



(19) 대한민국특허청(KR)
(12) 등록특허공보(B1)

(45) 공고일자 2016년11월30일
(11) 등록번호 10-1681250
(24) 등록일자 2016년11월24일

(51) 국제특허분류(Int. Cl.)
G06Q 50/00 (2006.01) G06Q 50/30 (2012.01)
H04L 29/08 (2006.01) H04W 4/20 (2009.01)
(52) CPC특허분류
G06Q 50/01 (2013.01)
G06Q 50/30 (2015.01)
(21) 출원번호 10-2016-7003560
(22) 출원일자(국제) 2014년07월07일
심사청구일자 2016년02월11일
(85) 번역문제출일자 2016년02월11일
(65) 공개번호 10-2016-0021460
(43) 공개일자 2016년02월25일
(86) 국제출원번호 PCT/US2014/045516
(87) 국제공개번호 WO 2015/006180
국제공개일자 2015년01월15일
(30) 우선권주장
13/939,093 2013년07월10일 미국(US)
(56) 선행기술조사문헌
US8312056 B1

(73) 특허권자
페이스북, 인크.
미국, 캘리포니아 94025, 멘로 파크, 윌로우 로드 1601
(72) 발명자
말로 캐머론 알렉산더
미국 캘리포니아 94025 멘로 파크 윌로우 로드 1601
에클레스 딘
미국 캘리포니아 94025 멘로 파크 윌로우 로드 1601
(뒷면에 계속)
(74) 대리인
방해철, 김용인

전체 청구항 수 : 총 19 항

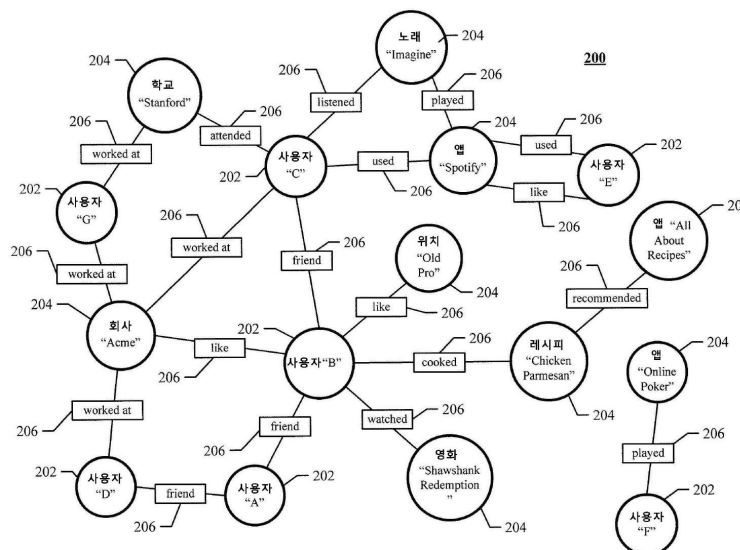
심사관 : 송미라

(54) 발명의 명칭 온라인 소셜 네트워크 내 네트워크-인지 제품 시판

(57) 요약

일실시예로, 본 방법은 복수의 노드 및 노드를 연결하는 복수의 에지를 포함하는 소셜 그래프에 액세스하는 단계; 그래프 클러스터링을 사용하여 소셜 그래프에서 복수의 클러스터를 식별하는 단계; 클러스터에 기반하여 제1 사용자 세트로 처리를 제공하는 단계; 및 각 사용자에게 대한 처리에의 네트워크 노출에 기반하여 제1 사용자 세트의 사용자들에 대한 처리의 처리 효과를 결정하는 단계를 포함하며, 각 노드는 온라인 소셜 네트워크의 사용자에게 대응한다.

대표도



(52) CPC특허분류

H04L 67/22 (2013.01)

H04W 4/206 (2013.01)

(72) 발명자

카리 브라이언

미국 캘리포니아 94025 멘로 파크 윌로우 로드
1601

우간더 요한

미국 캘리포니아 94025 멘로 파크 윌로우 로드
1601

백스트롬 라스 세렌

미국 캘리포니아 94025 멘로 파크 윌로우 로드
1601

클레인버그 존

미국 캘리포니아 94025 멘로 파크 윌로우 로드
1601

명세서

청구범위

청구항 1

하나 이상의 컴퓨팅 장치와 연관된 하나 이상의 프로세서에 의해;

복수의 노드 및 노드를 연결하는 복수의 에지를 포함하는 소셜 그래프에 액세스하는 단계;

그래프 클러스터링(clustering)을 사용하여 소셜 그래프에서 복수의 노드로부터의 별개의 노드 세트를 각각 포함하는 복수의 비-중첩(non-overlapping) 클러스터를 식별하는 단계;

복수의 클러스터 중 제1 클러스터 세트 및 제2 클러스터 세트에 각각 속하는 적어도 제1 사용자 세트 및 제2 사용자 세트로 처리(treatment)를 제공하는 단계; 및

적어도 제1 사용자 세트 및 제2 사용자 세트 각각에 대해, 각 사용자에게 대한 처리에의 네트워크 노출에 기반하여 사용자 세트의 사용자에게 대한 처리의 처리 효과를 결정하는 단계를 포함하며,

2개의 노드 사이의 각각의 에지는 노드 사이의 이격도 1도를 나타내고,

복수의 노드는 온라인 소셜 네트워크의 복수의 사용자에게 각각 대응하며,

제1 클러스터 세트는 제2 클러스터 세트와 별개이고,

각각의 개별 클러스터에 대해, 클러스터 내 노드의 네트워크 노출은 절대 k-이웃 노출, 절대 k-코어 노출, 부분 q-이웃 노출 또는 부분 q-코어 노출이며,

여기서, $d \geq k$ 인 차수 d의 노드 i는 i 및 i의 차수 k 이상의 이웃이 처리를 수신하면 그 처리에 대한 절대 k-이웃 노출(absolute k-neighborhood exposure)을 경험하고,

$d \geq k$ 인 차수를 가진 노드 i는 i가 처리를 수신하는 노드 V' 에서 유도된 G의 서브그래프인 $G[V']$ 그래프의 k-코어에 속한다면 그 처리에 대한 절대 k-코어 노출(absolute k-core exposure)을 경험하며,

차수 d의 노드 i는 i 및 i의 차수 q 이상인 이웃이 처리를 수신하면 그 처리에 대한 부분 q-이웃 노출(fractional q-neighborhood exposure)을 경험하고,

노드 i는 i가 처리를 수신하는 노드 V' 에서 유도된 G의 서브그래프인 $G[V']$ 그래프의 부분 q-코어에 속한다면 그 처리에 대한 부분 q-코어 노출(fractional q-core exposure)을 경험하는 방법.

청구항 2

제 1 항에 있어서,

처리 효과는 사용자에게 대한 처리의 네트워크 효과(network effect)와 사용자에게 대한 처리의 개별 효과(individual effect)의 함수인 방법.

청구항 3

제 1 항에 있어서,

처리의 결정된 처리 효과에 기반하여 처리를 수정하는 단계를 더 포함하는 방법.

청구항 4

제 1 항에 있어서,

각 클러스터에 대하여, 클러스터에서 노드의 네트워크 노출은 네트워크 노출의 임계 레벨에서 네트워크 노출의 최대 레벨까지의 특정된 분포를 가지는 방법.

청구항 5

제 4 항에 있어서,

처리의 처리 효과를 결정하는 단계는 특정된 분포 범위의 각기 다른 레벨의 네트워크 노출에 대한 처리의 처리 효과를 결정하는 단계를 포함하는 방법.

청구항 6

제 1 항에 있어서,

각 클러스터에 대하여, 클러스터에서 노드의 네트워크 노출은 완전 이웃 노출(full neighborhood exposure) 또는 컴포넌트 노출(component exposure)인 방법.

청구항 7

제 1 항에 있어서,

각 클러스터에 대하여, 클러스터 내 다른 노드들에 대한 클러스터 내 노드들의 소셜-그래프 친밀성은 임계 소셜-그래프 친밀성보다 더 큰 방법.

청구항 8

제 1 항에 있어서,

노드의 이격도 1도 내에 있는 노드들의 임계 부분들이 동일한 처리 조건에 있다면 특정 클러스터 내 노드가 네트워크-노출되는 방법.

청구항 9

제 1 항에 있어서,

임계 수의 노드는 모든 노드인 방법.

청구항 10

제 1 항에 있어서,

처리를 제공하는 단계는 복수의 클러스터 중 처리 클러스터와 제어(control) 클러스터 사이에서 랜덤화하는 단계를 포함하는 방법.

청구항 11

제 1 항에 있어서,

처리는 특정 광고인 방법.

청구항 12

제 1 항에 있어서,

처리는 제3자 시스템의 특정 제품이나 피처(feature)인 방법.

청구항 13

제 1 항에 있어서,

처리는 온라인 소셜 네트워크의 특정 제품이나 피처인 방법.

청구항 14

제 1 항에 있어서,

복수의 클러스터를 식별하는 단계는 각 클러스터 내 임계 수의 노드가 클러스터 내 다른 노드들에 대해 네트워

크-노출되도록 클러스터들을 식별하는 단계를 포함하는 방법.

청구항 15

복수의 노드 및 노드를 연결하는 복수의 에지를 포함하는 소셜 그래프에 액세스하고;

그래프 클러스터링을 사용하여 소셜 그래프에서 복수의 노드로부터의 별개의 노드 세트를 각각 포함하는 복수의 비-중첩 클러스터를 식별하며;

복수의 클러스터 중 제1 클러스터 세트 및 제2 클러스터 세트에 각각 속하는 적어도 제1 사용자 세트 및 제2 사용자 세트로 처리를 제공하고;

적어도 제1 사용자 세트 및 제2 사용자 세트 각각에 대해, 각 사용자에게 처리에의 네트워크 노출에 기반하여 사용자 세트의 사용자들에 대한 처리의 처리 효과를 결정하도록 실행될 때 동작하는 소프트웨어를 수록하는 하나 이상의 컴퓨터-판독가능한 비-일시적 저장매체로서,

2개의 노드 사이의 각각의 에지는 노드 사이의 이격도 1도를 나타내고,

복수의 노드는 온라인 소셜 네트워크의 복수의 사용자에게 각각 대응하며,

제1 클러스터 세트는 제2 클러스터 세트와 별개이고,

각각의 개별 클러스터에 대해, 클러스터 내 노드의 네트워크 노출은 절대 k-이웃 노출, 절대 k-코어 노출, 부분 q-이웃 노출 또는 부분 q-코어 노출이며,

여기서, $d \geq k$ 인 차수 d의 노드 i는 i 및 i의 차수 k 이상의 이웃이 처리를 수신하면 그 처리에 대한 절대 k-이웃 노출(absolute k-neighborhood exposure)을 경험하고,

$d \geq k$ 인 차수를 가진 노드 i는 i가 처리를 수신하는 노드 V' 에서 유도된 G의 서브그래프인 $G[V']$ 그래프의 k-코어에 속한다면 그 처리에 대한 절대 k-코어 노출(absolute k-core exposure)을 경험하며,

차수 d의 노드 i는 i 및 i의 차수 qd 이상의 이웃이 처리를 수신하면 그 처리에 대한 부분 q-이웃 노출(fractional q-neighborhood exposure)을 경험하고,

노드 i는 i가 처리를 수신하는 노드 V' 에서 유도된 G의 서브그래프인 $G[V']$ 그래프의 부분 q-코어에 속한다면 그 처리에 대한 부분 q-코어 노출(fractional q-core exposure)을 경험하는 컴퓨터-판독가능한 비-일시적 저장매체.

청구항 16

하나 이상의 프로세서; 및

프로세서와 연결되고 프로세서에 의해 실행가능한 명령어를 포함하는 메모리를 포함하는 시스템으로서,

상기 프로세서는:

복수의 노드 및 노드를 연결하는 복수의 에지를 포함하는 소셜 그래프에 액세스하고;

그래프 클러스터링을 사용하여 소셜 그래프에서 복수의 노드로부터의 별개의 노드 세트를 각각 포함하는 복수의 비-중첩 클러스터를 식별하며;

복수의 클러스터 중 제1 클러스터 세트 및 제2 클러스터 세트에 각각 속하는 적어도 제1 사용자 세트 및 제2 사용자 세트로 처리를 제공하고;

적어도 제1 사용자 세트 및 제2 사용자 세트 각각에 대해, 각 사용자에게 처리에의 네트워크 노출에 기반하여 사용자 세트의 사용자들에 대한 처리의 처리 효과를 결정하는 명령어를 실행할 때 동작하며,

2개의 노드 사이의 각각의 에지는 노드 사이의 이격도 1도를 나타내고,

복수의 노드는 온라인 소셜 네트워크의 복수의 사용자에게 각각 대응하며,

제1 클러스터 세트는 제2 클러스터 세트와 별개이고,

각각의 개별 클러스터에 대해, 클러스터 내 노드의 네트워크 노출은 절대 k-이웃 노출, 절대 k-코어 노출, 부분 q-이웃 노출 또는 부분 q-코어 노출이며,

여기서, $d \geq k$ 인 차수 d의 노드 i는 i 및 i의 차수 k 이상의 이웃이 처리를 수신하면 그 처리에 대한 절대 k-이웃 노출(absolute k-neighborhood exposure)을 경험하고,

$d \geq k$ 인 차수를 가진 노드 i는 i가 처리를 수신하는 노드 V'에서 유도된 G의 서브그래프인 $G[V']$ 그래프의 k-코어에 속한다면 그 처리에 대한 절대 k-코어 노출(absolute k-core exposure)을 경험하며,

차수 d의 노드 i는 i 및 i의 차수 qd 이상의 이웃이 처리를 수신하면 그 처리에 대한 부분 q-이웃 노출(fractional q-neighborhood exposure)을 경험하고,

노드 i는 i가 처리를 수신하는 노드 V'에서 유도된 G의 서브그래프인 $G[V']$ 그래프의 부분 q-코어에 속한다면 그 처리에 대한 부분 q-코어 노출(fractional q-core exposure)을 경험하는 시스템.

청구항 17

제 1 항에 있어서,

특정 클러스터 내 사용자들에 대한 처리의 결정된 처리 효과는 평균 처리 효과(τ)이고,

$$\tau(\bar{z} = \bar{1}, \bar{z}' = \bar{0}) = \frac{1}{n} \sum_{i=1}^n [Y_i(\bar{z} = \bar{1}) - Y_i(\bar{z} = \bar{0})] \quad \text{이며,}$$

$\bar{z} = \bar{1}$ 은 처리 그룹 내 사용자를 나타내고,

$\bar{z}' = \bar{0}$ 은 제어 그룹 내 사용자를 나타내며,

n은 특정 클러스터 내 노드의 수이고,

$Y_i(\bar{z})$ 은 처리 할당 벡터 \bar{z} 하에서 사용자 i의 잠재적 결과인 방법.

청구항 18

제 1 항에 있어서,

2개의 랜덤화된 클러스터들 사이의 추정된 평균 처리 효과($\hat{\tau}$)를 결정하는 단계를 더 포함하며,

$$\hat{\tau}(Z) = \frac{1}{n} \sum_{i=1}^n \left(\frac{Y_i(Z)1[Z \in \sigma_0^i]}{\Pr(Z \in \sigma_0^i)} - \frac{Y_i(Z)1[Z \in \sigma_1^i]}{\Pr(Z \in \sigma_1^i)} \right)$$

이고,

Z 는 $\{0, 1\}^n$ 의 값들을 포함하는 랜덤 벡터를 나타내며,

n은 특정 클러스터 내 노드의 수이고,

σ_1^i 은 $\bar{z} = \bar{1}$ 의 노출 조건을 나타내며,

$\bar{z} = \bar{1}$ 은 처리 그룹 내 사용자를 나타내고,

σ_0^i 은 $\bar{z} = \bar{0}$ 의 노출 조건을 나타내며,

$\bar{Z}' = \bar{0}$ 은 제어 그룹 내 사용자를 나타내고,

$Y_i(Z)$ 은 처리 할당 벡터 Z 하에서 사용자 i 의 잠재적 결과이며,

$\Pr(Z \in \sigma_1^i)$ 은 처리에 대한 네트워크 노출의 확률이고,

$\Pr(Z \in \sigma_0^i)$ 은 제어에 대한 네트워크 노출의 확률이며,

i 는 노드와 연관된 사용자를 나타내는 방법.

청구항 19

제 1 항에 있어서,

각각의 식별된 클러스터에 대하여, 각 클러스터 내 임계 수의 노드는 클러스터 내 다른 노드들에 대해 네트워크 노출되는 방법.

발명의 설명

기술 분야

[0001] 본 명세서는 일반적으로 온라인 소셜 네트워크 및 온라인 환경에서의 피처 또는 제품의 개시에 관한 것이다.

배경 기술

[0002] 소셜 네트워킹 웹사이트를 포함할 수 있는 소셜 네트워킹 시스템은 그 사용자들(가령, 사람들 또는 단체)이 소셜 네트워킹 시스템과 상호작용하고 소셜 네트워킹 시스템을 통해 서로 상호작용할 수 있게 해줄 수 있다. 소셜 네트워킹 시스템은 사용자로부터의 입력으로 사용자와 관련된 사용자 프로필을 생성하고 소셜 네트워킹 시스템에 저장할 수 있다. 사용자 프로필은 신상 정보, 통신-채널 정보 및 사용자의 개인 관심사에 대한 정보를 포함할 수 있다. 또한, 소셜 네트워킹 시스템은 사용자로부터의 입력으로 소셜 네트워킹 시스템의 다른 사용자들과의 사용자의 관계의 레코드를 생성하고 저장할 수 있을 뿐만 아니라, 서비스(예컨대, 담벼락-게시물, 사진 공유, 이벤트 조직, 메시징, 게임 또는 광고)를 제공하여 사용자들 사이의 소셜 상호작용을 용이하게 할 수 있다.

[0003] 소셜 네트워킹 시스템은 하나 이상의 네트워크를 통해 서비스에 관한 콘텐츠 또는 메시지를 사용자의 모바일이나 다른 컴퓨팅 장치로 송신할 수 있다. 또한, 사용자는 소셜 네트워킹 시스템 내 사용자의 사용자 프로필과 다른 데이터에 접근하기 위해 소프트웨어 애플리케이션을 사용자의 모바일이나 다른 컴퓨팅 장치에 설치할 수 있다. 소셜 네트워킹 시스템은 사용자에게 디스플레이하기 위해 가령 사용자와 연결된 다른 사용자들의 집계된 소식의 뉴스피드와 같이 개인화된 세트의 콘텐츠 객체를 생성할 수 있다.

[0004] A/B 테스트법은 온라인 실험(online experiments)의 효과를 평가하기 위한 표준 접근법이다. 일반적으로, A/B 테스트법은 (통상 A 및 B라고 하는) 2개의 변량(variants)을 가진 랜덤화된 실험을 사용하며, 이런 2개의 변량은 실험에서 제어(control) 및 처리(treatment)를 나타낸다. 이런 실험은 웹 개발 및 마케팅에서뿐 아니라 더 종래의 형태의 광고에서 공통으로 사용된다. 온라인 설정에서, 목적은 통상적으로 온라인 환경에 대한 변화가 환경에서의 사용자 행동에 어떻게 영향을 주는지를 결정하기 위함이다. 제어 그룹 및 처리 그룹이 비교되며, 사용자의 행동에 영향을 줄 수 있는 하나의 변량을 제외하고 일반적으로 동일하다.

발명의 내용

해결하려는 과제

[0005] 특정 실시예로, 소셜 네트워킹 시스템은 새로운 제품, 피처, 광고 또는 경험("처리")이 특정 사용자 클러스터로 그 개시를 스테거링(staggering)하여 온라인 소셜 네트워크의 사용자에게 개시되도록 하는 방법을 향상시킬 수 있다. 표준 A/B 테스트 접근법을 사용하여 새로운 처리를 테스트하는 대신에, 테스트 및 시판(rollout)은 네트워크-인지 방식으로 이루어질 수 있다. 예컨대, 별개의 사용자 클러스터가 적절한 클러스터링 알고리즘을 사용하여 식별될 수 있다. 이런 사용자 클러스터는 특정 클러스터 내 사용자들이 동일한 클러스터 내 다른 사용자들

에 대해 네트워크 노출되도록 정의될 수 있다. 이를 통해 처리 효과는 더 효과적으로 테스트될 수 있는데, 이는 특정 사용자 및 그 사용자의 대부분의 친구들이 처리를 수신할 것이라고 보장할 수 있기 때문이다. 이는 또 다른 사용자와 함께 사용될 수 있을 때(예컨대, 온라인 채팅 애플리케이션) 더 유용한 처리에 대해 특히 이점적일 수 있다. 사용자 클러스터가 식별된 후, 처리는 사용자 클러스터의 제1 세트에게 제공되고 그들과 함께 테스트될 수 있다. 이후, 이런 사용자 클러스터의 제1 세트로부터 수신된 데이터는 이런 사용자들에 대한 처리의 처리 효과를 결정하도록 분석될 수 있다. 처리는 사용자 클러스터의 다음 세트에 시판되기 전에 수정될 수 있다. 이런 제품을 시판하는 순차적인 접근법은 제품이 모든 사용자에게 개시될 때까지 시행될 수 있다.

과제의 해결 수단

[0006] 일 실시예로, 본 방법은 복수의 노드 및 노드를 연결하는 복수의 에지를 포함하는 소셜 그래프에 액세스하는 단계; 그래프 클러스터링을 사용하여 소셜 그래프에서 복수의 클러스터를 식별하는 단계; 클러스터에 기반하여 제1 사용자 세트로 처리를 제공하는 단계; 및 각 사용자에게 대한 처리에의 네트워크 노출에 기반하여 제1 사용자 세트의 사용자들에 대한 처리의 처리 효과를 결정하는 단계를 포함하며, 각 노드는 온라인 소셜 네트워크의 사용자에게 대응한다.

발명의 효과

[0007] 본 발명의 내용 중에 포함되어 있다.

도면의 간단한 설명

[0008] 도 1은 소셜 네트워킹 시스템과 연관된 예시적인 네트워크 환경을 도시한다.

도 2는 예시적인 소셜 그래프를 도시한다.

도 3은 소셜 그래프 내 클러스터의 예를 도시한다.

도 4는 예시적인 확률 분포를 도시한다.

도 5는 예시적인 사이클 그래프 및 분산 계산 도표를 도시한다.

도 6은 처리 효과를 결정하는 예시적인 방법을 도시한다.

도 7은 예시적인 컴퓨터 시스템을 도시한다.

발명을 실시하기 위한 구체적인 내용

[0009] 시스템 개요

[0010] 도 1은 소셜 네트워킹 시스템과 관련된 예시적인 네트워크 환경(100)을 도시한다. 네트워크 환경(100)은 네트워크(110)에 의해 서로 연결되는 클라이언트 시스템(130), 소셜 네트워킹 시스템(160) 및 제3자 시스템(170)을 포함한다. 도 1은 클라이언트 시스템(130), 소셜 네트워킹 시스템(160), 제3자 시스템(170) 및 네트워크(110)의 특정한 구성을 도시하지만, 본 명세서는 클라이언트 시스템(130), 소셜 네트워킹 시스템(160), 제3자 시스템(170) 및 네트워크(110)의 임의의 적절한 구성을 고려한다. 예로서 제한 없이, 2 이상의 클라이언트 시스템(130), 소셜 네트워킹 시스템(160) 및 제3자 시스템(170)은 네트워크(110)를 우회하여 서로 직접적으로 연결될 수 있다. 또 다른 예로서, 2 이상의 클라이언트 시스템(130), 소셜 네트워킹 시스템(160) 및 제3자 시스템(170)은 전체적으로 또는 부분적으로 서로 물리적으로 또는 논리적으로 같은 장소에 배치될 수 있다. 게다가, 도 1은 특정한 수의 클라이언트 시스템(130), 소셜 네트워킹 시스템(160), 제3자 시스템(170) 및 네트워크(110)를 도시하지만, 본 명세서는 임의의 적절한 수의 클라이언트 장치(130), 소셜 네트워킹 시스템(160), 제3자 시스템(170) 및 네트워크(110)를 고려한다. 예로서 제한 없이, 네트워크 환경(100)은 다수의 클라이언트 시스템(130), 소셜 네트워킹 시스템(160), 제3자 시스템(170) 및 네트워크(110)를 포함할 수 있다.

[0011] 본 명세서는 임의의 적절한 네트워크(110)를 고려한다. 예로서 제한 없이, 네트워크(110)의 하나 이상의 부분은 애드 혹 네트워크, 인트라넷, 엑스트라넷, 가상 사설 네트워크(VPN), 근거리 네트워크(LAN), 무선 LAN(WLAN), 광역 네트워크(WAN), 무선 WAN(WWAN), 대도시 네트워크(MAN), 인터넷의 일부, 공중 전화망(PSTN)의 일부, 셀룰러 전화 네트워크 또는 2 이상의 이들의 조합을 포함할 수 있다. 네트워크(110)는 하나 이상의 네트워크(110)를 포함할 수 있다.

