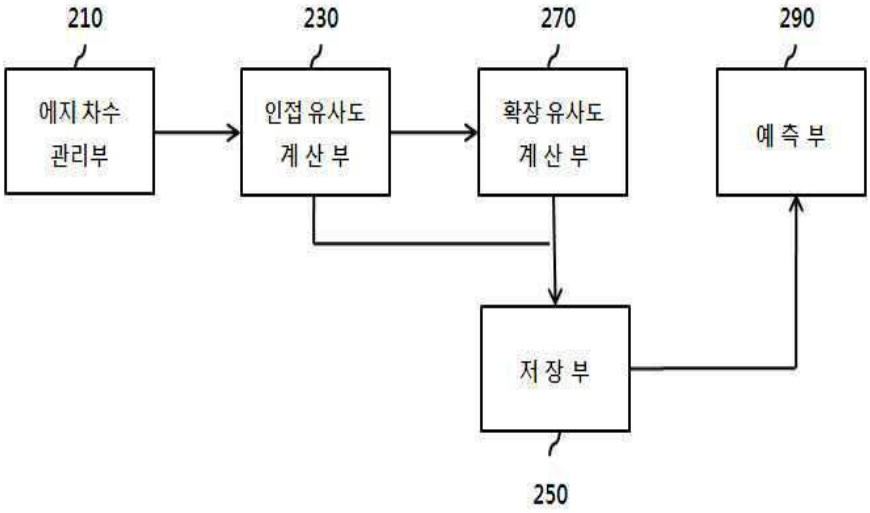
도면1



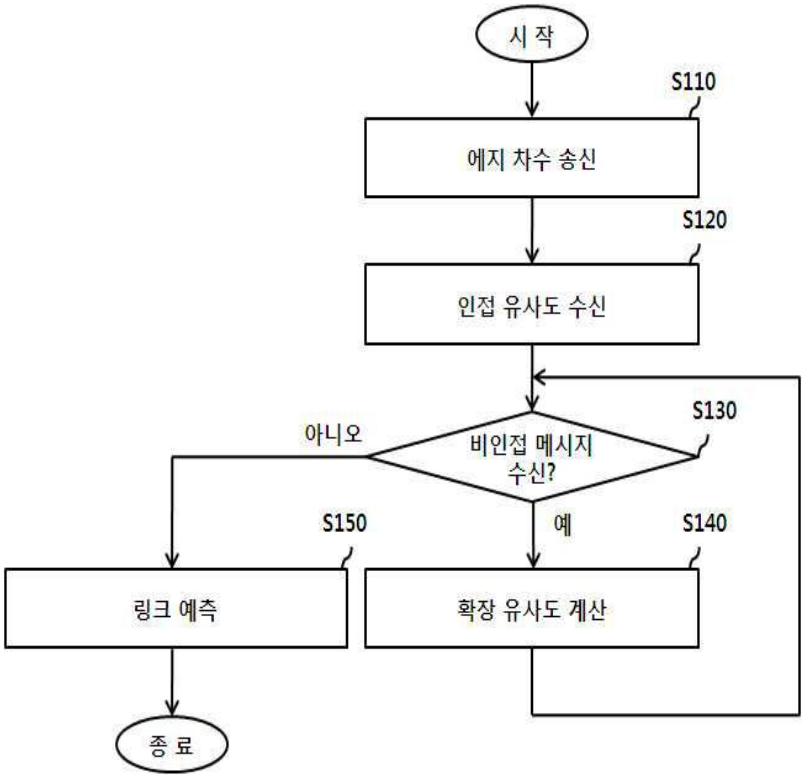
도면2



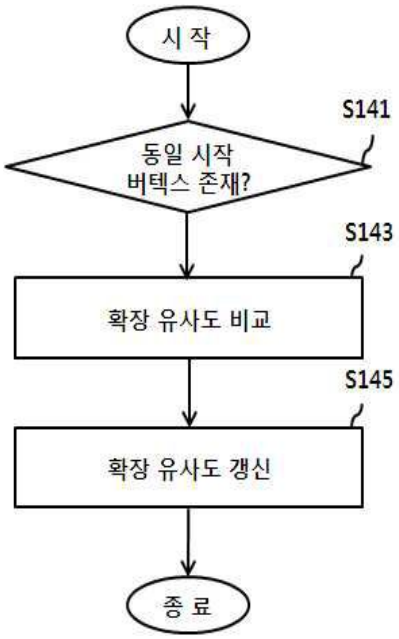
도면3


210
230
270
290
에지 차수 관리부
인접 유사도 계산부
확장 유사도 계산부
예측부
저장부
250

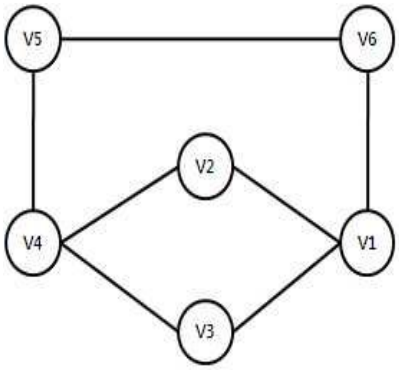
도면4


시작
S110
에지 차수 송신
S120
인접 유사도 수신
S130
아니오
비인접 메시지 수신?
예
S150
S140
링크 예측
확장 유사도 계산
종료

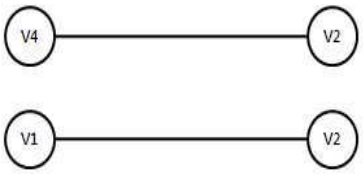
도면5


시 작
S141
동일 시작
버텍스 존재?
S143
확장 유사도 비교
S145
확장 유사도 갱신
종 료

도면6


V5
V6
V2
V4
V1
V3

도면7

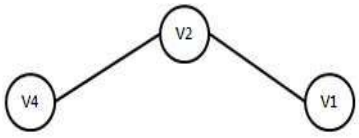


(a)

	V_1	V_2	V_3	V_4	V_5	V_6
V_1	1	0.25	0.25	0	0	0.25
V_2	0.25	1	0	0.25	0	0
V_3	0.25	0	1	0.25	0	0
V_4	0	0.25	0.25	1	0.25	0
V_5	0	0	0	0.25	1	0.33
V_6	0.25	0	0	0	0.33	1

(b)

도면8



(a)

	V_1	V_2	V_3	V_4	V_5	V_6
V_1	1	0.25	0.25	0.0625	0.0832	0.25
V_2	0.25	1	0.0625	0.25	0.0625	0
V_3	0.25	0.0625	1	0.25	0.0625	0.0625
V_4	0.0625	0.25	0.25	1	0.25	0.0832
V_5	0.0832	0.0625	0.0625	0.25	1	0.33
V_6	0.25	0.0625	0.0625	0.0832	0.33	1

(b)

도면9