- [0012] 링크(150)는 클라이언트 시스템(130), 소셜 네트워킹 시스템(160) 및 제3자 시스템(170)을 통신 네트워크(110)와 연결하거나 서로 연결할 수 있다. 본 명세서는 임의의 적절한 링크(150)를 고려한다. 특정 실시예로, 하나 이상의 링크(150)는 하나 이상의 유선라인(가령, 예컨대, 디지털 가입자 라인(DSL) 또는 DOCSIS(Data Over Cable Service Interface Specification)), 무선(가령, 예컨대, Wi-Fi 또는 WiMAX(Worldwide Interoperability for Microwave Access)) 또는 광학(가령, 예컨대, 동기식 광학 네트워크(Synchronous Optical Network, SONET) 또는 동기식 디지털 계층(Synchronous Digital Hierarchy, SDH)) 링크를 포함한다. 특정 실시예로, 하나 이상의 링크(150)는 애드 혹 네트워크, 인트라넷, 엑스트라넷, VPN, LAN, WLAN, WAN, WWAN, MAN, 인터넷의 일부, PSTN의 일부, 셀룰러 기술-기반 네트워크, 위성 통신 기술-기반 네트워크, 또 다른 링크(150) 또는 2 이상의 이런 링크(150)의 조합을 각각 포함한다. 링크(150)는 네트워크 환경(100)을 통해 반드시 동일할 필요는 없다. 하나 이상의 제1 링크(150)는 하나 이상의 측면에서 하나 이상의 제2 링크(150)와 다를 수 있다.
- [0013] 특정 실시예로, 클라이언트 시스템(130)은 하드웨어, 소프트웨어 또는 임베디드 로직 컴포넌트나 이들의 2 이상의 컴포넌트의 조합을 포함하고 클라이언트 시스템(130)에 의해 구현되거나 지원되는 적절한 기능을 수행할 수 있는 전자식 장치일 수 있다. 예로서 제한 없이, 클라이언트 시스템(130)은 가령 데스크톱 컴퓨터, 노트북이나 랩톱 컴퓨터, 넷북, 태블릿 컴퓨터, e-북 리더, GPS 장치, 카메라, 개인용 정보 단말기(PDA), 휴대용 전자 장치, 셀룰러 전화, 스마트폰, 다른 적절한 전자 장치 또는 이들의 임의의 적절한 조합과 같은 컴퓨터 시스템을 포함할 수 있다. 본 명세서는 임의의 적절한 클라이언트 시스템(130)을 고려한다. 클라이언트 시스템(130)은 클라이언트 시스템(130)에서 네트워크 사용자가 네트워크(110)에 접근할 수 있게 할 수 있다. 클라이언트 시스템(130)은 그 사용자가 다른 클라이언트 시스템(130)의 다른 사용자와 통신할 수 있게 할 수 있다.
- [0014] 특정 실시예로, 클라이언트 시스템(130)은, 가령 MICROSOFT INTERNET EXPLORER, GOOGLE CHROME 또는 MOZILLA FIREFOX와 같은 웹 브라우저(132)를 포함할 수 있고, 가령 TOOLBAR 또는 YAHOO TOOLBAR와 같은 하나 이상의 애드-온(add-ons), 플러그-인(plug-ins) 또는 다른 확장형(extensions)을 가질 수 있다. 클라이언트 시스템(130)에서의 사용자는 URL(Uniform Resource Locator) 또는 웹 브라우저(132)를 특정 서버(가령, 서버(162) 또는 제3자 시스템(170)과 관련된 서버)로 인도하는 다른 주소를 입력할 수 있고, 웹 브라우저(132)는 하이퍼텍스트 전송 프로토콜(HTTP) 요청을 생성하고 HTTP 요청을 서버로 통신할 수 있다. 서버는 HTTP 요청을 수락하고, HTTP 요청에 응답하여 하나 이상의 하이퍼텍스트 마크업 언어(HTML) 파일을 클라이언트 시스템(130)으로 통신할 수 있다. 클라이언트 시스템(130)은 사용자에게 표시하기 위해 서버로부터 HTML 파일에 기초한 웹페이지를 렌더링할 수 있다. 본 명세서는 임의의 적절한 웹페이지 파일을 고려한다. 예로서 제한 없이, 웹페이지는 특정한 필요에 따라 HTML 파일, 확장형 하이퍼텍스트 마크업 언어(XHTML) 파일 또는 확장형 마크업 언어(XML) 파일로부터 렌더링할 수 있다. 또한, 이런 페이지는, 예로서 제한 없이 JAVASCRIPT, JAVA, MICROSOFT SILVERLIGHT, 가령 AJAX(비동기식 JAVASCRIPT 및 XML)와 같은 마크업 언어와 스크립트의 조합 등과 같은 스크립트를 실행할 수 있다. 본 명세서에서, 웹페이지에 대한 레퍼런스는 (브라우저가 웹페이지를 렌더링하는데 사용할 수 있는) 하나 이상의 해당 웹페이지 파일을 포함하며, 적절한 경우, 그 역도 또한 같다.
- [0015] 특정 실시예로, 소셜 네트워킹 시스템(160)은 온라인 소셜 네트워크를 호스팅할 수 있는 네트워크-주소화 컴퓨팅 시스템일 수 있다. 소셜 네트워킹 시스템(160)은 예컨대 사용자-프로필 데이터, 컨셉-프로필 데이터, 소셜-그래프 정보 또는 온라인 소셜 네트워크에 관한 다른 적절한 데이터와 같은 소셜 네트워킹 데이터를 생성, 저장, 수신 및 송신할 수 있다. 소셜 네트워킹 시스템(160)은 직접적으로 또는 네트워크(110)를 통해 네트워크 환경(100)의 다른 컴포넌트들에 의해 접근될 수 있다. 특정 실시예로, 소셜 네트워킹 시스템(160)은 하나 이상의 서버(162)를 포함할 수 있다. 각 서버(162)는 일체형 서버(unitary server)일 수 있거나, 다수의 컴퓨터 또는 다수의 데이터센터에 걸쳐 있는 분산형 서버일 수 있다. 서버(162)는 예로서 제한 없이, 웹 서버, 뉴스 서버, 메일 서버, 메시지 서버, 광고 서버, 파일 서버, 애플리케이션 서버, 교환 서버, 데이터베이스 서버, 프록시 서버, 본 명세서에 기술된 기능이나 프로세스를 수행하는데 적절한 또 다른 서버 또는 이들의 임의의 조합과 같이, 다양한 타입일 수 있다. 특정 실시예로, 각 서버(162)는 서버(162)에 의해 구현되거나 지원되는 적절한 기능을 수행하기 위한 하드웨어, 소프트웨어 또는 임베디드 논리 소자 또는 2 이상의 이런 소자들의 조합을 포함할 수 있다. 특정 실시예로, 소셜 네트워킹 시스템(164)은 하나 이상의 데이터 스토어(164)를 포함할 수 있다. 데이터 스토어(164)는 다양한 타입의 정보를 저장하는데 사용될 수 있다. 특정 실시예로, 데이터 스토어(164)에 저장된 정보는 특정한 데이터 구조에 따라 구조화될 수 있다. 특정 실시예로, 각 데이터 스토어(164)는 관계형, 컬럼형, 상관계형 또는 다른 적절한 데이터베이스일 수 있다. 비록 본 명세서는 특정 타입의 데이터베이스를 기술하거나 도시하지만, 본 명세서는 임의의 적절한 타입의 데이터베이스를 고려한다. 특정 실시예는 클라

이언트 시스템(130), 소셜 네트워킹 시스템(160) 또는 제3자 시스템(170)이 데이터 스토어(164)에 저장된 정보를 관리, 검색, 변경, 추가 또는 삭제할 수 있게 하는 인터페이스를 제공할 수 있다.

[0016] 특정 실시예로, 소셜 네트워킹 시스템(160)은 하나 이상의 데이터 스토어(164)에 하나 이상의 소셜 그래프를 저장할 수 있다. 특정 실시예로, 소셜 그래프는 (특정 사용자에게 각각 해당하는) 다수의 사용자 노드 또는 (특정 컨셉에 각각 해당하는) 다수의 컨셉 노드를 포함할 수 있는 다수의 노드 및 노드를 연결하는 다수의 에지를 포함할 수 있다. 소셜 네트워킹 시스템(160)은 다른 사용자와 통신하고 상호작용하는 능력을 온라인 소셜 네트워크의 사용자에게 제공할 수 있다. 특정 실시예로, 사용자는 소셜 네트워킹 시스템(160)을 통해 온라인 소셜 네트워크에 가입한 후 연결들(즉, 관계들)을 그들이 연결되고자 하는 소셜 네트워킹 시스템(160)의 다수의 다른 사용자에게 추가할 수 있다. 본 명세서에서, "친구"란 용어는 사용자가 소셜 네트워킹 시스템(160)을 통해 연결, 유대 또는 관계를 형성했던 소셜 네트워킹 시스템(160)의 임의의 다른 사용자를 가리킬 수 있다.

[0017] 특정 실시예로, 소셜 네트워킹 시스템(160)은 소셜 네트워킹 시스템(160)에 의해 지원되는 다양한 유형의 아이템이나 객체에 대한 행위를 취할 수 있는 능력을 사용자에게 제공할 수 있다. 예로서 제한 없이, 아이템 및 객체는 소셜 네트워킹 시스템(160)의 사용자들이 속할 수 있는 그룹이나 소셜 네트워크, 사용자가 관심을 가질지도 모르는 이벤트 또는 캘린더 엔트리, 사용자가 사용할 수 있는 컴퓨터-기반 애플리케이션, 사용자가 서비스를 통해 아이템을 팔거나 구매할 수 있게 하는 거래, 사용자가 수행할 수 있는 광고와의 상호작용 또는 다른 적절한 아이템이나 객체를 포함할 수 있다. 사용자는 소셜 네트워킹 시스템(160) 또는 소셜 네트워킹 시스템(160)과 별개이거나 네트워크(110)를 통해 소셜 네트워킹 시스템(160)에 결합되는 제3자 시스템(170)의 외부 시스템에서 표현될 수 있는 모든 것과 상호작용할 수 있다.

[0018] 특정 실시예로, 소셜 네트워킹 시스템(160)은 다양한 엔티티를 링크할 수 있다. 예로서 제한 없이, 소셜 네트워킹 시스템(160)은 사용자가 서로 상호작용할 뿐 아니라 제3자 시스템(170)이나 다른 엔티티로부터 콘텐츠를 수신할 수 있게 하거나, 사용자가 응용 프로그래밍 인터페이스(API)나 다른 통신 채널을 통해 이런 엔티티와 상호작용할 수 있게 해줄 수 있다.

[0019] 특정 실시예로, 제3자 시스템(170)은 하나 이상의 타입의 서버, 하나 이상의 데이터 스토어, API들을 포함하나 이에 국한되지 않는 하나 이상의 인터페이스, 하나 이상의 웹 서비스, 하나 이상의 콘텐츠 소스, 하나 이상의 네트워크 또는 예컨대 서버가 통신할 수 있는 임의의 다른 적절한 컴포넌트를 포함할 수 있다. 제3자 시스템(170)은 소셜 네트워킹 시스템(160)을 운영하는 엔티티와는 다른 엔티티에 의해 운영될 수 있다. 그러나, 특정 실시예로, 소셜 네트워킹 시스템(160) 및 제3자 시스템(170)은 소셜 네트워킹 서비스를 소셜 네트워킹 시스템(160)이나 제3자 시스템(170)의 사용자에게 제공하도록 서로 함께 동작할 수 있다. 이런 의미에서, 소셜 네트워킹 시스템(160)은 가령 제3자 시스템(170)과 같은 다른 시스템들이 인터넷상의 사용자들에게 소셜 네트워킹 서비스와 기능을 제공하도록 사용할 수 있는 플랫폼 또는 백본(backbone)을 제공할 수 있다.

[0020] 특정 실시예로, 제3자 시스템(170)은 제3자 콘텐츠 객체 제공자를 포함할 수 있다. 제3자 콘텐츠 객체 제공자는 클라이언트 시스템(130)과 통신될 수 있는 하나 이상의 소스의 콘텐츠 객체를 포함할 수 있다. 예로서 제한 없이, 콘텐츠 객체는 가령 영화 쇼타임, 영화 리뷰, 레스토랑 리뷰, 레스토랑 메뉴, 제품 정보와 리뷰 또는 다른 적절한 정보와 같이 사용자가 관심 있는 사물이나 활동에 관한 정보를 포함할 수 있다. 또 다른 예로서 제한 없이, 콘텐츠 객체는 가령 쿠폰, 할인 티켓, 상품권 또는 다른 적절한 인센티브 객체와 같은 인센티브 콘텐츠 객체를 포함할 수 있다.

[0021] 특정 실시예로, 소셜 네트워킹 시스템(160)은 또한 소셜 네트워킹 시스템(160)과 사용자의 상호작용을 향상시킬 수 있는 사용자-생성된 콘텐츠 객체를 포함한다. 사용자-생성된 콘텐츠는 사용자가 소셜 네트워킹 시스템(160)에 추가, 업로드, 송신 또는 "게시"할 수 있는 어떤 것을 포함할 수 있다. 예로서 제한 없이, 사용자는 클라이언트 시스템(130)으로부터 소셜 네트워킹 시스템(160)으로 게시물을 통신할 수 있다. 게시물은 가령 상태 업데이트나 다른 텍스트형 데이터와 같은 데이터, 위치 정보, 사진, 비디오, 링크, 음악 또는 다른 유사한 데이터나 매체를 포함할 수 있다. 또한, 콘텐츠는 가령 뉴스피드 또는 스트림과 같이 "통신 채널"을 통해 제3자에 의해 소셜 네트워킹 시스템(160)에 추가될 수 있다.

[0022] 특정 실시예로, 소셜 네트워킹 시스템(160)은 다양한 서버, 서브-시스템, 프로그램, 모듈, 로그 및 데이터 스토어를 포함할 수 있다. 특정 실시예로, 소셜 네트워킹 시스템(160)은 다음의 하나 이상의: 웹 서버, 행위 로거, API 요청 서버, 관련성 및 순위화 엔진, 콘텐츠 객체 분류기, 알림 제어기, 행위 로그, 제3자 콘텐츠 객체 노출 로그, 추론 모듈, 인증/프라이버시 서버, 검색 모듈, 광고 타겟팅 모듈, 사용자 인터페이스 모듈, 사용자 프로필 스토어, 연결 스토어, 제3자 콘텐츠 스토어 또는 위치 스토어를 포함할 수 있다. 또한, 소셜 네트워킹 시스템(160)은 다양한 서버, 서브-시스템, 프로그램, 모듈, 로그 및 데이터 스토어를 포함할 수 있다.

템(160)은 가령 네트워크 인터페이스, 보안 메커니즘, 부하 균형기, 장애 서버, 관리 및 네트워크 운영 콘솔, 다른 적절한 컴포넌트 또는 이들의 임의의 적절한 조합과 같이 적절한 컴포넌트들을 포함할 수 있다. 특정 실시예로, 소셜 네트워킹 시스템(160)은 사용자 프로필을 저장하기 위한 하나 이상의 사용자 프로필 스토어를 포함할 수 있다. 사용자 프로필은 예컨대 인명정보, 인구학적 정보, 행동 정보, 소셜 정보 또는 가령 경력, 학력, 취미나 기호, 관심사, 친밀감, 위치와 같은 다른 유형의 설명적 정보를 포함할 수 있다. 관심사 정보는 하나 이상의 카테고리에 관한 관심사를 포함할 수 있다. 카테고리는 일반적이거나 구체적인 수 있다. 예로서 제한 없이, 사용자가 신발 브랜드에 관한 기사에 "좋아한다"면, 카테고리는 그 브랜드일 수 있거나 "신발" 또는 "옷"의 일반 카테고리일 수 있다. 연결 스토어는 사용자에게 대한 연결 정보를 저장하는데 사용될 수 있다. 연결 정보는 유사하거나 공통의 경력, 그룹 멤버십, 취미, 학력을 가지거나 임의의 방식으로 관련되거나 공통 속성을 공유하는 사용자들을 표시할 수 있다. 또한, 연결 정보는 (내부와 외부 모두의) 다른 사용자들과 콘텐츠 사이의 사용자-정의된 연결들을 포함할 수 있다. 웹 서버는 네트워크(110)를 통해 하나 이상의 클라이언트 시스템(130)이나 하나 이상의 제3자 시스템(170)과 소셜 네트워킹 시스템(160)을 링크하는데 사용될 수 있다. 웹 서버는 소셜 네트워킹 시스템(160)과 하나 이상의 클라이언트 시스템(130) 사이에서 메시지를 수신하고 라우팅하기 위한 메일 서버나 다른 메시징 기능을 포함할 수 있다. API 요청 서버는 제3자 시스템(170)이 하나 이상의 API를 호출하여 소셜 네트워킹 시스템(160)으로부터의 정보에 접근할 수 있게 해줄 수 있다. 행위 로거는 소셜 네트워킹 시스템(160) 내부나 외부에서의 사용자의 행위에 대한 웹 서버로부터의 통신을 수신하는데 사용될 수 있다. 행위 로그와 함께, 제3자 콘텐츠 객체 로그에서는 제3자 콘텐츠 객체에 대한 사용자 노출이 관리될 수 있다. 알림 제어기는 콘텐츠 객체에 관한 정보를 클라이언트 시스템(130)에 제공할 수 있다. 정보는 알림으로서 클라이언트 시스템(130)으로 푸싱(pushed)될 수 있거나, 정보는 클라이언트 시스템(130)으로부터 수신된 요청에 응답하여 클라이언트 시스템(130)으로부터 풀링(pulled)될 수 있다. 인증 서버는 소셜 네트워킹 시스템(160)의 사용자의 하나 이상의 프라이버시 설정을 강제하는데 사용될 수 있다. 사용자의 프라이버시 설정은 사용자와 관련된 특정 정보가 어떻게 공유될 수 있는지를 결정한다. 인증 서버는 예컨대 적절한 프라이버시 설정을 설정함으로써, 사용자의 행위를 소셜 네트워킹 시스템(160)에 의해 로그되게 하거나 다른 시스템(예컨대, 제3자 시스템(170))과 공유되게 하도록 사용자가 참여하거나 탈퇴할 수 있게 해줄 수 있다. 제3자 콘텐츠 객체 스토어는 가령 제3자 시스템(170)과 같은 제3자로부터 수신된 콘텐츠 객체를 저장하는데 사용될 수 있다. 위치 스토어는 사용자와 관련된 클라이언트 시스템(130)으로부터 수신된 위치 정보를 저장하는데 사용될 수 있다. 광고 가격결정 모듈은 소셜 정보, 현재시간, 위치 정보 또는 다른 적절한 정보를 결합하여 알림의 형태로 사용자에게 관련 광고를 제공할 수 있다.

[0023] 소셜 그래프

[0024] 도 2는 예시적인 소셜 그래프(200)를 도시한다. 특정 실시예로, 소셜 네트워킹 시스템(160)은 하나 이상의 데이터 스토어에 하나 이상의 소셜 그래프(200)를 저장할 수 있다. 특정 실시예로, 소셜 그래프(200)는 다수의 사용자 노드(202)나 다수의 컨셉 노드(204)를 포함할 수 있는 다수의 노드 및 노드를 연결하는 다수의 에지(206)를 포함할 수 있다. 도 2에 도시된 예시적인 소셜 그래프(200)는 혼시적 목적상 2차원 시각적 지도 표현으로 도시된다. 특정 실시예로, 소셜 네트워킹 시스템(160), 클라이언트 시스템(130) 또는 제3자 시스템(170)은 소셜 그래프(200) 및 적절한 애플리케이션에 대한 관련 소셜-그래프 정보에 접근할 수 있다. 소셜 그래프(200)의 노드 및 에지는 예컨대 데이터 스토어(가령, 소셜-그래프 데이터베이스)에 데이터 객체로서 저장될 수 있다. 이런 데이터 스토어는 소셜 그래프(200)의 노드 또는 에지의 하나 이상의 검색가능하거나 질의가능한 인덱스를 포함할 수 있다.

[0025] 특정 실시예로, 사용자 노드(202)는 소셜 네트워킹 시스템(160)의 사용자에게 해당할 수 있다. 예로서 제한 없이, 사용자는 개인(사람인 사용자), 엔티티(예컨대, 기업, 사업체 또는 제3자 애플리케이션) 또는 소셜 네트워킹 시스템(160)과 상호작용하거나 소셜 네트워킹 시스템에서 통신하는 (예컨대, 개인 또는 엔티티의) 그룹일 수 있다. 특정 실시예로, 사용자가 소셜 네트워킹 시스템(160)에서 계정을 등록하면, 소셜 네트워킹 시스템(160)은 사용자에게 해당하는 사용자 노드(202)를 생성할 수 있고, 하나 이상의 데이터 스토어에 사용자 노드(202)를 저장할 수 있다. 적절한 경우, 본 명세서에 기술되는 사용자들 및 사용자 노드들(202)은 등록 사용자들 및 등록 사용자들과 관련된 사용자 노드들(202)을 말할 수 있다. 추가로 또는 대안으로, 본 명세서에 기술되는 사용자들 및 사용자 노드들(202)은 적절한 경우 소셜 네트워킹 시스템(160)에 등록되지 않은 사용자들을 말할 수 있다. 특정 실시예로, 사용자 노드(202)는 사용자가 제공한 정보 및 소셜 네트워킹 시스템(160)을 포함하는 다양한 시스템이 수집한 정보와 관련될 수 있다. 예로서 제한 없이, 사용자는 그들의 이름, 프로필 사진, 연락 정보, 생일, 성별, 혼인 여부, 가족 관계, 직장, 학력, 기호, 관심사 또는 다른 신상 정보를 제공할 수 있다. 특정 실시예로, 사용자 노드(202)는 사용자와 관련된 정보에 해당하는 하나 이상의 데이터 객체와 관련될 수 있다. 특정

실시예로, 사용자 노드(202)는 하나 이상의 웹페이지에 해당할 수 있다.

[0026] 특정 실시예로, 컨셉 노드(204)는 컨셉에 해당할 수 있다. 예로서 제한 없이, 컨셉은 장소(가령, 예컨대, 영화관, 레스토랑, 명소 또는 도시); 웹사이트(가령, 예컨대, 소셜 네트워크 시스템(160)과 관련된 웹사이트 또는 웹-애플리케이션 서버와 관련된 제3자 웹사이트); 엔티티(가령, 예컨대, 사람, 사업체, 그룹, 스포츠 팀 또는 유명인사); 소셜 네트워킹 시스템(160) 또는 가령 웹-애플리케이션 서버와 같은 외부 서버에 위치할 수 있는 자원(가령, 예컨대, 오디오 파일, 비디오 파일, 디지털 사진, 텍스트 파일, 구조화된 문서 또는 애플리케이션); 물적 재산권 또는 지적 재산권(가령, 예컨대, 조각품, 미술품, 영화, 게임, 노래, 아이디어, 사진 또는 저서); 게임; 활동; 아이디어나 이론; 또 다른 적절한 컨셉; 또는 2 이상의 이런 컨셉들에 해당할 수 있다. 컨셉 노드(204)는 사용자에게 의해 제공된 컨셉의 정보 또는 소셜 네트워킹 시스템(160)을 포함하는 다양한 시스템에 의해 수집된 정보와 관련될 수 있다. 예로서 제한 없이, 컨셉의 정보는 이름이나 제목; 하나 이상의 이미지(예컨대, 책의 커버 페이지의 이미지); 위치(예컨대, 주소 또는 지리적 위치); (URL과 관련될 수 있는) 웹사이트; 연락 정보(예컨대, 전화번호 또는 이메일 주소); 다른 적절한 컨셉 정보; 또는 이런 정보의 임의의 적절한 조합을 포함할 수 있다. 특정 실시예로, 컨셉 노드(204)는 컨셉 노드(204)와 관련된 정보에 해당하는 하나 이상의 데이터 객체와 관련될 수 있다. 특정 실시예로, 컨셉 노드(204)는 하나 이상의 웹페이지에 해당할 수 있다.

[0027] 특정 실시예로, 소셜 그래프(200)에서 노드는 ("프로필 페이지"라고 할 수 있는) 웹페이지를 표현하거나, 그 웹페이지로 표현될 수 있다. 프로필 페이지는 소셜 네트워킹 시스템(160)에 의해 호스팅될 수 있거나, 접근될 수 있다. 또한, 프로필 페이지는 제3자 서버(170)와 관련된 제3자 웹사이트에 호스팅될 수 있다. 예로서 제한 없이, 특정한 외부 웹페이지에 해당하는 프로필 페이지는 특정한 외부 웹페이지일 수 있고, 이런 프로필 페이지는 특정 컨셉 노드(204)에 해당할 수 있다. 프로필 페이지는 모든 또는 선택된 서브세트의 다른 사용자에게 의해 열람될 수 있다. 예로서 제한 없이, 사용자 노드(202)는 해당 사용자가 콘텐츠를 추가할 수 있고, 선언을 할 수 있으며, 그렇지 않으면 그들 자신을 표현할 수 있는 해당 사용자 프로필-페이지를 가질 수 있다. 또 다른 예로서 제한 없이, 컨셉 노드(204)는 특히 컨셉 노드(204)에 해당하는 컨셉과 관련하여 하나 이상의 사용자들이 콘텐츠를 추가할 수 있거나, 선언을 할 수 있거나, 그들 자신을 표현할 수 있는 해당 컨셉-프로필 페이지를 가질 수 있다.

[0028] 특정 실시예로, 컨셉 노드(204)는 제3자 시스템(170)에 의해 호스팅된 제3자 웹페이지 또는 자원을 표현할 수 있다. 제3자 웹페이지 또는 자원은 다른 요소들 중에서 행위 또는 활동을 표현하는 (예컨대, JavaScript, AJAX 또는 PHP 코드로 구현될 수 있는) 콘텐츠, 선택가능하거나 다른 아이콘 또는 다른 상호작용가능한 객체를 포함할 수 있다. 예로서 제한 없이, 제3자 웹페이지는 가령 "좋아요", "체크인", "식사하기(eat)", "추천하기" 또는 다른 적절한 행위나 활동과 같은 선택가능한 아이콘을 포함할 수 있다. 제3자 웹페이지를 열람하는 사용자는 아이콘들 중 하나(예컨대, "식사하기")를 선택하여 행위를 수행할 수 있고, 클라이언트 시스템(130)이 사용자의 행위를 표시하는 메시지를 소셜 네트워킹 시스템(160)으로 송신하게 할 수 있다. 그 메시지에 응답하여, 소셜 네트워킹 시스템(160)은 사용자에게 해당하는 사용자 노드(202)와 제3자 웹페이지 또는 자원에 해당하는 컨셉 노드(204) 사이의 에지(예컨대, "식사하기" 에지)를 생성할 수 있고, 하나 이상의 데이터 스토어에 에지(206)를 저장할 수 있다.

[0029] 특정 실시예로, 소셜 그래프(200)에서 노드 쌍은 하나 이상의 에지(206)에 의해 서로 연결될 수 있다. 노드 쌍을 연결하는 에지(206)는 노드 쌍 사이의 관계를 표현할 수 있다. 특정 실시예로, 에지(206)는 노드 쌍 사이의 관계에 해당하는 하나 이상의 데이터 객체나 속성을 포함하거나 표현할 수 있다. 예로서 제한 없이, 제1 사용자는 제2 사용자가 제1 사용자의 "친구"라고 표시할 수 있다. 이런 표시에 응답하여, 소셜 네트워킹 시스템(160)은 "친구 요청"을 제2 사용자에게 송신할 수 있다. 제2 사용자가 "친구 요청"을 확인하면, 소셜 네트워킹 시스템(160)은 소셜 그래프(200)에서 제1 사용자의 사용자 노드(202)와 제2 사용자의 사용자 노드(202)를 연결하는 에지(206)를 생성할 수 있고, 하나 이상의 데이터 스토어(164)에 소셜-그래프 정보로서 에지(206)를 저장할 수 있다. 도 2의 예에서, 소셜 그래프(200)는 사용자 "A"와 사용자 "B"의 사용자 노드(202)들 사이의 친구 관계를 표시하는 에지(206)를 그리고 사용자 "C"와 사용자 "B"의 사용자 노드(202) 사이의 친구 관계를 표시하는 에지를 포함한다. 본 명세서가 특정 사용자 노드(202)를 연결하는 특정 속성을 가진 특정 에지(206)를 기술하거나 도시하지만, 본 명세서는 사용자 노드(202)를 연결하는 임의의 적절한 속성을 가진 임의의 적절한 에지(206)를 고려한다. 예로서 제한 없이, 에지(206)는 친구관계, 가족관계, 사업이나 고용 관계, 팬 관계, 팔로워 관계, 방문자 관계, 구독자 관계, 상위/하위 관계, 호혜 관계, 비-상호 관계, 또 다른 적절한 타입의 관계 또는 2 이상의 이런 관계들을 표현할 수 있다. 게다가, 본 명세서는 일반적으로 노드들이 연결되는 것을 기술하지만, 본 명세서는 또한 사용자 또는 컨셉이 연결되는 것을 기술한다. 본 명세서에서, 사용자 또는 컨셉이 연결되는 것에

대한 언급은 적절한 경우 이들 사용자 또는 컨셉에 해당하는 노드가 하나 이상의 에지(206)에 의해 소셜 그래프(200)에서 연결되는 것을 말할 수 있다.

[0030] 특정 실시예로, 사용자 노드(202)와 컨셉 노드(204) 사이의 에지(206)는 컨셉 노드(204)와 관련된 컨셉에 대해 사용자 노드(202)의 사용자가 수행한 특정 행위 또는 활동을 표현할 수 있다. 예로서 제한 없이, 도 2에 도시된 바와 같이, 사용자는 컨셉을 "좋아요(like)", "참여했음(attended)", "실행했음(played)", "청취했음(listened)", "요리했음(cooked)", "근무했음(worked at)", 또는 "시청했음(watched)"을 할 수 있고, 이들 각각은 에지 타입이나 서브타입에 해당할 수 있다. 컨셉 노드(204)에 해당하는 컨셉-프로필 페이지는 예컨대 선택 가능한 "체크인" 아이콘(가령, 예컨대, 클릭가능한 "체크인" 아이콘) 또는 선택가능한 "즐거찾기에 추가(add to favorites)" 아이콘을 포함할 수 있다. 마찬가지로, 사용자가 이런 아이콘을 클릭한 후, 소셜 네트워킹 시스템(160)은 각각의 행위에 해당하는 사용자의 행위에 응답하여 "즐거찾기" 에지 또는 "체크인" 에지를 생성할 수 있다. 또 다른 예로서 제한 없이, 사용자(사용자 "C")는 특정 애플리케이션(온라인 음악 애플리케이션인 SPOTIFY)을 사용하여 특정 노래("Imagine")를 들을 수 있다. 이 경우, 소셜 네트워킹 시스템(160)은 사용자에게 해당하는 사용자 노드(202)와 노래 및 애플리케이션에 해당하는 컨셉 노드(204) 사이에 "청취했음(listened)" 에지(206) 및 "사용했음(used)" 에지(도 2에 도시)를 생성하여, 사용자가 그 노래를 들었고 그 애플리케이션을 사용했음을 표시할 수 있다. 게다가, 소셜 네트워킹 시스템(160)은 노래와 애플리케이션에 해당하는 컨셉 노드(204) 사이에서 "실행했음(played)" 에지(206)(도 2에 도시)를 생성하여 특정 노래가 특정 애플리케이션에 의해 실행되었음을 표시할 수 있다. 이 경우, "실행했음(played)" 에지(206)는 외부 오디오 파일(노래 "Imagine")에 대해 외부 애플리케이션(SPOTIFY)이 수행한 행위에 해당한다. 본 명세서는 사용자 노드(202)와 컨셉 노드(204)를 연결하는 특정 속성을 가진 에지(206)를 기술하지만, 본 명세서는 사용자 노드(202)와 컨셉 노드(204)를 연결하는 임의의 적절한 속성을 가진 임의의 적절한 에지(206)를 고려한다. 게다가, 본 명세서는 단일의 관계를 표현하는 사용자 노드(202)와 컨셉 노드(204) 사이의 에지를 기술하지만, 본 명세서는 하나 이상의 관계를 표현하는 사용자 노드(202)와 컨셉 노드(204) 사이의 에지를 고려한다. 예로서 제한 없이, 에지(206)는 사용자가 특정 컨셉에서 좋아요 하고 사용했음을 모두 표현할 수 있다. 대안으로, 또 다른 에지(206)는 사용자 노드(202)와 컨셉 노드(204) 사이(도 2에 도시된 바와 같이, 사용자 "E"에 대한 사용자 노드(202)와 "SPOTIFY"에 대한 컨셉 노드(204) 사이)의 각 타입의 관계(또는 다수의 단일 관계)를 표현할 수 있다.

[0031] 특정 실시예로, 소셜 네트워킹 시스템(160)은 소셜 그래프(200)에서 사용자 노드(202)와 컨셉 노드(204) 사이의 에지(206)를 생성할 수 있다. 예로서 제한 없이, (가령, 예컨대, 사용자의 클라이언트 시스템(130)에 의해 호스팅되는 웹 브라우저 또는 특수목적 애플리케이션을 사용하여) 사용자가 컨셉-프로필 페이지를 열람하는 것은 사용자가 "좋아요" 아이콘을 클릭하거나 선택하여 컨셉 노드(204)가 표현한 컨셉을 좋아함을 나타낼 수 있는데, "좋아요" 아이콘은 사용자의 클라이언트 시스템(130)이 컨셉-프로필 페이지와 관련된 컨셉을 사용자가 좋아했다고 표시한 메시지를 소셜 네트워킹 시스템(160)으로 송신하게 할 수 있다. 그 메시지에 응답하여, 소셜 네트워킹 시스템(160)은 사용자와 컨셉 노드(204) 사이의 "좋아요" 에지(206)로 도시된 바와 같이 사용자와 관련된 사용자 노드(202)와 컨셉 노드(204) 사이의 에지(206)를 생성할 수 있다. 특정 실시예로, 소셜 네트워킹 시스템(160)은 하나 이상의 데이터 스토어에 에지(206)를 저장할 수 있다. 특정 실시예로, 에지(206)는 특정 사용자 행위에 응답하여 소셜 네트워킹 시스템(160)에 의해 자동 형성될 수 있다. 예로서 제한 없이, 제1 사용자가 사진을 업로드하거나, 영화를 시청하거나, 노래를 듣는다면, 에지(206)는 제1 사용자에게 해당하는 사용자 노드(202)와 이런 컨셉에 해당하는 컨셉 노드(204) 사이에 형성될 수 있다. 본 명세서는 특정 방식으로 특정 에지(206)를 형성하는 것을 기술하지만, 본 명세서는 임의의 적절한 방식으로 임의의 적절한 에지(206)를 형성하는 것을 고려한다.

[0032] 네트워크-인지 제품 시판

[0033] 특정 실시예로, 소셜 네트워킹 시스템(160)은 새로운 제품, 피처(feature), 광고 또는 경험("처리")이 특정 사용자 클러스터("클러스터들(clusters)")로 그 개시를 스테거링하여 온라인 소셜 네트워크의 사용자에게 개시되도록 하는 방법을 향상시킬 수 있다. 표준 A/B 테스트 접근법을 사용하여 새로운 처리를 테스트하는 대신에, 테스트 및 시판은 네트워크-인지 방식으로 이루어질 수 있다. 예로서 제한 없이, 소셜 그래프(200)로부터의 별개의 노드들의 클러스터가 적절한 클러스터링 알고리즘을 사용하여 식별될 수 있다. 이런 클러스터는 특정 클러스터 내 노드들이 동일한 클러스터 내 다른 사용자에게 대해 네트워크 노출되도록 정의될 수 있다. 이를 통해 처리의 처리 효과는 더 효율적으로 결정될 수 있는데, 이는 특정 사용자 및 (특정 클러스터 내 노드들에 대응하는) 그 사용자의 대부분의 친구들이 처리를 수신할 것이라고 보장할 수 있기 때문이다. 이는 또 다른 사용자와 함께 사용될 수 있을 때(예컨대, 온라인 채팅 애플리케이션) 더 유용한 처리에 대해 특히 이점적일 수

있다. 클러스터가 식별된 후, 처리는 클러스터의 제1 세트의 사용자에게 제공되고 그들과 함께 테스트될 수 있다. 이후, 이런 클러스터의 제1 세트로부터 수신된 데이터는 이런 사용자들에 대한 처리의 처리 효과를 결정하도록 분석될 수 있다. 이후, 처리는 클러스터의 다음 세트에서 사용자들에게 시판되기 전에 수정될 수 있다. 처리를 시판하는 순차적인 접근법은 처리가 온라인 소셜 네트워크의 모든 사용자에게 제공될 때까지 시행될 수 있다. 비록 본 명세서는 특정 방식으로 처리를 사용자에게 제공하고 처리 효과를 결정하는 것을 제공하지만, 본 명세서는 임의의 적절한 방식으로 사용자에게 처리를 제공하고 처리 효과를 결정하는 것을 고려한다.

[0034] A/B 테스트법은 온라인 실험의 효과를 평가하는 표준 접근법이다. 통상적으로 그 목표는 전체 모집단의 샘플을 그것에 노출시켜 새로운 제품, 피처 또는 조건(condition)의 평균 처리 효과를 추정하는 것이다. A/B 테스트법의 결점은 개인들의 처리가 기저를 이루는 소셜 네트워크를 따라 이웃하는 개인들로 확산될 때 소셜 간섭(social interference)을 포함하는 실험에 대해 적합하지 못하다는 점이다. 이런 결점은 소셜 간섭 하의 평균 처리 효과를 분석하는 그래프 클러스터링을 사용하여 극복될 수 있다. 우선, 우리는 개인들이 실험에 "네트워크 노출"되어 있다고 간주될 수 있는 그래프-이론 조건(graph-theoretic conditions)을 특징짓는다. 이후, 우리는 어떻게 그래프 클러스터 랜덤화가 효율적이고 정확한 알고리즘을 허용하여 몇몇의 이런 노출 조건하에 네트워크 노출되는 각 노드에 대한 확률을 계산하는지를 보여준다. 이후, 이런 확률을 가령 Horvitz-Thompson 추정치(estimator)와 같은 역 가중치(inverse weights), 추정치로 사용하는 것은 노출 모델이 적절히 특정되었다고 제시되는 비편향(unbiased)된 효과 추정(effect estimate)을 제공할 수 있다.

[0035] 비편향된 추정치가 주어지면, 우리는 분산을 최소화하는데 중점을 둔다. 먼저, 우리는 추정치의 분산에 대한 간단한 충분조건이 그래프의 크기 n 으로 점근적으로 작도록 개발한다. 그러나, 일반적인 랜덤화 기법에서, 이런 분산은 그래프의 수차 지수 함수에 의해 하계(lower bounded)될 수 있다. 반면에, 우리는 그래프가 이웃의 성장율에 대한 제한된-성장 조건을 만족한다면, 추정치의 분산이 수차 선형 함수에 의해 상계(upper bounded)될 수 있는 노드 이웃들에 기반하는 자연 클러스터링 알고리즘(natural clustering algorithm)이 존재함을 보여준다. 따라서, 우리는 적절한 클러스터 랜덤화가 간섭하에 평균 처리 효과를 실험적으로 측정할 때 지수적으로 낮은 추정치 분산에 이를 수 있음을 보여준다.

[0036] 팩스와 휴대폰에서 온라인 소셜 네트워크까지의 소셜 제품 및 서비스는 그들의 가치에 대해 "네트워크 효과"를 사용자에게 고유하게 내보인다. 사용자에게 이런 상품의 가치는 고유하게 비-로컬(non-local)적인데, 이는 그 가치가 통상적으로 사용자의 네트워크의 회원들이 제품을 또한 사용할 때 증가하기 때문이다. 그러나, 랜덤화된 실험(가령 A/B 테스트)은 소위 "고정 단위 처리 값 추정(stable unit treatment value assumption)(SUTVA)"이라는, 각각의 개인의 응답이 그들 자신의 처리뿐 아니라 임의의 다른 개인의 처리에 의해 영향을 받는다고 결정적으로 가정한다. A/B 테스트법의 형식론과 네트워크 상호작용의 비-로컬적 효과 사이의 이런 갈등을 참작하는 것이 온라인 행동의 분석 및 네트워크 실험의 설계에서 핵심 열린 질문(key open question)으로 떠올랐다.

[0037] 고정 단위 처리 값 추정이 타당한 근사(reasonable approximation)인 보통의 랜덤화된 시행(예컨대, 검색 엔진이 사용자들의 방문 시간에 대한 그들의 색 기법(color scheme)의 효과를 A/B 테스트하는 경우) 하에서, 모집단은 2개의 그룹: (새로운 색 기법 A를 보는) "처리" 그룹 및 (디폴트 색 기법 B를 보는) 제어 그룹으로 나뉜다. 사용자 사이의 간섭 효과를 무시할 정도라고 가정하면, 처리된 그룹 내 각각의 개인들은 전체 모집단이 처리된다면 마치 그들이 그럴 것처럼 응답하며, 제어 그룹 내 각각의 개인들은 전체 모집단이 제어된다면 마치 그들이 그럴 것처럼 응답한다. 이런 방식으로, 우리는 2개의 구별된 "평행 우주(parallel universes)"(예컨대, 색 기법 A가 모든 사람에게 사용되는 "우주 A" 및 색 기법 B가 모든 사람에게 사용되는 "우주 B")의 샘플들로부터의 결과를 동시에 관찰하고 있다고 생각할 수 있고, 각각의 우주에서 사용자의 행동의 성질에 대해 추론할 수 있다.

[0038] 이런 다루기 쉬운 구조는 테스트되는 피처 또는 제품이 임의의 종류의 소셜 성분(social component)을 가지는 경우와 같이 한 명의 사용자 i 의 행동이 다른 사용자 j 의 행동에 적지 않은 효과를 가질 수 있을 때 급격히 변화한다. 이제, i 가 우주 A에 놓이고 j 가 우주 B에 놓인다면, A에서 i 의 행동에 대한 우리의 분석은 B에서 j 의 행동의 성질에 의해 악영향을 받으며, 그 역도 또한 동일하다; 우리는 더 이상 2개의 평행 우주를 가지고 있지 못하다.

[0039] 평균 처리 및 네트워크 노출(Average Treatment and Network Exposure)

[0040] 특정 실시예로, 소셜 네트워킹 시스템(160)은 제1 사용자 세트의 사용자들에 대한 처리의 처리 효과를 결정할 수 있다. 하나의 목표는 이런 상호작용이 존재할 때 모집단에 대한 처리의 평균 효과를 분석하기 위한 기술을

개발하는 것이다. 소셜 네트워킹 시스템(160)은 온라인 소셜 네트워크의 사용자들의 서버세트에 처리를 제공함으로써 그 처리를 테스트할 수 있다. 처리는 처리에 대한 i 의 반응이 온라인 소셜 네트워크의 이웃 j 도 또한 처리를 받는지 여부에 의존한다는 점에서 "소셜" 성분을 가진다. 또한, 이런 소셜 성분은 처리의 네트워크 효과라고도 일컬어질 수 있다. 예로서 제한 없이, 특정 처리는 가령 소셜 광고, 텍스트/비디오 채팅, 메시징, 소셜 네트워킹 등과 같은 소셜 성분들을 나타낼 수 있다. 예컨대, 텍스트/비디오 채팅 피처에 대해, 이런 피처는 사용자 i 가 채팅할 한 명 이상의 이웃 j 를 가지는 경우에만 유용하다. 개인/사용자는 그 개인이 테스트에 대한 처리를 제공받는다면 처리 그룹에 있는 것으로 간주되며, 그렇지 않으면 제어 그룹에 있는 것으로 간주된다. 관심에 대한 근원적인 수치 응답 변수(예컨대, 각 조건에서 사용자의 time-on-site)가 있을 수 있고, 모집단이 처리군과 제어군으로 나뉘지기 때문에 어떤 가설상의 우주도 실제로 테스트되고 있지 않더라도 소셜 네트워킹 시스템(160)은 모든 사람이 서비스를 받는 우주 및 아무도 서비스를 받지 않는 우주 모두에서 이런 응답의 평균을 추

정할 수 있다. $\vec{z} \in \{0,1\}^n$ 를 처리 할당 벡터라고 하자. 여기서, $z_i = 1$ 은 사용자 i 가 처리 그룹에 있음을 의미하고, $z_i = 0$ 은 사용자 i 가 제어 그룹에 있음을 의미한다. $Y_i(\vec{z}') \in R$ 은 처리 할당 벡터 \vec{z}' 하에서 사용자 i 의 잠재적 결과(potential outcome)라고 하자. 우리가 관심을 가지는 기본량은 2개의 정반대의 우주들 $\vec{z} = \vec{1}$ 및 $\vec{z}' = \vec{0}$ 사이의 평균 처리 효과 τ 이다

(수식 1)

$$\tau(\vec{z} = \vec{1}, \vec{z}' = \vec{0}) = \frac{1}{n} \sum_{i=1}^n [Y_i(\vec{z} = \vec{1}) - Y_i(\vec{z}' = \vec{0})]$$

이 공식은 상기 약식의 항들로 논의되는 핵심 문제를 포함한다: 일반 A/B 테스트법과 달리, 2명의 사용자들 누구도 실제로 동시에 정반대의 우주에 있는 적은 없다. 특정 실시예로, 소셜 네트워킹 시스템(160)은 사용자들에 대한 처리의 개인 효과와 사용자들에 대한 처리의 네트워크 효과 모두를 결정할 수 있다. 특정 처리의 처리 효과는 네트워크 효과와 개인 효과 모두의 함수일 수 있다. 처리의 비-소셜 성분은 처리의 개인 효과라고 일컬어질 수 있다. 처리는 그 처리에 대한 i 의 반응이 온라인 소셜 네트워크의 이웃 j 도 또한 그 처리를 받는지 여부에 의존하지 않는다는 점에서 개인 성분을 가진다. 예로서 제한 없이, 사용자 i 에 대한 소셜 광고는 (물론, 소셜 광고도 또한 네트워크 효과를 가질 수 있다고 하더라도) 한 명 이상의 이웃 j 도 또한 소셜 광고에 노출되는지에 관계 없이 사용자 i 에 대한 개인 효과를 가질 수 있다. 특정 처리의 개인 효과와 네트워크 효과 모두를 구별하고 측정하는 것은 그 처리에 대한 사용자 참여를 증가시키거나 향상시키기 위해 처리를 최상으로 수정하는 방법을 결정하는데 유용할 수 있다. 비록 본 명세서는 특정 방식으로 처리 효과를 결정하는 것을 기술하지만, 본 명세서는 임의의 적절한 방식으로 처리 효과를 결정하는 것을 고려한다.

그래프 클러스터 랜덤화(Graph Cluster Randomization)

도 3은 소셜 그래프(200)의 클러스터들의 예를 도시한다. 특정 실시예로, 소셜 네트워킹 시스템은 그래프 클러스터 랜덤화를 사용하여 소셜 그래프(200)에서 복수의 클러스터를 식별할 수 있다. 각각의 식별된 클러스터는 소셜 그래프(200)의 복수의 노드 중 별개의 노드 세트를 포함할 수 있다. 도 3은 식별된 4개의 클러스터를 가진 상대적으로 소규모의 소셜 그래프(200)의 간단한 예를 도시한다. 각 클러스터는 이산되어 있다(즉, 중복되지 않는다). 특정 실시예로, 소셜 네트워킹 시스템(160)은 각 클러스터 내 임계 수의 노드가 클러스터 내 다른 노드들에 대해 네트워크 노출될 수 있도록 클러스터들을 특정할 수 있다. (수식 1)을 평가하기 위한 핵심 개념은 네트워크 노출의 개념이다. \vec{z}' 하의 i 의 응답이 그 할당에서 i 의 응답과 동일하다면 사용자 i 는 특정 할당 \vec{z}' 하의 처리에 "네트워크 노출"되며, 여기서 모든 사람은 처리를 수신한다. (또한, 우리는 이런 2가지 경우의 응답들이 작은 파라미터 ϵ 에 의해서만 차이가 있는 경우의 적용을 논의한다.) 우리는 제어 조건에 대한 네트워크 노출을 유사하게 정의한다. 이런 정의를 적소에 적용하여, 우리는 네트워크 노출을 구성하는 몇몇의 가능한 조건을 조사할 수 있다. 예로서 제한 없이, 사용자 i 는 i 및 i 의 이웃 모두가 처리된다면 그 처리에 네트워크 노출될 수 있다. 또 다른 예로서 제한 없이, 사용자 i 는, 고정 분수 $q > 0$ 에 대해 i 및 i 의 이웃의 적어도 q 분수가 처리된다면, 그 처리에 네트워크 노출될 수 있다. 네트워크 노출의 정의는 기본적으로 실험자에 의한 모델링 결정이며, 이런 작업에서 우리는 몇몇 집합의 노출 조건을 도입하는데, 각각은 사용자가 처리 우주 및 제어 우주

에 네트워크 노출된다고 가정될 수 있는 할당 벡터의 세트를 특정하고, 2개의 우주 사이의 연속체(continuum)의 몇몇 특성화(characterizations)를 제공한다. 네트워크 노출 조건을 선택하는 것은 사용자들을 실제로 처리 우주 또는 제어 우주에 놓지 않고 그들이 처리 우주 또는 제어 우주에 있는 것처럼 소셜 네트워킹 시스템(160)이 사용자의 잠재적 결과를 관찰할 수 있을 때를 특정하는데 중요할 수 있다. 특정 실시예로, 소셜 네트워킹 시스템(160)은 서로에 대한 사용자들의 소셜-그래프 친밀성에 적어도 부분적으로 기반하여 클러스터들을 식별할 수 있다. 소셜 그래프(200)에 이웃이 있음(즉, 에지(206)로 연결됨)에 기반하여 단지 사용자들을 클러스터링하는 대신에, 소셜 네트워킹 시스템(160)은 서로에 대해 상대적으로 높은 소셜-그래프 친밀성을 가진 이웃들을 클러스터링할 수 있다. 즉, 사용자 i 는 사용자 i 에 대해 임계치 이상의 친밀성을 가진 이웃들 j 를 가지는 클러스터들일 수 있는 반면, 사용자 i 에 대해 친밀성이 적은 이웃들은 그 클러스터에 반드시 포함되지는 않을 수 있다. 특정 실시예로, 소셜 네트워킹 시스템(160)은 클러스터 내 노드들의 네트워크 노출이 네트워크 노출의 특정된 분포 범위를 가지도록 클러스터들을 식별할 수 있다. 예로서 제한 없이, 네트워크 노출은 임계 레벨의 네트워크 노출(예컨대, 클러스터 내 모든 노드가 5%의 이웃 노출 이상임)에서 최대 레벨의 네트워크 노출(예컨대, 클러스터 내 모든 노드가 95% 이웃 노출 이하임)까지로 특정될 수 있다. 이후, 소셜 네트워킹 시스템(160)은 네트워크 노출의 분포 범위 상에서 클러스터 내 사용자들에 대한 처리의 처리 효과를 결정할 수 있다. 이런 처리 효과는 이후 처리의 개인 효과와 네트워크 효과를 결정하도록 그들의 제한들로 추정될 수 있다. 비록 본 명세서는 특정 방식으로 네트워크 노출을 가진 클러스터들을 식별하는 것을 기술하지만, 본 명세서는 임의의 적절한 방식으로 네트워크 노출을 가진 클러스터들을 식별하는 것을 고려한다.

[0046] 네트워크 노출의 공식에 이어서, 우리가 도입하는 제2의 핵심 개념은 그래프 클러스터 랜덤화라고 하는 그래프 클러스터링에 기반한 일반 그래프 랜덤화 기법(generic graph randomization scheme)이다. 그래프 클러스터 랜덤화는 그래프가 한 세트의 클러스터로 분할되는 기법이며, 처리와 제어 사이의 랜덤화가 클러스터 레벨에서 수행된다. 이후, 한 노드가 처리 또는 제어에 네트워크 노출되는 확률은 그 노드 근처의 로컬 그래프 구조와 클러스터 세트의 교집합에 대한 그래프-이론 질문(graph-theoretic question)을 통상적으로 포함할 것이다. 소셜 네트워킹 시스템(160)은 이런 랜덤화 하에서 네트워크 노출 조건에 진입하는 비-균등 확률(non-uniform probabilities)을 정확히 결정할 수 있다. 이후, 역 확률 가중화([참고문헌 9])를 사용하여, 소셜 네트워킹 시스템(160)은 우리가 명시적으로 확률을 계산할 수 있는 임의의 네트워크 노출 하에서 평균 처리 효과 τ 의 비편향된 추정치를 도출할 수 있다.

[0047] 우리는 그래프 클러스터 랜덤화가 점근적으로 작은 분산을 가진 추정치를 생성하는 조건을 제공함으로써 그래프 클러스터 랜덤화의 영향력에 동기를 부여한다. 먼저, 우리는 그래프가 한계 차수(bounded degree)를 가지고 모든 클러스터의 크기가 노드 수 n 과 무관하게 여전히 제한된다면 추정치 분산이 소규모에 대해 간단하지만 구체적인 충분 조건인 $O(1/n)$ 임을 관찰한다. 핵심 과제는 차수에 대한 종속성이다. 일반적으로, 제한된 크기(bounded-size)의 클러스터의 집합은 노드 차수로 지수적으로 증가하는 분산을 생성할 수 있다. 더 정확히, 단일-노드 클러스터로 그래프 클러스터 랜덤화를 수행하는 경우, 추정치의 분산은 차수에 지수적으로 종속하는 하계(lower bound)를 허용한다. 이는 클러스터링을 선택하는 방법에 대한 중요한 알고리즘 문제를 제기한다: 제한된 크기의 클러스터는 노드 수 n 에서 점근적으로 작은 분산을 제공하지만, 클러스터가 주의 깊게 선택되지 않으면 실제로 분산이 매우 커질 수 있는 노드 차수에 대한 지수적 종속성을 얻는다.

[0048] 제한된-증가 그래프에서 클러스터 랜덤화

[0049] 우리는, 비단순 클러스터링 알고리즘이 그래프 차수에 선형적인 추정치 분산에 대한 상계(upper bound)를 허용하는, 제한된-증가 그래프(restricted-growth graphs)라고 하는 중요한 클래스의 그래프들을 식별한다. 그래프에 대해 도입하는 제한된-증가 조건은 이웃이 지수적으로 증가할 수 있는 저-직경(low-diameter) 그래프를 포함하도록 설계되는, 거리 공간 내 최근접-이웃 알고리즘(nearest-neighbor algorithms)([참고문헌 10])을 연구하도록 이전에 도입된 한계-증가 조건(bounded-growth condition)의 확장이다. $B_r(v)$ 는 노드 v 의 τ 홉수(hops) 내에 있는 노드들의 세트라 한다; 제한된-증가 그래프는 그래프의 차수와 무관하게 상수 κ 가 존재함을 말하며, 모든 노드 v 및 모든 $\tau > 0$ 에 대하여 $|B_{r+1}(v)| \leq \kappa |B_r(v)|$ 를 가진다. 훨씬 강한 조건을 가지며 우리의 결과를 유지하는데 반드시 필요한 것은 아닌 $|B_{2r}(v)| \leq \kappa |B_r(v)|$ 을 필요로 하는 표준 한

계-증가 정의와의 비교를 주목하자.

[0050] 제한된-증가 그래프에 대해, 추정치 분산이 차수에 대해 단지 선형적으로 증가할 수 있는 클러스터링 알고리즘을 제공한다. 과제는 분산이 노드의 이웃에 교차하는 클러스터의 수에 지수적으로 증가한다는 점이다; 우리의 접근법은 잘-분리된 노드의 세트 주위에 그려진 고정된 반경의 볼(balls)로부터 클러스터들을 형성하는 것이다. 제한된 증가 조건은 볼이 임의의 하나의 노드 주위에 너무 근접하게 채워지는 것을 방지함으로써 노드 이웃이 너무 많은 클러스터와 만나는 것을 방지한다. 유클리드 공간 내 균밀도 임베딩(embedding)과 만나는 제한된-증가 그래프의 특수한 경우에 대하여 임베딩 내 노드의 위치를 사용하여 그 공간을 직접 클러스터들로 분할할 수 있다; 최근접-이웃 과제의 작업([참고문헌 10])으로서 요점은 기하학적 레벨보다는 그래프-이론적 레벨에서 이런 분할(carving-up)을 제어하는 것이며, 이것이 우리의 기법이 수행하는 것이다.

[0051] 제한된-증가 그래프의 우리의 클래스는 임의의 타입의 실세계 그래프에 대한 매력적인 모델을 제공한다. 제한된-증가 그래프는 가령 격자들 또는 랜덤 기하학 그래프와 같은 한계 차수(bounded dimension)의 유클리드 공간에서 대략적으로 균일한 밀도를 가진 노드들의 임베딩이 존재하는 그래프를 포함한다.

[0052] 네트워크 노출 모델

[0053] A/B 랜덤형 실험에 대해, 한 개인의 처리 조건은 그들이 개입(intervention)에 적용을 받는지 여부를 결정한다. 이것은 통상 2개의 값을 취한다: "처리" 또는 "제어". 대부분의 랜덤형 실험에서, 실험자는 처리 조건을 랜덤화하는 방법에 대해 명시적인 제어를 가지며, 일반적으로 개인들은 독립적으로 할당된다. 한편, 한 개인의 노출 조건은 세계가 어떻게 개입을 경험하는지에 충분히 연계하여 그들이 어떻게 개입을 경험하는지를 결정한다. 고정 단위 처리 값 추정 없이, 최악의 경우, 각각의 \vec{z} 의 2^n 의 가능한 값들은 각 사용자에게 대한 별개의 노출 조건을 정의한다. Aronow 및 Samii는 이것을 "임의 노출(arbitrary exposure)"이라고 하며([참고문헌 2]), 임의 노출 하에 실험을 분석하는 다루기 쉬운 방법을 없을 것이다.

[0054] 사용자 i 에 대한 잠재적 결과를 고려하자. "임의 노출"의 경우, $Y_i(\vec{z})$ 는 모든 가능한 \vec{z} 에 대해 완전히 다르다. 이는 우리가 모든 사용자를 처리 우주 또는 제어 우주에 놓지 않고서는 $\vec{z}=\vec{1}$ 또는 $\vec{z}=\vec{0}$ 중 하나에 대해 $Y_i(\vec{z}')$ 를 관찰할 수 없을 것임을 의미한다. 따라서, 임의의 다른 조건 하에서 평균 처리 효과를 추정하는 것을 진보시키기 위해, 우리는 추가적인 가정을 필요로 한다. 우리는 여기서 다수의 처리 벡터 \vec{z} 가 동일한 잠재적 결과에 맵핑할 수 있다고 가정한다: 본래, 하기에 정확히 나타나는 의미로, 처리 벡터 \vec{z} 및 \vec{z}' 가 노드 i 의 관점에서 "충분히 유사"하는 한, i 는 \vec{z} 및 \vec{z}' 하에서 동일한 응답을 가질 것이다.

[0055] 상세하게, σ_i^x 는 i 가 결과 \mathcal{X} 를 경험하는 모든 할당 벡터 \vec{z} 의 세트이다. σ_i^x 는 i 에 대한 노출 조건이라고 한다; 본래, σ_i^x 는 i 의 관점에서 "구별되지 않는(indistinguishable)" 할당 벡터의 세트로 구성된다. 우리의 관심사는 특히 노출 조건 σ_i^1 및 σ_i^0 이며, 각각은 $\vec{z}=\vec{1}$ 및 $\vec{z}=\vec{0}$ 를 포함하는 세트라고 정의한다. 이런 방식으로, 모든 $\vec{z}_1 \in \sigma_i^1$ 에 대하여 $Y_i(\vec{z}=\vec{z}_1)=Y_i(\vec{z}=\vec{1})$ 이며, 모든 $\vec{z}_0 \in \sigma_i^0$ 에 대하여 $Y_i(\vec{z}=\vec{z}_0)=Y_i(\vec{z}=\vec{0})$ 이라고 가정한다. $\vec{z}=\vec{1}$ 및 $\vec{z}=\vec{0}$ 은 동일한 한 노출 조건에 속할 수 있고, $\sigma_i^1=\sigma_i^0$ 는 아무런 효과를 가지지 않는 처리에 해당함을 주목하라. ("노출 조건"의 정의로 너무 제한적이라고 생각되면, 대신 \in -빈(bin)에서 잠재적 결과를 생성하는 모든 할당 벡터로서 "노출 조건"을 정의하는 작은 \in -크기의 빈(bins)을 사용하여 (할당 벡터의 공간을 분할하는 대신) 잠재적

결과의 공간을 분할하는 것을 고려하라. $\bar{z}=\bar{1}$ 또는 $\bar{z}=\bar{0}$ 에 대한 결과에 대응하는 다른 잠재적 결과가 없는 경우, 이 방식으로 잠재적 결과에 대한 \in 거리를 사용하여 편향(bias)을 관리하는 것이 더 적절할 수 있다.)

[0056] 사용자 i에 대한 노출 모델은 가능한 할당 벡터 \bar{z} 를 완전히 분할하는 한 세트의 노출 조건으로 정의된다. 모든 사용자에게 대해 모든 모델의 세트는 실험에 대한 노출 모델이다. 목적상, 극단의 우주들 사이의 평균 처리 효과를 결정하도록 단지 노력중이기 때문에 노출 모델을 전부 특정하는 것이 불필요하다. 주로 각 사용자 i가 처리 우주 또는 제어 우주에 대한 노출을 경험하는 노출 상태 σ_i^1 및 σ_i^0 에 주목한다. (각기 다른 노출 조건에서의 잠재적 결과들 사이의 함수 관계를 가정한다면, σ_i^1 및 σ_i^0 이외의 다른 노출 조건이 더 연관될 수 있다.)

[0057] 물론, 각 사용자에게 대한 참인 노출 조건 σ_i^1 및 σ_i^0 이 사전에 실험자에게 알려지지 않으며, 실험의 결과를 분석하려면 우리의 체계 안에서 이런 조건들을 선택해야 한다. 불충분한 노출 조건이 실험자에 의해 선택되면, 사용자는 $\bar{z}=\bar{1}$ 및 $\bar{z}=\bar{0}$ 에 대응하지 않는다는 방식으로 응답할 수 있고, 편향이 평균 처리 효과의 계산으로 도입될 수 있다. 이런 편향의 크기는 실제로 관찰된 결과가 우리가 관찰하고자 했던 $\bar{z}=\bar{1}$ 및 $\bar{z}=\bar{0}$ 에서의 결과와 얼마나 근접한지에 의존한다. 심지어, 실험의 결과에 대한 분산을 낮추기 위해 이런 편향을 가능하게 하는 것이 바람직할 수 있다.

[0058] 이웃 노출(Neighborhood Exposure)

[0059] 특정 실시예로, 소셜 네트워킹 시스템(160)은 다양한 노출 조건에 노출되도록 클러스터들을 식별할 수 있다. 임의의 적절한 노출 조건이 사용될 수 있다. 특히, 우리는 주로 2개의 할당이 i의 직접적인 그래프 이웃에 있어서 일치한다면 이들이 i에 대해 구별되지 않는 로컬 노출 조건(local exposure conditions)에 중점을 둔다. 처리된 이웃의 수에 대한 절대 조건 및 부분 조건을 고려한다. 이런 가능한 노출 조건들이 실험에서 실제의 잠재적 결과에 대한 실제의 노출 조건들이라는 점을 우리가 반드시 주장하는 것은 아니지만, 이들이 실험의 분석에 대한 유용한 추출(abstractions)을 제공하며 도입되는 편향의 차수는 다시 반사실적(counterfactual) 우주에 속하는 노출 조건들이 얼마나 잘 근사화되는지에 의존함에 유의하자. 예로서 제한 없이, 소셜 네트워킹 시스템(160)은 각 클러스터에 대하여 클러스터 내 노드의 네트워크 노출이 완전 이웃 노출(full neighborhood exposure)이도록 클러스터들을 정의할 수 있다. 완전 이웃 노출에서, 노드 i는 i 및 모든 i의 이웃이 어느 한 처리 조건을 수신하면 그 처리 조건에 대한 완전 이웃 노출을 경험한다. 또 다른 예로서 제한 없이, 소셜 네트워킹 시스템(160)은 각 클러스터에 대해 클러스터 내 노드의 네트워크 경험이 절대 k-이웃 노출(absolute k-neighborhood exposure)이도록 클러스터들을 정의할 수 있다. 절대 k-이웃 노출에서, $d \geq k$ 인 차수 d의 노드 i는 i 및 i의 $\geq k$ 이웃이 처리 조건을 수신하면 그 처리 조건에 대한 절대 k-이웃 노출을 경험한다. 또 다른 예로서 제한 없이, 소셜 네트워킹 시스템(160)은 각 클러스터에 대해 클러스터 내 노드의 네트워크 노출이 부분 q-이웃 노출(fractional q-neighborhood exposure)이도록 클러스터들을 정의할 수 있다. 부분 q-이웃 노출에서 차수 d의 노드 i는 i 및 i의 $\geq qd$ 이웃이 처리 조건을 수신하면 그 처리 조건에 대한 부분 q-이웃 노출을 경험한다. k-절대 이웃 노출 및 q-부분 이웃 노출은 i의 더 적은 이웃이 그 노출 조건에 속하는 것으로 간주되도록 i에 대한 고정된 처리 조건을 가지기를 요구한다는 점에서 노드 i에 대한 완전 이웃 노출의 완화(relaxations)로 간주될 수 있다. 실제로, k-절대 이웃 노출 및 q-부분 이웃 노출에 대응하는 할당 벡터의 세트는 파라미터 k 및 q 하에 각각 내포(nested)된다. k 또는 q가 증가하면 노드 i에 대한 완전 이웃 노출에 도달할 때까지 할당 벡터의 세트는 감소한다. 비록 본 명세서는 특정 노출 조건을 기술하지만, 본 명세서는 임의의 적절한 노출 조건을 고려한다.

[0060] 각 사용자에게 대해 다른 값들인 이중 값 k 또는 q를 고려하는 것이 당연하지만, 가능한 한 많이 사용자에게 대해 동종인 노출 조건으로 우리의 논의를 제한한다. 우리는 노드가 $d > k$ 차수를 가질 때 k-이웃 노출의 정의에 대해 가벼운 이중성을 통합할 수 있다; 대신 이런 노드들에 대해 우리는 완전 이웃 노출을 고려한다. 부분 노출은 이

런 조정을 필요로 하지 않는다.

[0061]

코어 노출(Core Exposure)

[0062]

충분 이웃 노출은 단지 우주에 대한 완전 몰입(full immersion)의 명백한 근사화일 뿐이다. 로컬 노출 조건을 넘어서, 우리는 또한 전면적 종속성(global dependence)을 가진 노출 조건을 고려한다. 하나의 접근법으로서, 개인들이 충분히 많은 처리된 이웃들에 의해 충분히 둘러싸여 있고 그 이웃들도 차례로 또한 충분히 많은 처리된 이웃들에 의해 둘러싸여 있는 동일 경우에만 처리에 노출되는 것으로 간주하자. 이런 회귀적 정의(recursive definition)는 처음에는 처리하기 어려워 보일 수 있지만, 이런 회귀적 노출은 실제로 처리 및 제어 개인들의 유도된 그래프에서 k-코어 그리고 더 일반적으로는 이종 k-코어(heterogeneous k-코어)를 분석하여 정확히 특성

화될 수 있다. 그래프 $G = (V, E)$ 의 k-코어는 모든 노드가 적어도 k 차수를 가지는 G의 최대 서브그래프이다([참고문헌 4]). 마찬가지로, 벡터 $\vec{k} = (k_1, \dots, k_{|V|})$ 파라미터로 나타내는 그래프 $G = (V, E)$ 의

이종 k-코어는 각 노드 $v_i \in V'$ 가 적어도 k_i 차수를 가지는 G의 최대 서브그래프 $H = (V', E')$ 이다

([참고문헌 5]). 마찬가지로, 부분 q-코어는 각 노드 $v_i \in V'$ 가 G에서 그 노드에 연결되었던 노드들의 적어도 분수 q에 연결되는 $G = (V, E)$ 의 최대 서브그래프 $H = (V', E')$ 이다. 따라서, 모든 $v_i \in V'$

에 대해, $\deg_H(v_i) \geq q \deg_G(v_i)$ 이다. 동등하게, d_i 가 노드 i의 차수라면, 부분 q-코어는 $\vec{k} = (qd_1, \dots, qd_{|V|})$ 에 대해 G의 이종 k-코어이다.

[0063]

이종 k-코어는 잘 정의된 객체이기 때문에, 부분 q-코어도 또한 그러하다. 이 정의를 사용하여, 우리는 이제 해당하는 이전의 이웃 조건의 모든 더 엄격한 버전인 노출 조건들을 정의한다. 예로서 제한 없이, 소셜 네트워킹 시스템(160)은 각 클러스터에 대해 클러스터에서 노드의 네트워크 노출이 컴포넌트 노출(component exposure)이도록 클러스터들을 정의할 수 있다. 컴포넌트 노출에서 노드 i는 i 및 그에 연결된 컴포넌트 내 모든 노드가 처리 조건을 수신하면 그 처리 조건에 대한 컴포넌트 노출을 경험한다. 또 다른 예로서 제한 없이, 소셜 네트워킹 시스템(160)은 각 클러스터에 대해 클러스터 내 노드의 네트워크 노출이 절대 k-코어 노출이도록 클러스터들을 정의할 수 있다. 절대 k-코어 노출에서, $d \geq k$ 인 차수를 가진 노드 i는 i가 처리 조건을 수신하는 노드 V'에서

유도된 G의 서브그래프인 $G[V']$ 그래프의 k-코어에 속한다면 그 처리 조건에 대한 절대 k-코어 노출을 경험한다. 또 다른 예로서 제한 없이, 소셜 네트워킹 시스템(160)은 각 클러스터에 대해 클러스터 내 노드의 네트워크 노출이 부분 q-코어 노출이도록 클러스터들을 정의할 수 있다. 부분 q-코어 노출에서 노드 i는 i가 처리 조건을

수신하는 노드 V'에서 유도된 G의 서브그래프인 $G[V']$ 그래프의 부분 q-코어에 속한다면 그 처리 조건에 대한 부분 q-코어 노출을 경험한다. 컴포넌트 노출은 네트워크 노출에 대해 생각할 수 있는 가장 강한 요건이며, 연구되는 간접 그래프가 많은 비연결 컴포넌트를 포함하는 경우에만 실현가능하다. 본 명세서에서 q=1에 대한 부분 q-코어 노출이 컴포넌트 노출로 감소함을 유의해야 함을 명확히 포함한다. 다시 이웃 노출의 경우와 같이, 절대 코어 노출은 모든 사용자에게 대해 유용한 조건이도록 사용자들에게 k의 이중성을 요구한다. k-이웃 노출에 대한 솔루션과 유사한 간명한 솔루션이 이중의 $\max(\text{degree}, k)$ -코어 노출을 고려하는 것일 수 있다. 한편으로, 부분 q-이웃 노출과 같은 부분 q-코어 노출은 이런 간명한 솔루션에서 자유롭다.

[0064]

코어 노출 조건은 상술한 관련 이웃 노출 조건보다 더 엄격하다. 실제로, 노드 i가 컴포넌트 또는 코어 노출되는 모든 할당 벡터는 이웃 노출에 대응하지만, 그 역은 그렇지 않다. 그래서, 코어와 컴포넌트 노출의 할당 벡터는 관련 이웃 노출의 할당 벡터에 완전히 포함된다.

[0065]

다른 노출 조건들

[0066]

다른 노출 조건이 특정 적용과 관련될 수 있다. 특히, h 홉수 내 노드의 모집단에 절대 또는 부분 조건을 설정하는 매개 개념(intermediate concept)이 관심사로서, 여기서 h=1은 상술한 이웃 노출 조건이다. 예로서 제한 없이, 소셜 네트워킹 시스템(160)은 i의 임계 이격도 내 노드들의 임계 분수가 처리에 노출되게 하는 각 클러스

터에 대해 클러스터 내 노드들의 네트워크 노출이도록 클러스터들을 정의할 수 있다. 또한, 매우 큰 차수를 가진 소셜 네트워크에 대해 많은 적용에서 단지 더 강한 유대를 고려하는 더 낮은 차수 네트워크의 관점에서 노출 조건을 정의하는 것이 더 관련성이 있을 수 있음에 유의하자.

[0067] 랜덤화 및 추정(Randomization and Estimation)

[0068] 특정 실시예로, 소셜 네트워킹 시스템(160)은 제1 클러스터 세트에 대응하는 제1 사용자 세트에 처리를 제공할 수 있다. 처리는 예컨대 특정 광고, 제품, 피쳐, 경험, 다른 적절한 처리 또는 이들의 임의의 조합일 수 있다. 처리는 온라인 소셜 네트워크 또는 제3자 시스템(170)을 통해 제공될 수 있다. 이후, 소셜 네트워킹 시스템(160)은 제1 사용자 세트의 사용자들에 대한 처리의 처리 효과를 결정할 수 있다. 네트워크 노출의 개념을 사용하여, 이제 랜덤형 실험을 사용하여 2개의 반사실적 우주들 사이의 평균 처리 효과 τ 를 추정하는 것을 고려할 수 있다. \vec{Z} 가 실험의 처리 할당 벡터라는 점을 상기하자. 실험을 랜덤화하기 위해, \vec{Z} 은 \vec{Z} 의 범위인 $\{0, 1\}^n$ 에서 값들을 취하는 랜덤 벡터 Z로부터 도출되게 한다. $\Pr(Z = \vec{Z})$ 에 의해 주어지는 $\{0, 1\}^n$ 에서의 Z의 분포는 우리의 랜덤화 기법을 정의하는 것이며, 또한 네트워크 노출의 관련 확률을 정확히 결정하는 것이다. 사용자 i에 대해, $\Pr(Z \in \sigma_i^1)$ 는 처리에 대한 네트워크 노출의 확률이며, $\Pr(Z \in \sigma_i^0)$ 은 제어에 대한 네트워크 노출의 확률이다. 통상, 이런 확률들은 각 사용자 및 각 처리 조건에 대해 다를 것이며, 이런 확률들을 알면 랜덤화 동안 편향의 할당을 정정할 수 있다. 특히, 하기에 주어진 바와 같이, τ 의 비편향된 추정치를 얻기 위해 Horvitz-Thompson 추정치 $\hat{\tau}$ 를 사용하는 것이 가능해진다:

[0069] (수식 2)

$$\hat{\tau}(Z) = \frac{1}{n} \sum_{i=1}^n \left(\frac{Y_i(Z)1[Z \in \sigma_i^1]}{\Pr(Z \in \sigma_i^1)} - \frac{Y_i(Z)1[Z \in \sigma_i^0]}{\Pr(Z \in \sigma_i^0)} \right)$$

[0070]

[0071] 여기서, $1[x]$ 는 지시 함수(indicator function)이다. 확률이 양수라고 가정하면, 노출 조건이 잘못 지정되지 않음을 가정함을 유의하더라도 Z에 대한 기대값(expectation)은 τ 를 제공한다. 특정 실시예로, 소셜 네트워킹 시스템(160)은 제2 클러스터 세트에 대응하는 제2 사용자 세트에 처리를 제공할 수 있다. 본 명세서에서 기술한 프로세스는 이후 필요한 경우 처리가 온라인 소셜 네트워크의 모든 사용자에게 제공될 때까지 반복될 수 있다. 이런 방식으로, 광고, 피쳐 또는 제품은 실제로 처리가 특정 사용자 세트에게만 제공되었을 때 마치 전체 소셜 네트워크에 효과적으로 제공된 것처럼 처리 그룹 내 사용자가 느끼도록 처리를 수신하는 사용자들의 각 클러스터가 처리를 경험하는 네트워크-인지 방식으로 온라인 소셜 네트워크의 사용자에게 시판될 수 있다. 특정 실시예로, 처리 효과가 제1 사용자 세트에 대해 결정된 후, 소셜 네트워킹 시스템(160)은 제1 사용자 세트의 사용자들에 대한 처리의 결정된 처리 효과에 기반하여 처리를 수정할 수 있다. 광고, 피쳐 또는 제품은 예컨대 처리와 사용자의 상호작용 또는 경험을 최적화하도록 결정된 처리 효과에 기반하여 수정될 수 있다. 이후, 수정된 처리는 추가적인 사용자 세트에게 제공될 수 있고, 이어서 예컨대 충분히 개선되고 전체 온라인 소셜 네트워크에 개시될 때까지 추가 사용자에게 수정 및 시판될 수 있다. 비록 본 명세서는 특정 방식으로 처리를 제공하고 처리 효과를 결정하는 것을 기술하지만, 본 명세서는 임의의 적절한 방식으로 처리를 제공하거나 처리 효과를 결정하는 것을 고려한다. 게다가, 비록 본 명세서는 특정 추정치를 사용하는 것을 기술하지만, 본 명세서는 임의의 적절한 추정치를 사용하는 것을 고려한다.

[0072] 예로서 제한 없이, 각 노드가 처리 또는 제어에 독립하여 할당되는 가장 간단한 랜덤화 기법인 독립 노드 랜덤화(independent node randomization) 하에서 가장 간단한 네트워크 노출 조건인 완전 이웃 노출에 대한 노출 확률을 검토하자. 모든 노드가 확률 $p \in (0,1)$ 로 독립하여 처리된다면, 차수 d_i 의 사용자 i에 대해 처리에 대한 완전 이웃 노출의 확률은 간단히 $\Pr(Z \in \sigma_i^1) = p^{d_i+1}$ 로 주어지며, 제어에 대한 완전 이웃 노출의 확률은 $\Pr(Z \in \sigma_i^0) = (1-p)^{d_i+1}$ 로 주어진다. 이는 네트워크 노출의 주된 과제를 강조한다: 높은 차수를 가

진 노드가 완전 이웃 노출이나 그에 근접한 어느 곳이든 도달하도록 관리하는 가능성이 d_i 로 지수적으로 작을 수 있다. 직관적으로, 이런 작은 노출 확률은 Horvitz-Thompson 추정치의 분산을 크게 증가시킬 것이며, 이는 더 지능적인 랜덤화를 사용해야 함을 나타낸다.

[0073] 이런 Horvitz-Thompson 추정치의 분산을 감소시키기 위해, 독립적으로 할당된 클러스터들로 노드 레벨 대신에 클러스터 레벨에서 그래프 클러스터를 생성하고 할당을 랜덤화하는 일반적인 그래프 클러스터 랜덤화 접근법을 도입한다. 이후, 연결된 노드는 독립된 할당으로 발생하는 것보다 더 빈번히 동일한 처리 조건에 할당될 것이며, 사용자의 노출 조건 사이의 증가된 상관(correlations) 비용으로 조건에 네트워크 노출되는 예상 사용자 수를 증가시킨다.

[0074] 노드를 n_c 클러스터 C_1, \dots, C_{n_c} 로 분할하자. 그래프 G에서 i의 이웃을 $N_i \subseteq V$ 이라 하고, i 또는 i의 이

웃을 포함하는 클러스터 세트를 $S_i = \{C_j : (i \cup N_i) \cap C_j \neq \emptyset\}$ 이라 하자; S_i 를 i가 연결되는 클러스터 세트라고 한다. 이런 표기법을 사용하여, 우리는 이제 각기 다른 네트워크 노출의 확률을 검토할 것이다.

[0075] 노출 확률

[0076] 예로서 제한 없이, 완전 이웃 노출 조건에 대해, 처리에 대한 네트워크 노출의 확률은 간단히 $P\tau(Z \in \sigma^0) = p^{|S_i|}$ 이며, 제어에 대한 네트워크 노출의 확률은 $P\tau(Z \in \sigma_i^1) = (1-p)^{|S_i|}$ 이다. 이제, 우리는 절대 및 부분 이웃 노출 조건에 대한 노출 확률을 계산하는 것도 또한 다룰 수 있음을 보여준다.

[0077] 클러스터 랜덤화 하에서 차수 d_i 를 가진 노드 i가 처리되고 k의 그 이웃 노드가 처리되는 확률을 계산하는 과제를 고려하자. 이는 절대 및 부분 이웃 노출 모두를 고려할 때 적용된다. 먼저, 클러스터를 다시 인덱싱하여, i가 $|S_1| = s$ 클러스터에 연결된다면, i 그 자체는 클러스터 s에 상주하며 다른 연결된 클러스터를 $j = 1, \dots, s-1$ 라고 하자. i가 각 클러스터에 가지는 연결 수를 w_{i1}, \dots, w_{is} 라고 하며, 각 클러스터와 연관된 독립된 동전 던지기를 Bernoulli(p) 랜덤 변수 X_1, \dots, X_s 라고 하자. 그러면:

$$\begin{aligned} P\tau[Z \in \sigma^1] &= P\tau[X_s = 1] \cdot P\tau\left[\sum_{j=1}^{s-1} w_{ij} X_j \geq k - w_{is}\right], \\ P\tau[Z \in \sigma_i^0] &= P\tau[X_s = 0] \cdot P\tau\left[\sum_{j=1}^{s-1} w_{ij} X_j \leq d_i - k\right]. \end{aligned}$$

[0078]

[0079] 여기서, 랜덤 양 $\sum_j w_{ij} X_j$ 은 포아송-이항 분포(Poisson-binomial distribution)의 가중 등식을 따르며, 해당 확률은 하기의 회귀식(recursion)으로 정의되는 동적 프로그램을 명시적으로 사용하여 계산될 수 있다:

$$P\tau\left[\sum_{j=1}^s w_j X_j \geq T\right] = p P\tau\left[\sum_{j=1}^{s-1} w_{ij} X_j \geq T - w_{is}\right] + (1-p) P\tau\left[\sum_{j=1}^{s-1} w_{ij} X_j \geq T\right]$$

[0080]

[0081] T는 최대 노드 차수 d_{\max} 에 의해 한정지어지며, 이는 실행시간 $O(d_{\max} s)$ 을 가진 다항 시간 동적 프로그램을 생성함을 유의하자.

[0082] 독립 클러스터 랜덤화 하에서 노드 i가 처리되고 $\geq k$ 이웃 노드가 처리되는 확률은 다음과 같이 주어진다:

$$P\tau[Z \in \sigma_i^1] = p f(s-1, k - w_{is}, p, \bar{w})$$

[0083]

[0084] 여기서:

$$f(1, T; p, \bar{w}_i) = p\mathbb{1}[T < w_{i1}],$$

$$f(j, T; p, \bar{w}_i) = pf(j-1, T - w_{ij}; p, \bar{w}_i) + (1-p)f(j-1, T; p, \bar{w}_i).$$

독립 클러스터 랜덤화 하에서 노드 i 가 제어되고 $\geq k$ 이웃 노드가 제어되는 확률은 다음과 같이 주어진다:

$$P\tau[Z \in \sigma_i^1] = (1-p)[1-f](s-1, d_i - k + 1; p, \bar{w})$$

이런 부분 이웃 노출 조건(절대 및 부분)들이 내포되어 있음을 상기하자. 실제로, 소정의 노드 i 에 대해, 회귀 식은 단일의 $O(d_{\max} s)$ 이중 루프(double for-loop)를 고려하여 모든 가능한 임계값에 대한 확률을 도출하는데 사용될 수 있다. 실제로 이런 계산은 각 개인에 대한 노출 공간에서의 확률 분포를 반환한다.

도 4는 예시적인 확률 분포를 도시한다. 단일의 개인에 대한 노출 공간에 대한 확률 분포인 노출 조건 σ_i^0 및 σ_i^1 은 (a) i.i.d. 노드 랜덤화 및 (b) 이상적인 클러스터 랜덤화 모두에 대해 노란색으로 도시되며, 여기서 확률 질량(probability mass)은 관심 노출 조건에 모인다. 비록 도 4 및 본 명세서는 특정 확률 분포를 도시하고 기술하지만, 본 명세서는 임의의 적절한 확률 분포를 고려한다.

상술한 동적 프로그램은 단지 절대 및 부분 이웃 노출 조건에 대한 노출 확률을 정확히 계산하는 수단만을 제공한다. 불행하게도, k -코어 및 부분 q -코어 노출 조건의 정확한 확률을 효율적으로 계산하는 방법은 불명확하지만, 이런 노출 조건들은 공식적으로 해당하는 이웃 노출 조건들의 서브세트에 내포되었음을 상기하자. 이를 통해 적어도 코어 노출 확률을 상계(upper bound)할 수 있고, 하기의 명제를 통해 이런 연결을 공식화한다. 우리는 일반적으로 매우 작은 노출 확률에 관심이 있기 때문에, 이런 상계는 이미 이웃 노출 하에서 문제적으로 작은 확률을 가진 노드를 식별하는데 유용할 수 있다.

확률 노드 i 는 코어 노출 하에서 처리 조건에 네트워크 노출되며 유사한 이웃 노출 하의 확률보다 작거나 동일하다:

$$\begin{aligned} P\tau(Z \in \sigma_i^x | k\text{-core}) &\leq P\tau(Z \in \sigma_i^x | k\text{-hood}), \\ P\tau(Z \in \sigma_i^x | \text{fracq-core}) &\leq P\tau(Z \in \sigma_i^x | \text{fracq-hood}), \end{aligned}$$

코어 노출 확률의 직접적인 유용한 추정치 랜덤화의 Monte Carlo 샘플링을 통해 획득될 수 있지만, 여기서는 그 확률을 검토하지 않는다.

추정치 분산(Estimator variance)

간접 하의 Horvitz-Thompson 추정치의 분산은 Aronow 및 Samii([참고문헌 2])에 의해 연구되었으며, 또한 몇몇의 분산 감소 기법을 제시한다. 이런 접근법 하에서 분산을 추정하는 것은 노드 i 가 처리/제어에 네트워크 노출되고 노드 j 가 처리/제어에 네트워크 노출되는 결합 확률(joint probability)인 결합 노출 조건의 지식을 필요로 한다. 이는 랜덤 벡터 Z 가 노드 i 및 노드 j 에 대해 동시에 노출 조건에 있는 확률, 즉 처리에 대한 결합 노

출 조건에 대해 $P\tau(Z \in (\sigma_i^1 \cap \sigma_j^1))$ 이다. 분석적으로 추정치의 분산을 계산하는 것에 관심이 있다면, 2개의 세트의 교집합이 공집합일 수 있다는 점을 제외하고는 하나의 노드의 노출 확률과 비교할 때 이런 확률 계산에 대해 기본적으로 다른 것은 없다.

$$\hat{Y}^x(Z) = \frac{1}{n} \sum_i [Y_i(Z) \mathbb{1}[Z \in \sigma_i^x] / P\tau(Z \in \sigma_i^x)]$$

인 효과 추정치의 분산은 다음과 같이 주어진다:

[0097] (수식 3)

$$[0098] \text{ } Var[\bar{\tau}(Z)] = [Var[\bar{Y}^1(Z)] + Var[\bar{Y}^0(Z)] - 2Cov[\bar{Y}^1(Z), \bar{Y}^0(Z)]]$$

[0099] 노출 조건이 적절히 특정된다고 가정하면, 즉 $Y_i(\vec{z})$ 가 모든 $\vec{z} \in \sigma_i^x$ 에 대해 상수라고 가정하면, $Y_i(\sigma_i^x := Y_i(\vec{z} \in \sigma_i^x))$ 라는 표기를 도입할 수 있다. 추가적인 표기 $\pi_i^x := \Pr[Z \in \sigma_i^x]$ 및 $\pi_{ij}^{xy} := \Pr[Z \in (\sigma_i^x \cup \sigma_j^y)]$ 를 사용하면, 다음을 얻는다:

[0100] (수식 4)

$$[0101] \text{ } Var[\hat{Y}^x(Z)] = \frac{1}{n^2} \left[\sum_{i=1}^n \frac{1-\pi_i^x}{\pi_i^x} Y_i(\sigma_i^x)^2 + \sum_{i=1}^n \sum_{\substack{j=1 \\ j \neq i}}^n \frac{\pi_{ij}^{xx} - \pi_i^x \pi_j^x}{\pi_i^x \pi_j^x} Y_i(\sigma_i^x) Y_j(\sigma_j^x) \right],$$

[0102] 및

[0103] (수식 5)

$$[0104] \text{ } Cov[\hat{Y}^1(Z), \hat{Y}^0(Z)] = \frac{1}{n^2} \left[\sum_{i=1}^n \sum_{\substack{j=1 \\ j \neq i}}^n \frac{\pi_{ij}^{10} - \pi_i^1 \pi_j^0}{\pi_i^1 \pi_j^0} Y_i(\sigma_i^1) Y_j(\sigma_j^0) - \sum_{i=1}^n Y_i(\sigma_i^1) Y_i(\sigma_i^0) \right],$$

[0105] 상기 수식은 분산이 노출의 확률에 의해 매우 엄격히 제어되며 분산을 상계하기 위해 모든 노드 쌍 및 x와 y의 모든 조합에 대해 확률 π_i^x 에 대한 하계 및 결합 확률 π_{ij}^{xy} 에 대한 상계가 필요함을 자명하게 한다. 이제, 이웃 노출에 대해, 추정치의 분산이 그래프 클러스터 랜덤화에 대해 n에서 점근적으로 $O(1/n)$ 인 기본 충분 조건을 작성할 수 있다.

[0106] 잠재적 결과 $Y_i(\cdot)$ 는 n에서 모두 $O(1)$ 이라고 가정하자. G가 최대 차수 $O(1)$ 를 가지고 각 클러스터의 크기가 $O(1)$ 라면, 그래프 클러스터 랜덤화 하에서 충분, k-이웃 및 q-부분 이웃 노출에 대한 Horvitz-Thompson 추정치의 분산은 $O(1/n)$ 이다. G가 최대 차수 $O(1)$ 를 가지고 각 클러스터의 크기가 $O(1)$ 라고 가정하자. 모든 단일 합은 명백히 $O(n)$ 이다: 모든 노드는 한계 차수(bounded degree)를 가지므로 π_i^x 는 $O(1)$ 이다. 이중 합

(double sums)에 대해, i 및 j가 어떤 공통 클러스터 이웃도 가지지 않는, $|S_i \cap S_j| = 0$ 의 경우에만

$\pi_{ij}^{xx} = \pi_i^x \pi_j^x$ 임을 유의하자. $|S_i \cap S_j| > 0$ 일 때, 완전, k-이웃 및 q-부분 이웃 노출에 대해

$\pi_{ij}^{xx} > \pi_i^x \pi_j^x$ 이다. 게다가, $|S_i \cap S_j| > 0$ 라면 $\pi_{ij}^{10} < \pi_i^1 \pi_j^0$ 이며, 그 밖의 경우에는

$\pi_{ij}^{10} = \pi_i^1 \pi_j^0$ 이다. 그래서, 이중 합의 항은 $\pi_{ij} = \pi_i \pi_j$ 일 때 0이며, 그 항이 0이 아닌 경우

($|S_i \cap S_j| > 0$), 이들은 모두 양수이고 한계 차수로 인해 $O(1)$ 위로 유계된다. 이제, ($|S_i \cap S_j| > 0$)에

대한 노드의 수 j를 한계짓는다. 최대 노드 i는 $O(1)$ 에 연결되므로, $|S_i| = O(1)$ 이다. 모든 $C \in S_i$ 에

대해, 임의의 $j \in C$ 에 대해 그리고 클러스터 C 내 한 노드에 인접한 임의의 노드 j 에 대해 $(|S_i \cap S_j| > 0)$ 를 가진다. 이런 기여(contributions) 모두는 $O(1)$ 이며, 각 $C \in S_i$ 에 대한 노드의 $O(1)$ 기여를 제공한다. 이런 $O(1)$ 클러스터가 있기 때문에, 이것은 $(|S_i \cap S_j| > 0)$ 이도록 노드 i 에 대해 여전히 총 $O(1)$ 노드 j 이다. 따라서, 각 노드에 대해, 이중 합의 항들의 최대 $O(1)$ 는 양수이고 총 분산 $O(1/n)$ 를 생성한다.

[0107] 제한된-증가 그래프에 대한 분산

[0108] 특정 실시예로, 소셜 네트워킹 시스템(160)은 고차 노드에 대한 추정치 분산을 포함할 수 있는 랜덤화 기법을 사용하여 대규모의 그래프에 대한 간접 하에서 평균 처리 효과를 측정할 수 있다. 제한된-증가 그래프를 만족하는 임의의 그래프는 모두 $O(1)$ 이고 그래프의 차수에 선형적인 비편향된 효과 추정치를 생성할 수 있는 클러스터링을 허용한다. 반면에, 덜 주의깊은 클러스터링으로는 분산이 차수에 지수적으로 증가하기 쉽다. 먼저, 제한된-증가 그래프를 정의하자. 노드 v 의 r 홉 내에 있는 노드의 세트를 $B_r(v)$ 라고 하자. 모든 노드 $v \in V$ 및 모든 $r > 0$ 에 대해 $|B_{r+1}(v)| \leq k|B_r(v)|$ 라면 $G = (V, E)$ 는 제한된-증가 그래프이다. 상술한 바와 같이, m 차원의 유클리드 공간에서 균밀도 임베딩으로부터 도출된 그래프는 차수와 무관하게 증가 상수 $k+2^m$ 을 가진 제한된 증가를 나타낸다. 제한된-증가 가정에 대한 직관을 발전시키기 위해, 먼저 사이클(cycle)의 k 제곱인 특히 다루기 쉬운 제한된-증가 그래프군에 대한 그래프 클러스터 랜덤화를 사용하여 분산을 분석한다. 일반적인 제한된-증가 그래프에 대한 분산에 한계를 증명하여 이런 분석을 이어간다. 비록 본 명세서는 특정 방식으로 처리 효과를 측정하는 것을 기술하지만, 본 명세서는 임의의 적절한 방식으로 처리 효과를 측정하는 것을 고려한다.

[0109] 사이클 예들의 사이클 및 제곱

[0110] 예로서 제한 없이, 먼저 n 노드를 가진 단일의 사이클로 구성된 간단한 그래프를 고려할 것이다. 이런 그래프에서는 완전 이웃 노출 모델을 고려하는데, 여기서 우리의 관심사는 σ_i^1 사이의 평균 처리 효과, 노드가 처리되고 이웃들 모두가 처리되는 시간, 및 σ_i^1 , 노드가 처리되지 않고 어떤 이웃들도 처리되지 않는 시간이다. 처리 및 제어에 대한 노드의 고정된 응답에 대해, 모든 노드가 처리에 대한 네트워크 노출에 $Y_i(\sigma_i^1) = \bar{Y}$ 로 그리고 제어에 대한 네트워크 노출에 $Y_i(\sigma_i^0) = 0$ 로 균일하게 응답한다고 가정하자. 사이클 그래프는 함께 랜덤화되는 c 노드들의 인접 블록을 가지고 사이클 구조를 사용하는 직관적으로 자명한 클러스터링을 허용한다. 마지막 가정으로서, 클러스터들은 $p=1/2$ 로 균형된 랜덤화 하에서 선택된다고 가정하자. 우리의 목표는 Horvitz-Thompson 평균 처리 효과의 분산이 어떻게 이런 클러스터들의 크기 c 에 의존하는지를 결정하는 것이다. 그래프, 노출 조건, 응답 및 클러스터링의 이런 기본적인 조합에 대해, 점근적 분산(asymptotic variance)을 정확히 도출할 수 있다.

[0111] 상기 (수식 3)에 제시된 분산을 고려하자. 우리의 예에서 모든 노드가 제어 조건에 대해 0으로 응답하기 때문에, 제어 조건에 대한 노출 확률이 엄격히 양의 값인 한, $Var(\hat{Y}(\sigma_0))$ 및 $Cov(\hat{Y}(\sigma_1), \hat{Y}(\sigma_0))$ 모두 0이다. 계산들은 단지 처리 조건에의 노출에 대한 확률 π_i^1 에만 의존할 것이기 때문에, 윗첨자를 생략한다. 그러면, 분산은 다음과 같다:

[0112] (수식 6)

[0113]

$$Var[\hat{r}(Z)] = \frac{\bar{Y}^2}{n^2} \left[\sum_{i=1}^n \left(\frac{1}{\pi_i} - 1 \right) + \sum_{i=1}^n \sum_{\substack{j=1 \\ j \neq i}}^n \left(\frac{\pi_{ij}}{\pi_i \pi_j} - 1 \right) \right]$$

[0114]

이중 합의 항들은 노드 쌍들에 대해 오직 0이 아님, 즉 $\pi_{ij} \neq \pi_i \pi_j$ 임을 유의하자.

[0115]

먼저, 각 노드가 그 자신의 클러스터인 경우를 고려하자. 한 노드 및 그의 두 이웃들 모두가 노출되는 확률은 3개의 독립된 동전이 앞이 나올 확률과 동일하다. 랜덤화가 균형화되는 경우(예컨대, $p=1/2$), $\forall i$ 에 대해 $\pi_i = 1/8$ 를 얻는다. 공동-할당(co-assignment) 확률은 노드 i 및 j 가 이웃들이거나 이웃을 공유하는지에 의존함에 유의하자. 이로부터, $|i-j|=1$ 이라면 $\pi_{ij} = 1/16$ 이고 $|i-j|=2$ 이라면 $\pi_{ij} = 1/32$ 이며 $|i-j| > 2$ 이라면 확률이 독립적임을 도출한다. 우리는 $Var(\hat{r}(Z)) = (15/2)\bar{Y}^2 \frac{1}{n} + O(1/n^2)$ 를 얻는다.

[0116]

이제, $c \geq 2$ 노드의 블록을 랜덤화하는 것을 고려하자, 여기서 c 는 n 에 의존하지 않는다. 균등하게 사이클적으로 변이된 확률 중 하나를 고려하자. 그 계산은 $c=2$ 및 $c \geq 3$ 을 별도로 다루는 것을 필요로 하지만, c 의 함수로서 $c \geq 3$ 에 대한 식은 $c=2$ 에 대해 유지하며, 따라서 간결화를 위해 특별한 경우를 생략한다. 분산 계산은 $\Delta = c+1$ 까지 거리 $\Delta = |i-j|$ 에 의존하며, $c \geq 3$ 에 대해 이는 다음과 같이 평가된다:

[0117]

$$Var[\hat{r}(Z)] = \frac{\bar{Y}^2}{n^2} \left[\left(n + \frac{4n}{c} \right) + \underbrace{\frac{2n}{c}(c+2)}_{\Delta=1} + \underbrace{\frac{2n}{c} \sum_{k=2}^{c-2} (c-k+2)}_{1 < \Delta < c-1} + \underbrace{\frac{2n}{c} 3}_{\Delta=c-1} + \underbrace{\frac{2n}{c} 2}_{\Delta=c} + \underbrace{\frac{2n}{c}}_{\Delta=c+1} \right] + O\left(\frac{1}{n^2}\right)$$

[0118]

이는 모든 $c \geq 2$ 에 대해 유지하는 $Var(\hat{r}(Z)) = \left(\frac{c}{2} + 2 + \frac{4}{c} \right) \bar{Y}^2 \frac{1}{n} + O(1/n^2)$ 로 감소한다.

[0119]

도 5는 예시적인 사이클 그래프 및 분산 계산 도표를 도시한다. 상술한 계산을 고려하면, 모든 c 에 대한 추정치의 점근적 분산이 도 5에 도시된다. 분산은 간단한 사이클에 대한 이웃의 크기에 정확히 대응하는 $c=3$ 크기의 클러스터를 랜덤화할 때 최소화된다. (a) $c=2$ 노드의 그룹에서 클러스터링되는, 노드가 처리에 대해 \bar{Y} 으로 그리고 제어에 대해 0으로 응답하는 사이클 그래프가 도시된다. (b) 클러스터당 노드의 수의 함수로서 이런 그래프에 대한 추정치의 점근적 분산이 클러스터당 $c=1$ 노드에 대한 추정치 분산에 의해 정규화된다. (c) 클러스터당 노드 수의 함수로서 $k=1, \dots, 5$ 에 대한 사이클 그래프의 k 제곱에 대한 추정치의 분산이 시뮬레이션된다. 각 k 에 대해, 클러스터 크기 $c=2k+1$ 에 대한 분산은 k 에 대해 선형적으로 증가한다. 비록 도 5 및 본 명세서는 특정 그래프 및 분산 계산을 도시하고 기술하지만, 본 명세서는 임의의 적절한 그래프 및 분산 계산을 고려한다.

[0120]

이런 관찰을 구축하기 위해, 이제 분석 도출(analytic derivation)이 이미 다루기 어려운, 소위 사이클의 k 제곱이라고 하는 사이클의 고차 확장에 대한 시뮬레이션된 분산이 검토된다. 따라서, 효과 크기 추정치의 분산이 어떻게 이런 고차 그래프에 대한 클러스터 크기에 의존하는지를 검토하는 클러스터 랜덤화 절차의 시뮬레이션이 사용된다.

[0121]

사이클 그래프의 k 제곱은 각 노드가 각 측면에서 k 최근접 이웃들과 연결되는 사이클로 구성되며, 모든 노드가 $d=2k$ 인 차수를 가지는 정규 그래프를 산출한다. $n=5000$ 노드를 가진 그래프에 대한 일백만 클러스터 랜덤화를 샘플링함으로써, 이런 샘플들에 대한 추정치의 샘플 분산이 계산될 수 있다. 그 결과가 $k=1$ 내지 $k=5$ 에 대해 도 5에 도시된다. $k=1$ 에 대한 시뮬레이션은 중첩된 점근적 계산에 정확히 일치한다.

[0122]

최적 클러스터 크기 c 는 차수에 대략 선형적으로 스케일링(scale)하는 것으로 보이며, k 가 증가할 때의 각 곡선의 최소값인 최적 클러스터링 크기의 분산은 k 에 선형적으로 스케일링하는 것으로 보인다. 클러스터 크기 c 의

함수로서 정확한 분산이 도출하기 어려운 한편, 하기의 상계를 제공할 수 있고, 크기 $c=d+1$ 의 클러스터에 대한 추정치의 분산이 어떻게 그래프의 차수 d 에 선형적으로 스케일링하는지를 보여준다. 이는 노드의 이웃의 크기에 맞춰지는 사이클의 인접 블록을 다루어야 함을 제안한다.

[0123] 이런 상계를 도출할 때, 균일한 응답 $Y_i(\sigma_i^1) = \bar{Y}$ 을 가정하는 것이 더 이상 필요한 것은 아니며, 그 대신 그 응답이 임의의 값 $Y_i(\sigma_i^1) \leq Y_M$ 에 의해 상계된다고 간단히 가정된다.

[0124] 클러스터가 크기 $c=d+1$ 을 가질 때, 각 노드는 최대 2개의 클러스터에 연결될 수 있으며, 이는 모든 i 에 대해 $1/\pi_i \leq 1/p^2$ 를 의미한다. 그래서:

[0125]

$$Var[\hat{\tau}(Z)] \leq \frac{Y_M^2}{n^2} \left[\sum_{i=1}^n (p^{-2} - 1) + \sum_{i=1}^n \sum_{\substack{j=1 \\ j \neq i}}^n \left(\frac{\pi_{ij}}{\pi_i \pi_j} - 1 \right) \right]$$

[0126] 이제, 각 노드는 최대 $3d+1$ 개 다른 노드들(이들이 2개의 클러스터, 좌측 클러스터에서 좌측으로 $d/2$ 그리고 우측 클러스터의 우측으로 $d/2$ 에 인접한 경우 최대 $2d+1$ 개의 다른 노드들)을 가진 비-독립 결합 할당 확률 $(\pi_{ij} \neq \pi_i \pi_j)$ 을 가진다. 결합 할당 확률 π_{ij} 은 2개의 노드가 모두 클러스터의 중심에 있지 않을 수 있기 때문에 최대 p^2 이다. 각각의 i 에 대해, j 로 인접성되는 합은 한계지어질 수 있고, 다음을 산출한다:

[0127]

$$Var[\hat{\tau}(Z)] \leq Y_m^2 (p^{-2} - 1)(3d + 2) \frac{1}{n}$$

[0128] 이 결과는 단지 d 에 선형적인 분산을 가진 매우 고차인 d 의 사이클 그래프에 대한 네트워크 효과를 실험적으로 측정할 수 있음을 보여주며, 여기서 노드는 $d+1$ 노드의 인접 블록에 클러스터링된다. 이제, 분산을 경계짓는 (bounding) 전략이 임의의 기하학적 구조의 지식을 필요로 하지 않는 클러스터링 알고리즘을 사용하여 어떻게 훨씬 더 일반적인 클래스의 그래프에 적용되는지를 보여준다.

[0129] 제한된-증가 그래프의 클러스터링

[0130] 이제, 이번 섹션의 주요 결과인 제한된-증가 그래프의 클래스에 대한 클러스터 랜덤화 기법을 발전시킨다. 제1 컴포넌트는 각 노드가 그 노드의 차수에 관계없이 최대 상수의 클러스터에 연결되는 그래프들에 대한 클러스터링 알고리즘이다. 이후, 이는 임의의 제한된-증가 그래프에 대한 분산이 차수에 선형인 함수에 의해 상계될 수 있음을 의미할 것이다. 우리의 클러스터링은 인접 지역에 의한 사이클의 훌륭한 분해가 우리의 클래스에서 임의의 그래프로 일반화될 수 있음을 보여준다. 즉, 기하학 구조가 중요한 것이 아니다; 제한된-증가 성질이면 충분하다.

[0131] 제한된-증가 그래프 $G = (V, E)$ 를 고려하자. G 는 d -정규(d -regular)인 경우가 제시되지만, 후술하는 바와 같이 정규성(regularity)은 연결된 클러스터의 수에서 더 약화되지만 여전히 일정한 한계의 비용에서의 임의의 차수의 분포로 완화될 수 있다.

[0132] 제한된-증가 조건이란 모든 v 및 모든 $r > 0$ 에 대해 $|B_{r+1}(v)| \leq \kappa |B_r(v)|$ 이도록 하는 κ 가 존재하는 것을 말함을 상기하자. 중요하게, $\tau = 0$ 은 다르다: $B_0(v)$ 은 단집합 $\{v\}$ 인 한편, $B_1(v)$ 은 v 의 이웃이며 따라서 크기 $d+1$ 을 가진다. 따라서, $|B_1(v)|/|B_0(v)| = d+1$ 이며, $\tau > 0$ 에 대해 $|B_{r+1}(v)|/|B_r(v)|$ 비에서 κ 의 한계보다 잠재적으로 훨씬 더 크다. 이것이 제한된-증가 조건의 핵심이다: 반경 0에서 1까지 비-제한된 증가($d+1$ 의 인자)를 가지지만, 이후 그 증가는 d 와는 별도로 한계될 수 있는 κ 의 인자로 줄어든다.

[0133] 거리 공간의 언어로, 우리는 G 의 최단-경로 거리에 대한 3-망(net)을 사용하여 그래프를 클러스터링할 것이다 ([참고문헌 8]). 형식상, 거리 공간 X 에서, τ -net $Y \subset X$ 은 서로로부터 공통으로 적어도 r 의 거리에 있는 점들의 집합이다. 따라서, 우리의 구성을 그래프의 3-망 클러스터링이라고 한다. 3-망 클러스터링을 구축하기 위해, 반복적으로 노드들 v_1, v_2, \dots 을 식별하며, 그렇게 함으로써 노드들을 "표시(marking)"한다. 이후, 이런 노드들에 동반되는 클러스터들 C_1, C_2, \dots 을 식별한다. 더 상세하게는, 2개의 원칙 단계로 구성되는 다음의 절차를 수행한다:

[0134] ● 초기에 모든 노드는 표시되지 않음.

[0135] ● 미표시된 노드들이 있는 동안, 단계 j 에서 임의의 미표시된 노드 v 를 찾고 v 를 노드 v_j 로 선택하며, 모든 노드를 $B_2(v_j)$ 에 표시한다.

[0136] ● k 개의 이런 노드가 정의되면, $S = \{v_1, v_2, \dots, v_k\}$ 라고 한다.

[0137] ● G 의 모든 노드 w 에 대해, w 를 최근접 노드 $v_i \in S$ 에 할당하고, (예컨대, 최저 인덱스의 순서로) 지속적으로 연대(ties)를 끊는다.

[0138] ● 모든 v_j 에 대해, v_j 에 할당되는 모든 노드의 세트를 C_j 라고 한다.

[0139] 그러면, 세트 C_1, \dots, C_k 는 우리의 3-망 클러스터링이다. 이런 클러스터링의 핵심 특성은 각 노드가 차수에 무관하게 K 의 함수에 의해 한계지어질 수 있는 다수의 클러스터에 연결됨을 확립하는 다음의 결과이다. 예로서 제한 없이, 그래프 $G = (V, E)$ 의 임의의 3-망 클러스터링을 고려하자. 모든 $w \in V$ 에 대해, 이웃 $B_1(w)$ 는 최대 K^3 의 별개의 클러스터를 가진 공집합이 아닌 교집합을 가진다. 먼저, 모든 $v_i \in S$ 에 대해 우리가 가진다고 말한다. 실제로, C_j 에서 임의의 노드 $w \neq v_j$ 를 고려하자. 우리는 $w \notin S$ 를 가지는데, 그렇지 않으면 w 는 그 자체를 가지고 식별되는 클러스터에 속하기 때문이다. 이제, w 가 표시되었던 반복 i 를 고려한다: 우리는 $w \in B_2(v_i)$ 를 가진다. $w \in C_j$ 이고 S 에서 최근접 노드에 할당되기 때문에, $w \in B_2(v_i)$ 를 따른다. 따라서, $C_j \subseteq B_2(v_i)$ 이다. 다음으로, 모든 $v_i, v_j \in S$ 에 대하여 세트 $B_1(v_i)$ 및 $B_1(v_j)$ 가 비결합(disjoint)이라고 말한다. 모순으로서 $B_1(v_i) \cap B_1(v_j) \neq \emptyset$ 라고 가정하면, $v_i, v_j \in S$ 이며 그 역도 또한 같다. 그러나, 이후 먼저 S 에 추가되었던 v_i 및 v_j 중의 노드를 고려하면, v_i 및 v_j 중 다른 것이 그 반복에서 표시되었을 수 있고, 따라서 물론 S 로 추가되지 않았을 수 있다. 이런 모순(contradiction)은 $B_1(v_i)$ 및 $B_1(v_j)$ 가 비결합임을 확립한다. 그 증거를 완전히 하기 위해, 모순으로서 $B_1(w)$ 는 최대 K^3 의 별개의 클러스터를 가진 공집합이 아닌 교집합을 가진다고 가정하자: 일부 $t > K$ 에 대해, $B_1(w)$ 에서 별개의 노드들을 u_1, u_2, \dots, u_t 라고 하고, $h=1, 2, \dots, t$ 에 대해 $u_h \in C_{i_h}$ 이도록 S 내 별개의 노드들을 v_{i_1}, \dots, v_{i_t} 라고 하자. $C_{i_h} \subseteq B_2(v_{i_h})$ 및 C_{i_h} 는 w 에 인접한 노드를 포함(또는 w 그 자체를 포함)하기 때문에, $v_{i_h} \in B_3(w)$ 를 가지며, 따라서, $B_1(v_{i_h}) \subseteq B_4(w)$ 이다. 이웃들 $B_1(v_{i_1}), B_1(v_{i_2}), \dots, B_1(v_{i_t})$ 은 상술한 바대로 모두 쌍으로 비결합이며, 이들은 모두 $|B_4(w)| \geq t(d+1) > K^3(d+1)$ 를 의미하는

$B_4(w)$ 에 포함된다. 그러나, 한계 증가 부등식 $|B_{r+1}(w)| \leq \kappa |B_r(w)|$ 을 3회 적용하면, 모순인 $|B_4(w)| \leq \kappa^3(d+1)$ 를 가진다. 이것은 $B_1(w)$ 이 최대 κ^3 의 별개의 클러스터를 가진 공집합이 아닌 교집합을 가질 수 있음을 확립한다.

[0140] 상기 결과는 d-정규 그래프에 대해 공식화된다. 그러나, 실제로 차수에 대한 임의의 조건 없이 심지어 임의의 제한된-증가 그래프에 대해 단지 κ 에 의존하는 약화된 한계(weaker bound)를 보여줄 수 있다. κ^6 의 이런 약화된 한계는 임의의 제한된-증가 그래프가 노드 차수에서 "한계 변화도(bounded gradient)"를 나타내며 이로써 그래프에서 서로에 근접해 있는 노드들이 유사한 차수를 가져야 함을 관찰함으로써 확립될 수 있다. 이런 사실을 이전의 것과 결합하면 소정의 한계를 얻는다.

[0141] 분산 한계(Variance bounds)

[0142] 이제, 상기 결과를 적용하여 효과 추정치 $\hat{\tau}$ 의 분산을 한계짓는다. 하기의 섹션에서, 모든 응답이 처리 및 제어에의 노출, $x = 0, 1$ 모두에 대해 상계 및 양의 하계 $Y_i^x \in [Y_m, Y_M]$ 를 따른다고 가정하자. 양의 하계에 대한 근거는 이들 없이 사용자가 모든 처리에 대한 모두 0을 응답할 수 있다는 것이며 처리 기법에 관계없이 분산을 0으로 만든다. 또한, 랜덤화 확률 p 가 저하되지 않는다고, 즉 $p \in (0, 1)$ 이라고 가정하자. 관리가능한 표시들을 유지하도록 d-정규 그래프에 대한 결과가 제시되지만, 유사한 결과가 임의의 차수에 대해 도출될 수 있다.

[0143] 먼저, 노드-레벨 랜덤화 하에서 분산에 대한 지수적 하계가 확립되며, 이후 우리의 3-망 클러스터 랜덤화 기법 하에서 분산에 대한 대조적인 선형적 상계가 나타난다. 예로서 제한 없이, n개의 노드를 가진 그래프의 노드 랜덤화에 대한 완전 이웃 노출 하에서 Horvitz-Thompson 추정치의 분산이 그래프의 차수 d에서의 지수 함수에 의해 하계된다고, 즉 $Var[\hat{\tau}(Z)] \geq 0(1/n)(p^{-(d+1)} + (1-p)^{-(d+1)} - 1)$ 라고 가정하자. 동일한 노출을 가지는 2개의 노드에 대한 결합 할당 확률은 적어도 각각의 확률의 곱, 즉 $x = 0, 1$ 에 대해 $\pi_{ij}^{xx} \geq \pi_i^x \pi_j^x$ 이다. 따라서, (수식 4)의 이중 합은 음이 아니다. 마찬가지로, 노출 조건을 대비시키기 위해, (수식 5)를 (수식 3)에 대한 음이 아닌 기여로 만드는, $x \neq y$ 에 대한 $\pi_{ij}^{xy} \leq \pi_i^x \pi_j^y$ 를 가진다. 우리는 (수식 4)의 주요 항에 대한 하계에 집중한다. 확률들 $\pi_i^1 = p^{d+1}$ 및 $\pi_i^0 = (1-p)^{d+1}$ 및 하계의 응답들을 입력하는 것은 소정의 결과를 제공한다:

$$Var[\hat{\tau}(Z)] \geq \frac{1}{n^2} \left[\sum_{i=1}^n \left(\frac{1}{\pi_i^1} - 1 \right) (Y_i^1)^2 + \sum_{i=1}^n \left(\frac{1}{\pi_i^0} - 1 \right) (Y_i^0)^2 \right] \geq \frac{Y_m^2}{n} (p^{-(d+1)} + (1-p)^{-(d+1)} - 2)$$

[0144]

[0145] 임의의 차수의 분포를 가진 그래프에 대해, 이런 한계는 각 노드의 차수에 지수적인 $Var[\hat{\tau}(Z)] \geq O(1/n) \sum_{i=1}^n (p^{-(d_i+1)} + (1-p)^{-(d_i+1)} - 2)$ 가 되며, 이는 심지어 하나의 고차 노드가 용이하게 분산을 증가시킬 수 있음을 의미한다.

[0146] 이제, 우리의 3-망 클러스터링을 사용할 때 증가-제한 그래프에 대한 선형적 상계를 살펴본다. 예로서 제한 없이, 제한된-증가 그래프의 3-망 클러스터 랜덤화에 대한 완전, q-부분 또는 k-절대 이웃 노출하에서 Horvitz-Thompson 추정치의 분산이 그래프의 차수 d에 선형적인 함수에 의해 상계된다고 가정하자. 추정치의 분산은 $Var(\hat{\tau}(Z)) = Var(\hat{Y}^1) + Var(\hat{Y}^0) - 2Cov(\hat{Y}^1, \hat{Y}^0)$ 로 주어짐을 상기하자. 우리는 $\hat{Y}^1(Z)$ 의 분산을 상계하는 것으로 시작하고, $\hat{Y}^0(Z)$ 에 대한 상계는 동일한 원칙을 따른다. 공분산(covariance) 항을 한계지음

으로써 결론이 난다. 각 노드는 최대 κ^3 클러스터에 연결된다. 따라서, 완전과 부분 이웃 노출 모두에 대해 하계 $\pi_i^1 \geq p \wedge \kappa^3$ 를 가진다:

$$Var[\hat{Y}^1(Z)] \leq \frac{Y_M^2}{n^2} \left[n \left(\frac{1}{p \wedge \kappa^3} - 1 \right) + \sum_{i=1}^n \sum_{\substack{j=1 \\ j \neq i}}^n \left(\frac{\pi_{ij}^1}{\pi_i^1 \pi_j^1} - 1 \right) \right]$$

각 노드 i 에 대해, 내부의 2개의 합은 할당이 종속적인 이런 노드들 j 에서 단지 0이 아니다. i 및 j 에 대한 할당이 종속적이라면, 이들은 클러스터 센터의 세트에서 노드 v_h 와 연관된 동일한 클러스터 C_h 에 이웃들을 각각 가져야 한다. 우리는 이전에 $C_h \subseteq B_2(v_h)$ 로 설정했기 때문에, i 및 j 는 각각 v_h 의 거리 3 내에 있으며, 따라서 서로 거리 6의 내에 있다. 따라서, 그 할당이 i 의 할당에 종속적인 임의의 j 는 $B_6(i)$ 내에 있어야 하며, 그런 제한된-증가 조건에 의해, 그런 노드 j 가 최대 $|B_6(i)| \leq \kappa^5 |B_1(i)| = \kappa^5(d+1)$ 로 있을 수 있다. 따라서, 이런 j 에 대한 합은 최대 $\kappa^5(d+1)$ 항들을 가진다. 또한, 2개의 노드가 적어도 하나의 클러스터에 종속해야 하기 때문에, $\pi_{ij}^1 \leq p$ 를 적용한다. 우리는 다음을 얻는다:

$$Var[\hat{Y}^1(Z)] \leq Y_M^2 \left[\left((p \wedge \kappa^3) - 1 \right) + \kappa^5(d+1) \left((p \wedge (-2\kappa^3 - 1)) - 1 \right) \right] \frac{1}{n}$$

이제, 양의 값, $-2Cov(\hat{Y}^1, \hat{Y}^0)$ 인 분산에 대한 공분산 항의 기여를 고려하자. (수식 5)로부터 시작하며, 응답 Y_i 에 대한 상계를 적용하여 다음을 얻는다:

$$-2Cov[\hat{Y}^1(Z), \hat{Y}^0(Z)] \leq -\frac{2Y_M^2}{n^2} \sum_{i=1}^n \sum_{\substack{j=1 \\ j \neq i}}^n \left(\frac{\pi_{ij}^1}{\pi_i^1 \pi_j^1} - 1 \right) + \frac{2Y_M^2}{n}$$

이전의 유사한 식에 대하여, 각 i 에 대해, 내부의 합은 최대 $\kappa^5(d+1)$ 의 다른 노드들 j 에 대해 0이 아니다.

나머지 항들에 대해, $-(\pi_{ij}^{10} / (\pi_i^1 \pi_j^0) - 1)$ 의 값은 기껏해야 1에 의해 상계된다. 따라서 우리는 다음을 얻는다:

$$-2Cov[\hat{Y}^1(Z), \hat{Y}^0(Z)] \leq \frac{2Y_M^2}{n} [\kappa^5(d+1) + 1]$$

상계를 결합하여, 원하는 대로, 차수에 선형적인 총 상계가 얻어진다.

우리가 사용한 제한된-증가 조건은 정규 그래프에 대해 도출되었지만, 상술한 바와 같이, 임의의 차수 분포를 가진 제한된-증가 그래프에 대해, 클러스터 종속성에 대한 약화되지만 여전히 일정한 한계를 적용하여 차수에 여전히 선형적인 분산 한계를 얻을 수 있다.

도 6은 처리 효과를 결정하는 예시적인 방법(600)을 도시한다. 이 방법은 단계 610에서 시작할 수 있는데, 소셜 네트워킹 시스템(160)은 복수의 노드 및 노드를 연결하는 복수의 에지를 포함하는 소셜 그래프(200)에 액세스할 수 있다. 2개의 노드 사이의 각각의 에지는 노드 사이의 이격도 1도를 나타낼 수 있다. 복수의 노드는 온라인 소셜 네트워크와 연관된 복수의 사용자에게 각각 대응할 수 있다. 특정 실시예로, 복수의 노드는 온라인 소셜 네트워크와 연관된 복수의 사용자 또는 컨셉에 각각 대응할 수 있다. 단계 620에서, 소셜 네트워킹 시스템(160)은 그래프 클러스터링을 사용하여 소셜 그래프에서 복수의 클러스터를 식별할 수 있다. 각각의 클러스터는 복수의

노드로부터의 별개의 노드 세트를 포함할 수 있다. 각 클러스터 내 임계 수의 노드는 클러스터 내 다른 노드들에 대해 네트워크 노출될 수 있다. 단계 630에서, 소셜 네트워킹 시스템(160)은 복수의 클러스터 중 제1 클러스터 세트에 대응하는 제1 사용자 세트에 처리를 제공할 수 있다. 단계 640에서, 소셜 네트워킹 시스템(160)은 각 사용자에게 대한 처리에의 네트워크 노출에 기반하여 제1 사용자 세트의 사용자들에 대한 처리의 처리 효과를 결정할 수 있다. 적절한 경우, 특정 실시예는 도 6의 방법의 하나 이상의 단계를 반복할 수 있다. 비록 본 명세서는 도 6의 방법의 특정 단계가 특정 순서로 발생하는 것을 기술하고 도시하지만, 본 명세서는 도 6의 방법의 임의의 적절한 단계가 임의의 적절한 순서로 발생하는 것을 고려한다. 게다가, 비록 본 명세서는 도 6의 방법의 특정 단계를 포함하는 처리 효과를 결정하기 위한 예시적인 방법을 기술하고 도시하지만, 본 명세서는 적절한 경우 도 6의 방법의 단계들의 전부나 일부를 포함하거나 어떤 단계도 포함하지 않을 수 있는 임의의 적절한 단계를 포함하는 처리 효과를 결정하는 임의의 적절한 방법을 고려한다. 게다가, 비록 본 명세서는 도 6의 방법의 특정 단계를 수행하는 특정 컴포넌트, 장치 또는 시스템을 기술하고 도시하지만, 본 명세서는 도 6의 방법의 임의의 적절한 단계를 수행하는 임의의 적절한 컴포넌트, 장치 또는 시스템의 임의의 적절한 조합을 고려한다.

[0157] 참고문헌

- [0158] [참고문헌 1] E. Airoldi, E. Kao, P. Toulis, D. Rubin. "Causal estimation of peer influence effects. In ICML, 2013".
- [0159] [참고문헌 2] P. Aronow and C. Samii. "Estimating average causal effects under general interference. Working Paper, May 27, 2013", available at <http://arxiv.org/abs/1305.6156>.
- [0160] [참고문헌 3] L. Backstrom and J. Kleinberg. "Network bucket testing. In WWW, 2011".
- [0161] [참고문헌 4] B. Bollobas. "Random graphs, Chapter 6. Cambridge Univ. Press, 2001 (2nd ed.)".
- [0162] [참고문헌 5] D. Cellai, A. Lawlor, K. Dawson, J. Gleeson. "Critical phenomena in heterogeneous k-core percolation. Phys Rev E, 87(2):022134, 2013".
- [0163] [참고문헌 6] S. Fienberg. "A brief history of statistical models for network analysis and open challenges. J. Comp. Graph. Stat., 2012".
- [0164] [참고문헌 7] S. Fortunato. "Community detection in graphs. Physics Reports, 486(3):75-174, 2010".
- [0165] [참고문헌 8] A. Gupta, R. Krauthgamer, J. Lee. "Bounded geometries, fractals, and low-distortion embeddings. In FOCS, 2003".
- [0166] [참고문헌 9] D. Horvitz, D. Thompson. "A generalization of sampling without replacement from a finite universe. JASA, 1952".
- [0167] [참고문헌 10] D. Karger, M. Ruhl. "Finding nearest neighbors in growth-restricted metrics. In STOC, 2002".
- [0168] [참고문헌 11] L. Katzir, E. Liberty, O. Somekh. "Framework and algorithms for network bucket testing. In WWW, 2012".
- [0169] [참고문헌 12] R. Kohavi, A. Deng, B. Frasca, R. Longbotham, T. Walker, Y. Xu. "Trustworthy online controlled experiments: five puzzling outcomes explained. In KDD, 2012".
- [0170] [참고문헌 13] C. Manski. "Identification of treatment response with social interactions. The Econometrics Journal, 16(1):S1-S23, 2013".
- [0171] [참고문헌 14] D. Rubin. "Estimating causal effects of treatments in randomized and nonrandomized studies. J. Ed. Psych., 1974".
- [0172] [참고문헌 15] E. Tchetgen, T. VanderWeele. "On causal inference in the presence of interference. Stat. Meth. Med. Res., 2012".
- [0173] [참고문헌 16] J. Ugander, L. Backstrom. "Balanced label propagation for partitioning massive graphs. In WSDM, 2013".
- [0174] [참고문헌 17] D. J. Watts and S. H. Strogatz. "Collective dynamics of 'small-world' networks. Nature,

393(6684):440-442, 1998".

[0175] 본 명세서에서 인용된 모든 참고문헌은 전체로서 제시되는 바대로 참조로 통합된다.

[0176] 소셜 그래프 친밀성 및 계수

[0177] 특정 실시예로, 소셜 네트워킹 시스템(160)은 서로에 대한 다양한 소셜 그래프 엔티티의 소셜 그래프 친밀성(본 명세서에서는 "친밀성"이라 할 수 있음)을 결정할 수 있다. 친밀성은 가령 사용자들, 컨셉들, 콘텐츠들, 행위들, 광고들, 온라인 소셜 네트워크와 관련된 다른 객체들 또는 이들의 임의의 적절한 조합과 같은 온라인 소셜 네트워크와 관련된 특정 객체들 사이의 관계의 세기 또는 관심의 정도를 나타낼 수 있다. 친밀성은 또한, 제3자 시스템(170) 또는 다른 적절한 시스템과 연관된 객체에 관해 결정될 수 있다. 각각의 사용자, 대상물, 또는 콘텐츠의 타입에 대해 소셜 그래프 엔티티를 위한 전체 친밀도가 확립될 수 있다. 전체 친밀도는 소셜 그래프 엔티티와 연관된 행위 또는 관계의 지속적인 모니터링에 기초하여 변경될 수 있다. 본 명세서는 특정 방식으로 특정 친밀도를 결정하는 것을 기술하지만, 본 명세서는 임의의 적절한 방식으로 임의의 적절한 친밀도를 결정하는 것을 고려한다.

[0178] 특정 실시예로, 소셜 네트워킹 시스템(160)은 친밀성 계수(본 명세서에서는 "계수"라고 할 수 있음)를 사용하여 소셜 그래프 친밀성을 측정하거나 수량화할 수 있다. 계수는 온라인 소셜 네트워크와 관련된 특정 객체들 사이의 관계의 세기를 표현하거나 수량화할 수 있다. 또한, 계수는 사용자가 특정 행위를 수행할 예상 확률을 그 행위에 대한 사용자의 관심에 기초하여 측정하는 확률 또는 합수를 나타낼 수 있다. 이 방식으로, 사용자의 장래의 행위는 사용자의 이전의 행위들에 기초하여 예측될 수 있는데, 이때 계수는 적어도 부분적으로 사용자의 행위의 이력에 의해 계산될 수 있다. 계수는 온라인 소셜 네트워크의 내부 또는 외부에 있을 수 있는 임의의 수의 행위를 예측하는데 사용될 수 있다. 예로서 제한 없이, 이런 행위들은 가령 메시지를 송신하기, 콘텐츠를 게시하기 또는 콘텐츠에 대해 코멘트하기와 같은 다양한 타입의 통신들; 가령 프로필 페이지, 미디어 또는 다른 적절한 콘텐츠에 접근하거나 열람하는 것과 같은 다양한 타입의 관찰 행위들; 가령 동일한 그룹에 있기, 동일한 사진에 태그하기, 동일한 위치에 체크인하기 또는 동일한 이벤트에 참여하기와 같은 다양한 타입의 하나 이상의 소셜 그래프 엔티티에 대한 일치 정보(coincidence information); 또는 다른 적절한 행위들을 포함할 수 있다. 본 명세서는 특정 방식으로 친밀성을 측정하는 것을 기술하지만, 본 명세서는 임의의 적절한 방식으로 친밀성을 측정하는 것을 고려한다.

[0179] 특정 실시예로, 소셜 네트워킹 시스템(160)은 다양한 인자를 사용하여 계수를 계산할 수 있다. 이런 인자들은 예컨대 사용자 행위, 객체 사이의 관계 타입, 위치 정보, 다른 적절한 인자들 또는 이들의 임의의 조합을 포함할 수 있다. 특정 실시예로, 각기 다른 인자는 계수를 계산할 때 다르게 가중될 수 있다. 각 인자에 대한 가중치는 정적일 수 있거나, 가중치는 예컨대 사용자, 관계의 타입, 행위의 타입, 사용자의 위치 등에 따라 변할 수 있다. 인자들에 대한 평점(ratings)이 사용자에게 대한 총 계수를 결정하는데 그 가중치에 따라 결합될 수 있다. 예로서 제한 없이, 특정 사용자 행위는 평점과 가중치 모두를 할당받을 수 있는 한편, 특정 사용자 행위와 관련된 관계는 평점 및 상관형 가중치(correlating weight)(예컨대, 그래서 총 가중치는 100%)를 할당받는다. 특정 객체에 대한 사용자의 계수를 계산하기 위해, 사용자의 행위에 할당된 평점은 예컨대 총 계수의 60%를 포함할 수 있는 한편, 사용자와 객체 사이의 관계는 총 계수의 40%를 포함할 수 있다. 특정 실시예로, 소셜 네트워킹 시스템(160)은, 예컨대 정보에 접근한 이후의 시간, 감쇠 인자(decay factors), 접근 빈도, 정보에 대한 관계 또는 정보에 접근한 객체에 대한 관계, 객체와 연결된 소셜 그래프 엔티티에 대한 관계, 사용자 행위의 단기적 또는 장기적 평균, 사용자 피드백, 다른 적절한 변수들 또는 이들의 임의의 조합과 같이, 계수를 계산하는데 사용되는 다양한 인자들에 대한 가중치를 결정할 때 다양한 변수를 고려할 수 있다. 예로서 제한 없이, 더 최근의 행위가 계수를 계산할 때 더 관련되도록, 계수는 특정 행위가 제공한 신호의 세기가 시간에 따라 감쇠하게 하는 감쇠 인자를 포함할 수 있다. 평점 및 가중치는 계수가 기초하는 행위의 계속되는 추적에 기초하여 연속적으로 업데이트될 수 있다. 임의의 타입의 프로세스 또는 알고리즘은 각 인자에 대한 평점 및 인자에 할당된 가중치를 할당, 결합, 평균화 등을 하는데 이용될 수 있다. 특정 실시예로, 소셜 네트워킹 시스템(160)은 이력 행위 및 과거 사용자 응답 또는 다양한 옵션에 사용자를 노출시키고 응답을 측정하여 사용자로부터 파밍된(farmed) 데이터에 대하여 훈련되는 기계-학습 알고리즘을 사용하여 계수를 결정할 수 있다. 본 명세서는 특정 방식으로 계수를 계산하는 것을 기술하지만, 본 명세서는 임의의 적절한 방식으로 계수를 계산하는 것을 고려한다.

[0180] 특정 실시예로, 소셜 네트워킹 시스템(160)은 사용자의 행위에 기초하여 계수를 계산할 수 있다. 소셜 네트워킹 시스템(160)은 온라인 소셜 네트워크에서, 제3자 시스템(170)에서, 다른 적절한 시스템들에서 또는 이들의 임의의 조합에서 이런 행위들을 모니터링할 수 있다. 임의의 적절한 타입의 사용자 행위들이 추적되거나 모니터링될

수 있다. 일반적인 사용자 행위들은 프로필 페이지를 열람하기, 콘텐츠를 생성하거나 게시하기, 콘텐츠와 상호 작용하기, 태그하거나 이미지에 태그되기, 그룹에 가입하기, 이벤트에의 참여를 목록화하고 확인하기, 위치에 체크인하기, 특정 페이지를 좋아하기, 페이지를 생성하기 및 소셜 행위를 용이하게 하는 다른 업무들을 수행하기를 포함한다. 특정 실시예로, 소셜 네트워킹 시스템(160)은 특정 타입의 콘텐츠를 가진 사용자의 행위에 기초하여 계수를 계산할 수 있다. 콘텐츠는 온라인 소셜 네트워크, 제3자 시스템(170) 또는 다른 적절한 시스템과 관련될 수 있다. 콘텐츠는 사용자, 프로필 페이지, 게시물, 뉴스 소식, 헤드라인, 인스턴트 메시지, 채팅방 대화, 이메일, 광고, 사진, 비디오, 음악, 다른 적절한 객체들 또는 이들의 임의의 조합을 포함할 수 있다. 소셜 네트워킹 시스템(160)은 하나 이상의 행위들이 주제, 콘텐츠, 다른 사용자들 등에 대한 친밀성을 표시하는지를 결정하도록 사용자의 행위를 분석할 수 있다. 예로서 제한 없이, 사용자가 "커피" 또는 그 유사물에 관한 콘텐츠를 빈번히 게시할 수 있다면, 소셜 네트워킹 시스템(160)은 컨셉 "커피"에 대하여 사용자가 높은 계수를 가진다고 결정할 수 있다. 특정 행위 또는 행위 타입은 다른 행위들보다 더 높은 가중치 및/또는 평점을 할당받을 수 있으며, 이는 계산되는 총 계수에 영향을 줄 수 있다. 예로서 제한 없이, 제1 사용자가 제2 사용자에게 이메일 하는 경우, 그 행위에 대한 가중치 또는 평점은 제1 사용자가 제2 사용자에 대한 사용자 프로필 페이지를 단순히 열람한 경우보다 더 높을 수 있다.

[0181] 특정 실시예로, 소셜 네트워킹 시스템(160)은 특정 객체들 사이의 관계의 타입에 기초하여 계수를 계산할 수 있다. 소셜 그래프(200)를 참조하면, 소셜 네트워킹 시스템(160)은 계수를 계산할 때 특정 사용자 노드(202)와 컨셉 노드(204)를 연결하는 에지(206)의 수 및/또는 타입을 분석할 수 있다. 예로서 제한 없이, (2명의 사용자가 결혼했음을 표현하는) 배우자-타입 에지(spouse-type edge)로 연결되는 사용자 노드(202)는 친구-타입 에지로 연결되는 사용자 노드(202)보다 더 높은 계수를 할당받을 수 있다. 즉, 특정 사용자에 대한 행위 및 관계에 할당된 가중치에 따라, 총 친밀성은 사용자의 친구에 대한 콘텐츠보다 사용자의 배우자에 대한 콘텐츠에 대하여 더 높도록 결정될 수 있다. 특정 실시예로, 사용자가 또 다른 객체와 맺은 관계는 그 객체에 대한 계수를 계산할 때 사용자의 행위의 가중치 및/또는 평점에 영향을 줄 수 있다. 예로서 제한 없이, 사용자가 제1 사진에 태그되지만 단지 제2 사진만을 좋아한다면, 콘텐츠와 태그-타입(tagged-in-type) 관계를 가지는 것이 콘텐츠와 좋아요-타입 관계를 가지는 것보다 더 높은 가중치 및/또는 평점을 할당받을 수 있기 때문에, 소셜 네트워킹 시스템(160)은 사용자가 제2 사진보다 제1 사진에 대하여 더 높은 계수를 가진다고 결정할 수 있다. 특정 실시예로, 소셜 네트워킹 시스템(160)은 한 명 이상의 제2 사용자가 특정 객체와 맺은 관계에 기초하여 제1 사용자에 대한 계수를 계산할 수 있다. 즉, 다른 사용자들이 객체와 맺은 연결관계 및 계수는 그 객체에 대한 제1 사용자의 계수에 영향을 줄 수 있다. 예로서 제한 없이, 제1 사용자가 한 명 이상의 제2 사용자들과 연결되거나 그 제2 사용자들에 대한 높은 계수를 가지며, 이들 제2 사용자는 특정 객체와 연결되거나 그 특정 객체에 대한 높은 계수를 가진다면, 소셜 네트워킹 시스템(160)은 제1 사용자도 또한 그 특정 객체에 대해 상대적으로 높은 계수를 가져야 한다고 결정할 수 있다. 특정 실시예로, 계수는 특정 객체들 사이의 이격도를 기초로 할 수 있다. 더 낮은 계수는 제1 사용자가 소셜 그래프(200)에서 제1 사용자와 간접적으로 연결되는 사용자의 콘텐츠 객체에 대한 관심을 공유할 가능성이 감소함을 표현할 수 있다. 예로서 제한 없이, 소셜 그래프(200)에서 더 가까운(즉, 더 적은 이격도의) 소셜 그래프 엔티티들은 소셜 그래프(200)에서 더 멀리 떨어져 있는 엔티티들보다 더 높은 계수를 가질 수 있다.

[0182] 특정 실시예로, 소셜 네트워킹 시스템(160)은 위치 정보에 기초하여 계수를 계산할 수 있다. 서로 지리적으로 더 가까운 객체들은 더 먼 객체들보다 서로 더 관련이 있거나 더 관심이 있다고 간주될 수 있다. 특정 실시예로, 특정 객체에 대한 사용자의 계수는 사용자와 관련된 현재 위치(또는 사용자의 클라이언트 시스템(130)의 위치)에 대한 객체의 위치의 근접성을 기초로 할 수 있다. 제1 사용자는 제1 사용자와 더 가까운 다른 사용자들이나 컨셉들에 더 관심이 있을 수 있다. 예로서 제한 없이, 사용자가 공항으로부터는 1마일 그리고 주 유소로부터는 2마일 떨어져 있다면, 소셜 네트워킹 시스템(160)은 사용자와 공항의 근접성에 기초하여 사용자가 주유소보다 공항에 대해 더 높은 계수를 가진다고 결정할 수 있다.

[0183] 특정 실시예로, 소셜 네트워킹 시스템(160)은 계수 정보에 기초하여 사용자에게 특정 행위를 수행할 수 있다. 계수는 사용자가 특정 행위를 수행할 것인지 여부를 그 행위에 대한 사용자의 관심에 기초하여 예측하는데 사용될 수 있다. 계수는 가령 광고, 검색 결과, 뉴스 소식, 미디어, 메시지, 알림 또는 다른 적절한 객체와 같이 사용자에게 임의의 타입의 객체들을 생성하거나 제시할 때 사용될 수 있다. 또한, 계수는 적절한 경우 이런 객체들을 순위화 및 순서화하는데 이용될 수 있다. 이 방식으로, 소셜 네트워킹 시스템(160)은 사용자의 관심 및 현재 상황과 관련이 있는 정보를 제공할 수 있고, 사용자들이 관심을 가질 이런 정보를 찾을 가능성을 증가시킨다. 특정 실시예로, 소셜 네트워킹 시스템(160)은 계수 정보에 기초하여 콘텐츠를 생성할 수 있다. 콘텐츠 객체는 사용자에게 특화된 계수에 기초하여 제공되거나 선택될 수 있다. 예로서 제한 없이, 계수는 사용자에게

대한 미디어를 생성하는데 사용될 수 있는데, 이때 사용자는 사용자가 미디어 객체에 대해 높은 총 계수를 가지는 미디어를 제시받을 수 있다. 또 다른 예로서 제한 없이, 계수는 사용자에게 대한 광고를 생성하는데 사용될 수 있는데, 이때 사용자는 사용자가 광고된 객체에 대해 높은 총 계수를 가지는 광고를 제시받을 수 있다. 특정 실시예로, 소셜 네트워킹 시스템(160)은 계수 정보에 기초하여 검색 결과를 생성할 수 있다. 특정 사용자에게 대한 검색 결과는 질의중인 사용자에게 대하여 검색 결과와 관련된 계수에 기초하여 점수화되거나 순위화될 수 있다. 예로서 제한 없이, 더 높은 계수를 가진 객체에 해당하는 검색 결과는 더 낮은 계수를 가진 객체에 해당하는 결과보다 검색 결과 페이지에서 더 높게 순위화될 수 있다.

[0184] 특정 실시예로, 소셜 네트워킹 시스템(160)은 특정 시스템이나 프로세스로부터의 계수에 대한 요청에 응답하여 계수를 계산할 수 있다. 사용자가 소정의 상황에서 행할 수 있는(주제일 수 있는) 가능한 행위를 예측하기 위해, 임의의 프로세스는 사용자에게 대해 계산된 계수를 요청할 수 있다. 또한, 요청은 계수를 계산하는데 사용되는 다양한 인자들에 대해 사용하도록 한 세트의 가중치를 포함할 수 있다. 이 요청은 온라인 소셜 네트워크에서 실행하는 프로세스로부터, (예컨대, API 또는 다른 통신 채널을 통해) 제3자 시스템(170)으로부터 또는 다른 적절한 시스템으로부터 비롯될 수 있다. 그 요청에 응답하여, 소셜 네트워킹 시스템(160)은 계수를 계산(또는 이미 계산되고 저장되었다면 계수 정보에 접근)할 수 있다. 특정 실시예로, 소셜 네트워킹 시스템(160)은 특정 프로세스에 대한 친밀성을 측정할 수 있다. (온라인 소셜 네트워크 내부와 외부 모두의) 다른 프로세스들은 특정 객체 또는 객체들의 세트에 대한 계수를 요청할 수 있다. 소셜 네트워킹 시스템(160)은 친밀도를 요청한 특정 프로세스와 관련이 있는 친밀도를 제공할 수 있다. 이 방식으로, 각 프로세스는 프로세스가 친밀도를 사용하는 다른 컨텍스트에 대하여 맞춤화되는 친밀도를 수신한다.

[0185] 소셜 그래프 친밀성 및 친밀성 계수와 관련하여, 특정 실시예들은 참조로 각각 통합되는, 2006년 8월 11일자로 출원된 미국특허출원 제11/503093호, 2010년 12월 22일자로 출원된 미국특허출원 제12/977027호, 2010년 12월 23일자로 출원된 미국특허출원 제12/978265호 및 2012년 10월 1일자로 출원된 미국특허출원 제13/632869호에 개시되는 하나 이상의 시스템, 컴포넌트, 구성요소, 기능, 방법, 동작 또는 단계를 이용할 수 있다.

[0186] 광고

[0187] 특정 실시예로, 광고는, 하나 이상의 웹 페이지에서 제시되는 임의의 적절한 디지털 포맷에서, 하나 이상의 이메일에서 또는 사용자에게 의해 요청되는 검색 결과와 관련하여, (HTML-링크될 수 있는) 텍스트, (HTML-링크될 수 있는) 하나 이상의 이미지, 하나 이상의 비디오, 오디오, 하나 이상의 ADOBE FLASH 파일, 이들의 적절한 조합 또는 임의의 다른 적절한 광고일 수 있다. 추가로 또는 대안으로, 광고는 하나 이상의 스폰서 소식(예컨대, 소셜 네트워킹 시스템(160)의 뉴스-피드 또는 티커 아이템)일 수 있다. 스폰서 소식은 예컨대 소셜 행위를 사용자의 프로필 페이지나 다른 페이지의 기결정된 영역 내에 제시되게 하거나, 광고자와 관련된 추가 정보와 함께 제시되게 하거나, 다른 사용자들의 뉴스피드나 티커 내에 올리거나 하이라이팅되게 하거나, 홍보되게 함으로써, 광고자가 홍보하는 사용자에게 의한 소셜 행위(가령, 페이지를 "좋아요" 하기, 페이지의 게시물을 "좋아요" 하거나 코멘트하기, 페이지와 관련된 이벤트에 대해 회답(RSVP)하기, 페이지에 게시된 질문에 투표하기, 한 장소로 체크인하기, 애플리케이션을 사용하거나 게임을 실행하기 또는 웹사이트를 "좋아요" 하거나 공유하기)일 수 있다. 광고자는 소셜 행위가 홍보되도록 지불할 수 있다. 예로서 제한 없이, 광고는 검색-결과 페이지의 검색 결과들 중에 포함될 수 있으며, 이때 스폰서 콘텐츠는 비-스폰서 콘텐츠보다 더 홍보된다. 특정 실시예로, 하나 이상의 광고는 상술한 네트워크-인지 제품 시판 기술을 사용하여 온라인 소셜 네트워크의 사용자의 하나 이상의 클러스터에 대해 테스트될 수 있다.

[0188] 특정 실시예로, 광고는 소셜 네트워킹 시스템 웹페이지, 제3자 웹페이지 또는 다른 페이지 내에서 디스플레이하기 위해 요청될 수 있다. 광고는, 가령 페이지 상단의 배너 영역에서, 페이지 측면의 종렬에서, 페이지의 GUI에서, 팝-업창에서, 드롭-다운 메뉴에서, 페이지의 입력 필드에서, 페이지의 콘텐츠의 상단 위에서 또는 그 밖의 페이지의 어느 곳에서, 페이지의 전용 부분에 디스플레이될 수 있다. 추가로 또는 대안으로, 광고는 애플리케이션 내에 디스플레이될 수 있다. 광고는 사용자가 페이지에 접근하거나, 애플리케이션을 이용할 수 있기 전에 광고를 상호작용하거나 광고를 시청하도록 사용자에게 요구하는 전용 페이지 내에 디스플레이될 수 있다. 예컨대, 사용자는 웹 브라우저를 통해 광고를 열람할 수 있다.

[0189] 사용자는 임의의 적절한 방식으로 광고와 상호작용할 수 있다. 사용자는 광고를 클릭하거나 선택할 수 있다. 광고를 선택함으로써, 사용자는 광고와 관련된 페이지로 (또는, 사용자가 사용하는 브라우저 또는 다른 애플리케이션)를 인도될 수 있다. 광고와 관련된 페이지에서, 사용자는 가령 광고와 관련된 제품이나 서비스를 구매하기, 광고와 관련된 정보를 수신하기 또는 광고와 관련된 뉴스레터를 구독하기와 같은 추가의 행위들을 행

할 수 있다. 오디오 또는 비디오를 가진 광고는 광고의 컴포넌트를 선택("실행 버튼"을 좋아요)함으로써 실행될 수 있다. 대안으로, 광고를 선택함으로써, 소셜 네트워킹 시스템(160)은 사용자의 특정 행위를 실행하거나 수정할 수 있다.

[0190] 또한, 광고는 사용자가 상호작용할 수 있는 소셜 네트워킹 시스템 기능을 포함할 수 있다. 예로서 제한 없이, 광고는 보증과 관련된 아이콘이나 링크를 선택하여 사용자가 광고를 "좋아요"하거나 보증하게 할 수 있다. 또 다른 예로서 제한 없이, 광고는 (예컨대, 질의를 실행하여) 사용자가 광고자에 관한 콘텐츠를 검색하게 할 수 있다. 마찬가지로, 사용자는 (예컨대, 소셜 네트워킹 시스템(160)을 통해) 다른 사용자와 광고를 공유할 수 있거나, (예컨대, 소셜 네트워킹 시스템(160)을 통해) 광고와 관련된 이벤트에 회답(RSVP)할 수 있다. 추가로 또는 대안으로, 광고는 사용자에게 관한 소셜 네트워킹 시스템 콘텐츠를 포함할 수 있다. 예로서 제한 없이, 광고는 광고의 내용과 관련된 행위를 행했던 소셜 네트워킹 시스템(160) 내 사용자의 친구에 대한 정보를 디스플레이할 수 있다.

[0191] 시스템 및 방법

[0192] 도 7은 예시적인 컴퓨터 시스템(700)을 도시한다. 특정 실시예로, 하나 이상의 컴퓨터 시스템들(700)은 본 명세서에 기술되거나 도시된 하나 이상의 방법들의 하나 이상의 단계들을 수행한다. 특정 실시예로, 하나 이상의 컴퓨터 시스템들(700)은 본 명세서에 기술되거나 도시된 기능을 제공한다. 특정 실시예로, 하나 이상의 컴퓨터 시스템들(700)을 실행하는 소프트웨어는 본 명세서에 기술되거나 도시된 하나 이상의 방법들의 하나 이상의 단계들을 수행하거나, 본 명세서에 기술되거나 도시된 기능을 제공한다. 특정 실시예는 하나 이상의 컴퓨터 시스템들(700)의 하나 이상의 부분들을 포함한다. 본 명세서에서, 적절한 경우 컴퓨터 시스템에 대한 언급은 컴퓨팅 장치를 포함할 수 있고 그 역도 또한 동일하다. 게다가, 컴퓨터 시스템에 대한 언급은 적절한 경우 하나 이상의 컴퓨터 시스템을 포함할 수 있다.

[0193] 본 명세서는 임의의 적절한 수의 컴퓨터 시스템(700)을 고려한다. 본 명세서는 임의의 적절한 물리적 형태를 취하는 컴퓨터 시스템(700)을 고려한다. 예로서 제한 없이, 컴퓨터 시스템(700)은 임베디드 컴퓨터 시스템, 시스템-온-칩(SOC), 단일-보드 컴퓨터 시스템(SBC)(예컨대, 컴퓨터-온-모듈(COM) 또는 시스템-온-모듈(SOM)), 데스크톱 컴퓨터 시스템, 랩톱 또는 노트북 컴퓨터 시스템, 상호작용형 키오스크(kiosk), 메인 프레임, 컴퓨터 시스템 메쉬(mesh), 모바일 전화, 개인 정보 단말기(PDA), 서버, 태블릿 컴퓨터 시스템 또는 이들의 2 이상의 조합일 수 있다. 적절한 경우, 컴퓨터 시스템(700)은 하나 이상의 컴퓨터 시스템(700)들을 포함할 수 있거나; 일체형 또는 분산형일 수 있거나; 다수의 위치에 걸쳐 있거나, 다수의 기계에 걸쳐 있거나; 다수의 데이터 센터에 걸쳐 있거나; 하나 이상의 네트워크에 하나 이상의 클라우드 성분을 포함할 수 있는 클라우드에 상주할 수 있다. 적절한 경우, 하나 이상의 컴퓨터 시스템(700)은 본 명세서에 기술되거나 도시되는 하나 이상의 방법의 하나 이상의 단계를 실질적으로 공간적 또는 시간적 제한 없이 실행할 수 있다. 예로서 제한 없이, 하나 이상의 컴퓨터 시스템(700)은 본 명세서에 기술되거나 도시되는 하나 이상의 방법의 하나 이상의 단계를 실시간으로 또는 일괄 모드로 실행할 수 있다. 적절한 경우, 하나 이상의 컴퓨터 시스템(700)은 본 명세서에 기술되거나 도시되는 하나 이상의 방법의 하나 이상의 단계를 다른 시기에 또는 다른 위치에서 실행할 수 있다.

[0194] 특정 실시예로, 컴퓨터 시스템(700)은 프로세서(702), 메모리(704), 저장소(706), 입력/출력(I/O) 인터페이스(708), 통신 인터페이스(710) 및 버스(712)를 포함한다. 본 명세서가 특정 배열로 특정한 수의 특정 구성요소를 갖는 특정 컴퓨터 시스템을 기술하고 도시하지만, 본 명세서는 임의의 적절한 배열로 임의의 적절한 수의 임의의 적절한 구성요소를 갖는 임의의 적절한 컴퓨터 시스템을 고려한다.

[0195] 특정 실시예로, 프로세서(702)는 가령 컴퓨터 프로그램을 구성하는 명령어와 같은 명령어를 실행하기 위한 하드웨어를 포함한다. 예로서 제한 없이, 명령어를 실행하기 위해, 프로세서(702)는 내부 레지스터, 내부 캐시, 메모리(704) 또는 저장소(706)로부터 명령어를 검색(또는 페치(fetch))할 수 있고; 명령어를 디코딩하고 실행한 후; 하나 이상의 결과를 내부 레지스터, 내부 캐시, 메모리(704) 또는 저장소(706)에 기록할 수 있다. 특정 실시예로, 프로세서(702)는 데이터용, 명령어용 또는 주소용 하나 이상의 내부 캐시를 포함할 수 있다. 본 명세서는 적절한 경우 임의의 적절한 수의 임의의 적절한 내부 캐시들을 포함하는 프로세서(702)를 고려한다. 예로서 제한 없이, 프로세서(702)는 하나 이상의 명령어 캐시들, 하나 이상의 데이터 캐시들 및 하나 이상의 변환 색인 버퍼(translation lookaside buffers, TLBs)를 포함할 수 있다. 명령어 캐시에 저장된 명령어들은 메모리(704)나 저장소(706) 내 명령어들의 사본일 수 있고, 명령어 캐시는 프로세서(702)에 의한 이런 명령어들의 검색 속도를 높일 수 있다. 데이터 캐시 내의 데이터는 프로세서(702)에서 실행하는 다음 명령들에 의해 접근하거나 메모리(704)나 저장소(706)로 기록하기 위해 프로세서(702)에서 실행되는 이전 명령들의 결과; 또는 다른 적절

한 데이터를 동작하는데 프로세서(702)에서 실행하는 명령어를 위한 메모리(704)나 저장소(706) 내의 데이터의 사본일 수 있다. 데이터 캐시는 프로세서(702)에 의한 관독 또는 기록 동작의 속도를 높일 수 있다. TLB들은 프로세서(702)에 의한 가상 주소 변환의 속도를 높일 수 있다. 특정 실시예로, 프로세서(702)는 데이터용, 명령어용 또는 주소용 하나 이상의 내부 레지스터를 포함할 수 있다. 본 명세서는 적절한 경우 임의의 적절한 수의 임의의 적절한 내부 레지스터들을 포함하는 프로세서(702)를 고려한다. 적절한 경우, 프로세서(702)는 하나 이상의 산술 논리 유닛(ALUs)을 포함할 수 있거나; 멀티-코어 프로세서일 수 있거나; 하나 이상이 프로세서들(702)을 포함할 수 있다. 본 명세서가 특정 프로세서를 기술하고 도시하지만, 본 명세서는 임의의 적절한 프로세서를 고려한다.

[0196] 특정 실시예로, 메모리(704)는 프로세서(702)가 실행하는 명령어 또는 프로세서(702)가 운영하는 데이터를 저장하기 위한 메인 메모리를 포함한다. 예로서 제한 없이, 컴퓨터 시스템(700)은 저장소(706)나 또 다른 소스(가령, 예컨대 또 다른 컴퓨터 시스템(700))에서 메모리(704)로 명령어를 로딩할 수 있다. 이후, 프로세서(702)는 메모리(704)에서 내부 레지스터나 내부 캐시로 명령어를 로딩할 수 있다. 명령어를 실행하기 위해, 프로세서(702)는 내부 레지스터나 내부 캐시로부터 명령어를 검색하고 이들을 디코딩할 수 있다. 명령어의 실행 중 또는 실행 후, 프로세서(702)는 (중간 결과 또는 최종 결과일 수 있는) 하나 이상의 결과를 내부 레지스터나 내부 캐시로 기록할 수 있다. 이후, 프로세서(702)는 하나 이상의 이런 결과를 메모리(704)에 기록할 수 있다. 특정 실시예로, 프로세서(702)는 (저장소(706) 또는 다른 곳과는 대조적으로) 하나 이상의 내부 레지스터나 내부 캐시에서 또는 메모리(704)에서 단지 명령어만을 실행하며, (저장소(706) 또는 다른 곳과는 대조적으로) 하나 이상의 내부 레지스터나 내부 캐시에서 또는 메모리(704)에서 단지 데이터만을 운영한다. (주소 버스 및 데이터 버스를 각각 포함할 수 있는) 하나 이상의 메모리 버스는 프로세서(702)를 메모리(704)로 연결할 수 있다. 하기에 기술되는 바와 같이, 버스(712)는 하나 이상의 메모리 버스를 포함할 수 있다. 특정 실시예로, 하나 이상의 메모리 관리 유닛(MMUs)은 프로세서(702)와 메모리(704) 사이에 상주하며, 프로세서(702)에 의해 요청되는 메모리(704)로의 접근을 용이하게 한다. 특정 실시예로, 메모리(704)는 랜덤 액세스 메모리(RAM)를 포함한다. 적절한 경우, 이런 RAM은 휘발성 메모리일 수 있다. 적절한 경우, 이런 RAM은 동적 RAM(DRAM) 또는 정적 RAM(SRAM)일 수 있다. 게다가, 적절한 경우, 이런 RAM은 단일 포트형 또는 다중-포트형 RAM일 수 있다. 본 명세서는 임의의 적절한 RAM을 고려한다. 적절한 경우, 메모리(704)는 하나 이상의 메모리(704)를 포함할 수 있다. 본 명세서가 특정 메모리를 기술하고 도시하지만, 본 명세서는 임의의 적절한 메모리를 고려한다.

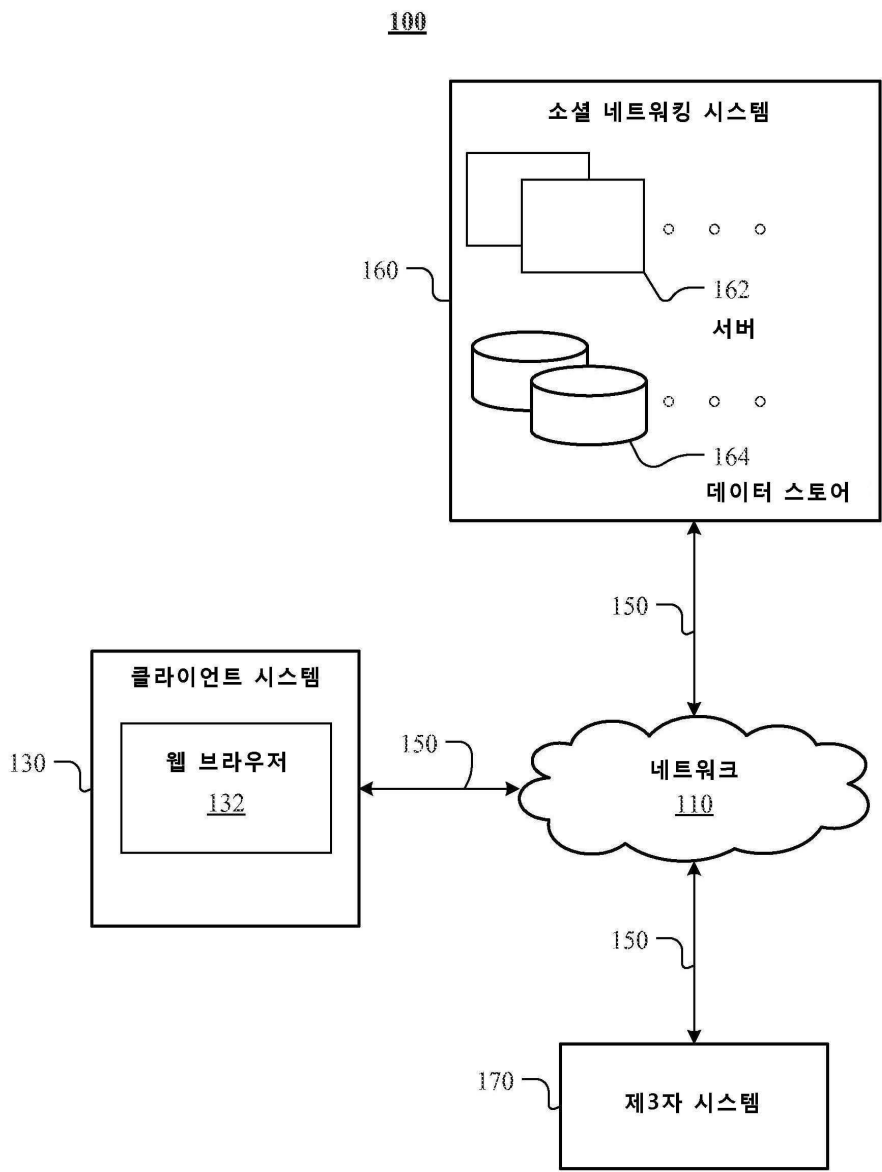
[0197] 특정 실시예로, 저장소(706)는 데이터용 또는 명령어용 대용량 저장소를 포함한다. 예로서 제한 없이, 저장소(706)는 하드 디스크 드라이브(HDD), 플로피 디스크 드라이브, 플래시 메모리, 광디스크, 자기-광학 디스크, 자기 테이프, 범용 직렬 버스(USB) 드라이브 또는 이들의 2 이상의 조합을 포함할 수 있다. 적절한 경우, 저장소(706)는 착탈식 또는 비-착탈식(또는 고정) 매체를 포함할 수 있다. 적절한 경우, 저장소(706)는 컴퓨터 시스템(700)의 내부 또는 외부에 있을 수 있다. 특정 실시예로, 저장소(706)는 비휘발성, 고체-상태(solid-state) 메모리이다. 특정 실시예로, 저장소(706)는 읽기 전용 메모리(ROM)를 포함한다. 적절한 경우, 이런 ROM은 마스크-프로그래밍된 ROM, 프로그램가능 ROM(PROM), 소거가능 PROM(EPROM), 전기적 소거가능 PROM(EEPROM), 전기적 변경가능 ROM(EROM), 플래시 메모리 또는 이들의 2 이상의 조합일 수 있다. 본 명세서는 임의의 적절한 물리적 형태를 취하는 대용량 저장소(706)를 고려한다. 적절한 경우, 저장소(706)는 프로세서(702)와 저장소(706) 사이의 통신을 용이하게 하는 하나 이상의 저장소 제어 유닛을 포함할 수 있다. 적절한 경우, 저장소(706)는 하나 이상의 저장소(706)를 포함할 수 있다. 본 명세서가 특정 저장소를 기술하고 도시하지만, 본 명세서는 임의의 적절한 저장소를 고려한다.

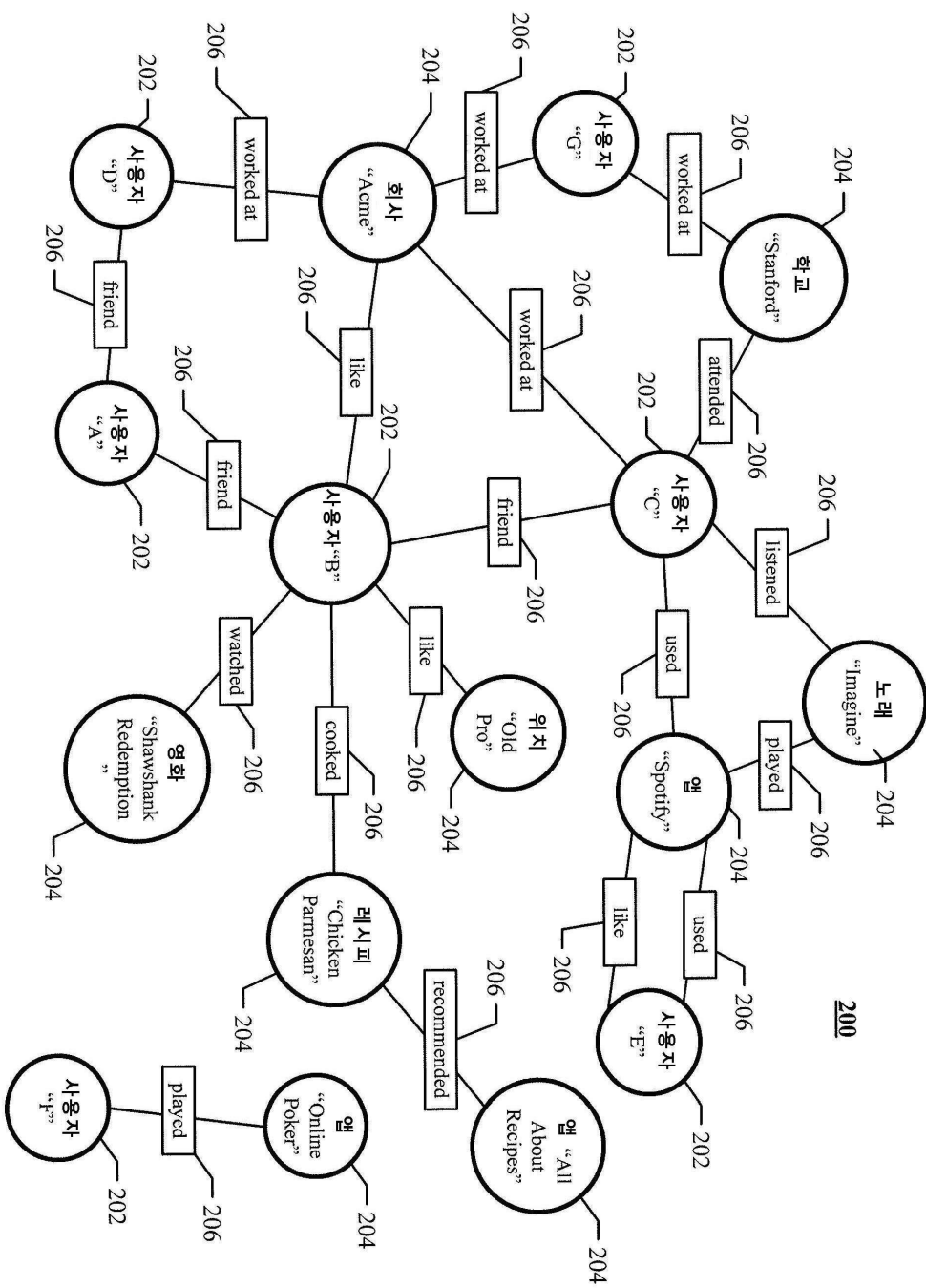
[0198] 특정 실시예로, I/O 인터페이스(708)는 컴퓨터 시스템(700)과 하나 이상의 I/O 장치 사이의 통신을 위한 하나 이상의 인터페이스를 제공하는 하드웨어, 소프트웨어 또는 이들 모두를 포함한다. 적절한 경우, 컴퓨터 시스템(700)은 하나 이상의 이들 I/O 장치를 포함할 수 있다. 하나 이상의 이들 I/O 장치는 사람과 컴퓨터 시스템(700) 사이의 통신을 가능하게 할 수 있다. 예로서 제한 없이, I/O 장치는 키보드, 키패드, 마이크로폰, 모니터, 마우스, 프린터, 스캐너, 스피커, 스틸 카메라(still camera), 스타일러스(stylus), 태블릿, 터치 스크린, 트랙볼(trackball), 비디오 카메라, 다른 적절한 I/O 장치 또는 이들의 2 이상의 조합을 포함할 수 있다. I/O 장치는 하나 이상의 센서를 포함할 수 있다. 본 명세서는 임의의 적절한 I/O 장치 및 이에 대한 적절한 I/O 인터페이스(708)를 고려한다. 적절한 경우, I/O 인터페이스(708)는 프로세서(702)가 하나 이상의 이들 I/O 장치를 구동할 수 있도록 하는 하나 이상의 장치 또는 소프트웨어 드라이버를 포함할 수 있다. 적절한 경우, I/O 인터페이스(708)는 하나 이상의 I/O 인터페이스(708)를 포함할 수 있다. 본 명세서가 특정 I/O 인터페이스를 기술하고 도시하지만, 본 명세서는 임의의 적절한 I/O 인터페이스를 고려한다.

- [0199] 특정 실시예로, 통신 인터페이스(710)는 컴퓨터 시스템(700)과 하나 이상의 다른 컴퓨터 시스템(700)이나 하나 이상의 네트워크 사이의 통신(가령, 예컨대 패킷-기반 통신)을 위한 하나 이상의 인터페이스를 제공하는 하드웨어, 소프트웨어 또는 이들 모두를 포함한다. 예로서 제한 없이, 통신 인터페이스(710)는 이더넷이나 다른 유선-기반 네트워크로 통신하기 위한 네트워크 인터페이스 제어장치(NIC)나 네트워크 어댑터 또는 가령 WI-FI 네트워크와 같이 무선 네트워크로 통신하기 위한 무선 NIC(WNIC)나 무선 어댑터를 포함할 수 있다. 본 명세서에서는 임의의 적절한 네트워크 및 이에 대한 임의의 적절한 통신 인터페이스(710)를 고려한다. 예로서 제한 없이, 컴퓨터 시스템(700)은 애드 혹 네트워크(ad hoc network), 개인 영역 네트워크(PAN), 근거리 네트워크(LAN), 광역 네트워크(WAN), 대도시 네트워크(MAN), 인터넷의 하나 이상의 부분 또는 2 이상의 이런 네트워크들의 조합으로 통신할 수 있다. 하나 이상의 이런 네트워크의 하나 이상의 부분은 유선 또는 무선일 수 있다. 예로서, 컴퓨터 시스템(700)은 무선 PAN(WPAN)(가령, 예컨대 BLUETOOTH WPAN), WI-FI 네트워크, WI-MAX 네트워크, 셀룰러 전화 네트워크(가령, 예컨대 GSM(Global System for Mobile Communication) 네트워크), 다른 적절한 무선 네트워크 또는 2 이상의 이런 네트워크들의 조합으로 통신할 수 있다. 적절한 경우, 컴퓨터 시스템(700)은 임의의 이들 네트워크에 대한 임의의 적절한 통신 인터페이스(710)를 포함할 수 있다. 적절한 경우, 통신 인터페이스(710)는 하나 이상의 통신 인터페이스(710)를 포함할 수 있다. 본 명세서가 특정 통신 인터페이스를 기술하고 도시하지만, 본 명세서는 임의의 적절한 통신 인터페이스를 고려한다.
- [0200] 특정 실시예로, 버스(712)는 컴퓨터 시스템(700)의 구성요소를 서로 연결하는 하드웨어, 소프트웨어 또는 이들 모두를 포함한다. 예로서 제한 없이, 버스(712)는 AGP(Accelerated Graphics Port)이나 다른 그래픽 버스, EISA(Enhanced Industry Standard Architecture) 버스, FSB(front-side bus), HT(HYPERTRANSPORT) 인터커넥트, ISA(Industry Standard Architecture) 버스, INFINIBAND 인터커넥트, LPC(low-pin-count) 버스, 메모리 버스, MCA(Micro Channel Architecture) 버스, PCI(Peripheral Component Interconnect) 버스, PCIe(PCI-Express) 버스, SATA(serial advanced technology attachment) 버스, VLB(Video Electronics Standard Association local) 버스, 다른 적절한 버스 또는 2 이상의 이런 버스의 조합을 포함할 수 있다. 적절한 경우, 버스(712)는 하나 이상의 버스(712)를 포함할 수 있다. 본 명세서가 특정 버스를 기술하고 도시하지만, 본 명세서는 임의의 적절한 버스나 인터커넥트를 고려한다.
- [0201] 본 명세서에서, 컴퓨터-판독가능한 비-일시적 저장매체 또는 저장매체들은 하나 이상의 반도체 기반 또는 다른 집적회로(ICs)(가령, 예컨대 FPGAs(field-programmable gate arrays) 또는 ASICs(application-specific ICs)), 하드 디스크 드라이브(HDDs), 하이브리드 하드 디스크(HHDs), 광학 디스크, 광학 디스크 드라이브(ODDs), 자기-광학 디스크, 자기-광학 드라이브, 플로피 디스크, 플로피 디스크 드라이브(FDDs), 자기 테이프, 고체-상태 드라이브(SSDs), RAM-드라이브, SECURE DIGITAL 카드나 드라이브, 임의의 다른 적절한 컴퓨터-판독가능한 비-일시적 저장매체 또는, 적절한 경우, 2 이상의 이들의 임의의 적절한 조합을 포함할 수 있다. 적절한 경우, 컴퓨터-판독가능한 비-일시적 저장매체는 휘발성, 비휘발성 또는 휘발성과 비휘발성의 조합일 수 있다.
- [0202] 본 명세서에서, "또는"은 명시적으로 다르게 지시하거나 문맥상 달리 지시되지 않는 한, 포괄적인 것이며 배타적인 것이 아니다. 따라서, 본 명세서에서 "A 또는 B"는 명시적으로 다르게 지시하거나 문맥상 달리 지시되지 않는 한, "A, B 또는 둘 모두"를 의미한다. 게다가, "및"은 명시적으로 다르게 지시하거나 문맥상 달리 지시되지 않는 한, 공동 및 별개 모두이다. 따라서, 본 명세서에서 "A 및 B"는 명시적으로 다르게 지시하거나 문맥상 달리 지시되지 않는 한, "A 및 B가 공동이든 별개이든 상관없이 모두"를 의미한다.
- [0203] 본 명세서의 범위는 당업자가 이해할 수 있는 본 명세서에 기술되거나 도시된 예시적인 실시예들에 대한 모든 변화, 치환, 변형, 대체 및 변경을 포함한다. 본 명세서의 범위는 본 명세서에 기술되거나 도시된 예시적인 실시예들로 국한되지 않는다. 게다가, 본 명세서는 특정 컴포넌트, 구성요소, 기능, 동작 또는 단계를 포함하는 것으로 본 명세서의 각각의 실시예들을 기술하고 도시하지만, 임의의 이런 실시예들은 당업자가 이해할 수 있는 본 명세서에 어디든 기술되거나 도시되는 임의의 컴포넌트, 구성요소, 기능, 동작 또는 단계의 임의의 조합이나 치환을 포함할 수 있다. 게다가, 첨부된 청구범위에서 특정 기능을 수행하도록 설계되거나, 배치되거나, 할 수 있거나, 구성되거나, 할 수 있게 하거나, 동작할 수 있거나, 동작하는 장치나 시스템 또는 장치나 시스템의 구성요소에 대한 언급은 장치, 시스템 또는 구성요소가 그렇게 설계되거나, 배치되거나, 할 수 있거나, 구성되거나, 가능하거나, 동작할 수 있거나 동작하는 한, 장치, 시스템, 구성요소, 그 또는 그러한 특정 기능이 활성화되었는지, 턴온 되었는지, 잠금 해제되었는지 여부를 포함한다.

도면

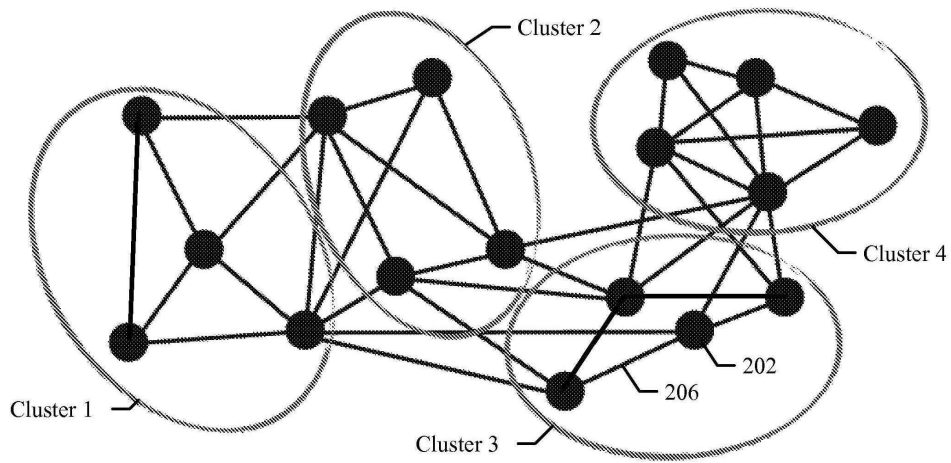
도면1



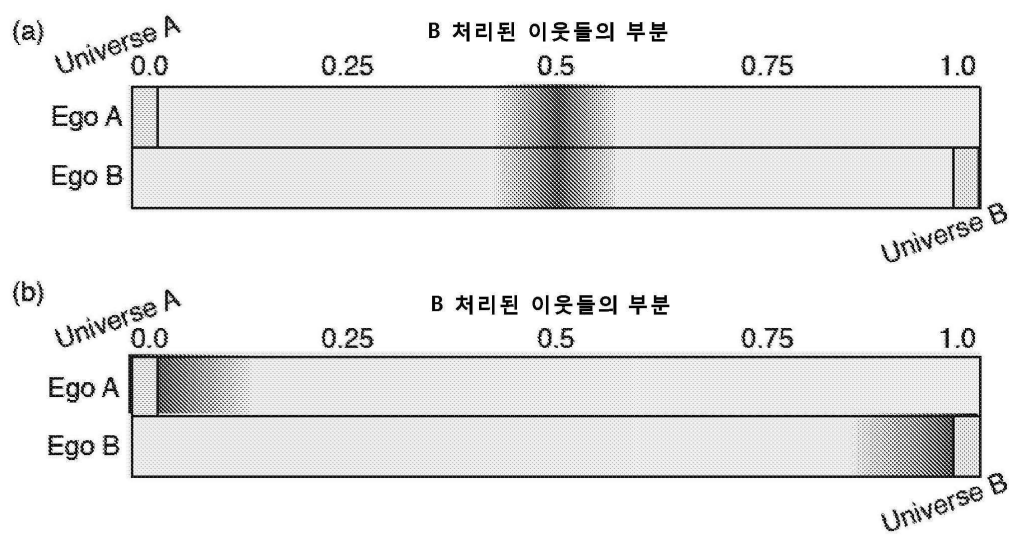


도면2

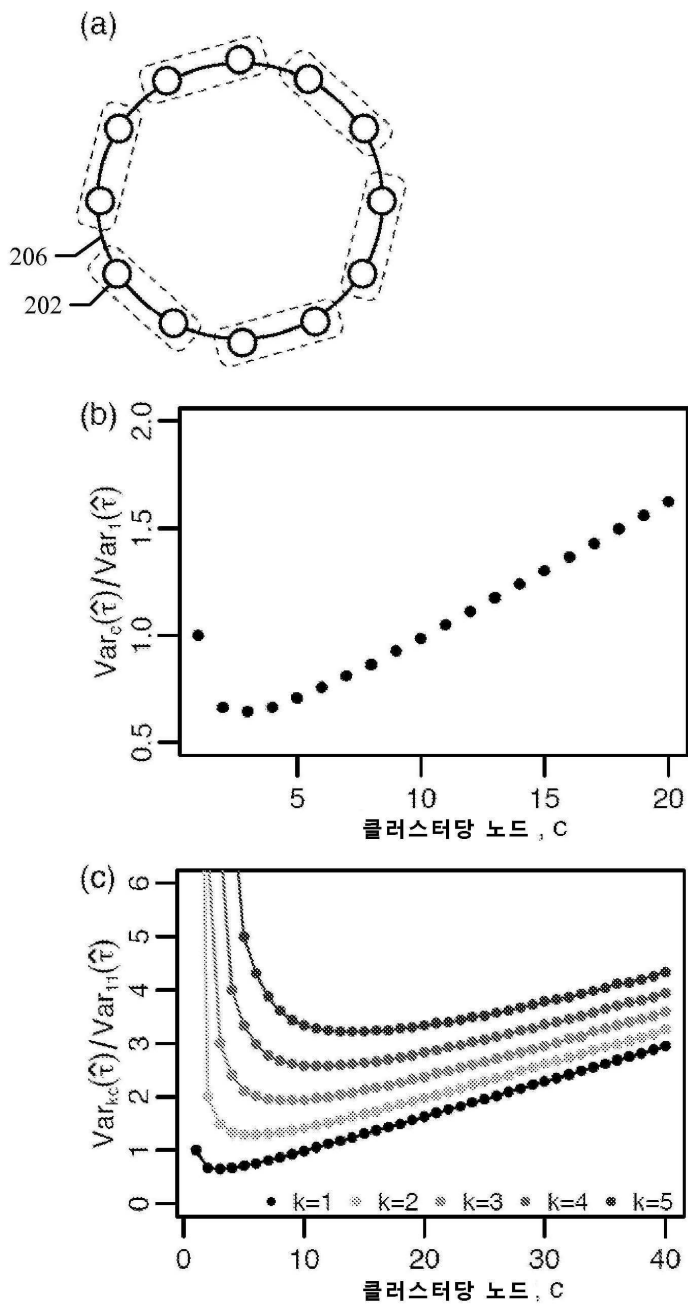
도면3



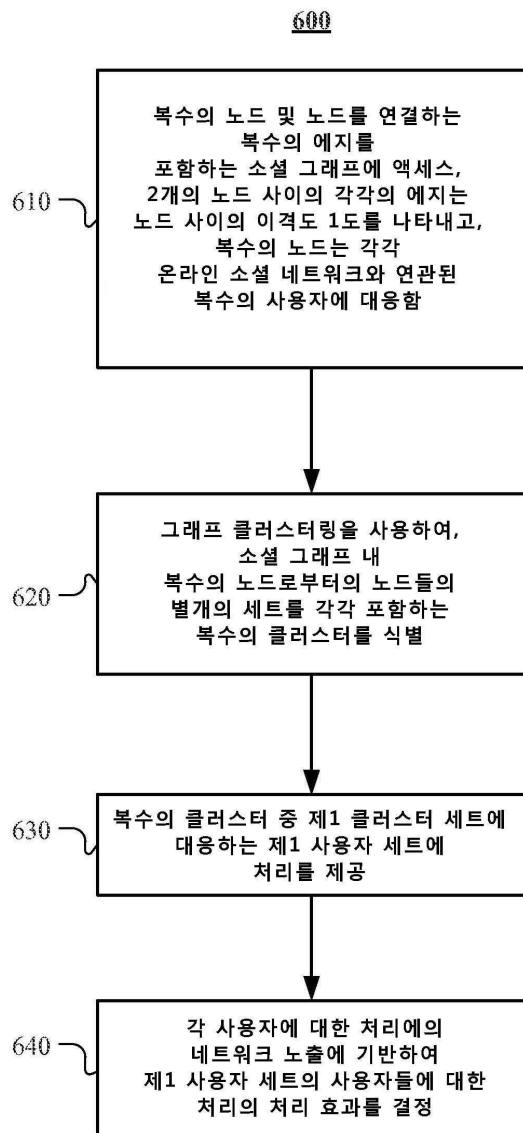
도면4



도면5



도면6



도면7

