■ (네이버 후원) 2017년도 학생논문 경진대회 수상작

소셜 네트워크에서 효율적인 영향력 최대화 방안 (Fast Influence Maximization in Social Networks)

고 윤 용 [†] 조 경 재 ^{††} 김 상 옥 ^{†††}
(Yun-Yong Ko) (Kyung-Jae Cho) (Sang-Wook Kim)

요 약 영향력 최대화란 소셜 네트워크에서 최대의 영향력을 갖는 k 개의 시드(seed) 노드로 이루어진 집합을 선출하는 문제이다. 이 문제를 해결한 기존 방법들이 갖는 가장 큰 문제는 시드 집합을 선출하는 데 너무 많은 시간이 소요된다는 점이다. 이러한 성능 문제는 미시적, 거시적 두 가지 측면에서 발생한다. 본 논문은 미시적, 거시적 측면의 성능 문제 동시에 해결하는 효율적인 영향력 최대화 방안을 제안한다. 또한, 양질의 시드 집합을 선출하기 위한 새로운 경로 기반 커뮤니티 탐지 기법을 제안한다. 네 가지 실제 계 데이터를 이용한 실험을 통해, 제안하는 방안이 미시적, 거시적 측면의 문제를 모두 해결하는 동시에 양질의 시드 집합을 선출함을 확인하였다.

키워드: 소셜 네트워크, 정보 파급, 영향력 최대화, 커뮤니티 탐지 기법

Abstract Influence maximization (IM) is the problem of finding a seed set composed of k nodes that maximizes the influence spread in social networks. However, one of the biggest problems of existing solutions for IM is that it takes too much time to select a k -seed set. This performance issue occurs at the micro and macro levels. In this paper, we propose a fast hybrid method that addresses two issues at micro and macro levels. Furthermore, we propose a path-based community detection method that helps to select a good seed set. The results of our experiment with four real-world datasets show that the proposed method resolves the two issues at the micro and macro levels and selects a good k -seed set.

Keywords: social network, information diffusion, influence maximization, community detection

· 본 연구는 과학기술정보통신부 및 정보통신기술진흥센터의 대학ICT연구센터 육성지원사업의 연구결과로 수행되었음(IITP-2017-2013-0-00881). 이 성과는 2017년도 정부(미래창조과학부)의 재원으로 한국연구재단의 지원을 받아 수행된 연구임(No. NRF-2017R1A2B3004581)

[†] 학생회원 : 한양대학교 컴퓨터 소프트웨어학과
yyko@agape.hanyang.ac.kr

^{††} 비 회 원 : 한양대학교 컴퓨터 소프트웨어학과
kcho035@agape.hanyang.ac.kr

^{†††} 종신회원 : 한양대학교 컴퓨터 소프트웨어학과 교수
(Hanyang Univ.)
wook@hanyang.ac.kr
(Corresponding author임)

논문접수 : 2017년 5월 30일

(Received 30 May 2017)

논문수정 : 2017년 8월 21일

(Revised 21 August 2017)

심사완료 : 2017년 8월 22일

(Accepted 22 August 2017)

Copyright©2017 한국정보과학회 : 개인 목적이나 교육 목적인 경우, 이 저작물의 전체 또는 일부에 대한 복사본 혹은 디지털 사본의 제작을 허가합니다. 이 때, 사본은 상업적 수단으로 사용할 수 없으며 첫 페이지에 본 문구와 출처를 반드시 명시해야 합니다. 이 외의 목적으로 복제, 배포, 출판, 전송 등 모든 유형의 사용행위를 하는 경우에 대하여는 사전에 허가를 얻고 비용을 지불해야 합니다.
정보과학회논문지 제44권 제10호(2017. 10)

1. 서론

Facebook, Twitter와 같은 소셜 네트워크 서비스(SNS)의 급격한 발전으로 이러한 서비스를 이용하는 유저들이 크게 증가하였다. 이때, SNS는 유저들이 다양한 의견, 지식 등을 서로 공유하는 매개체로써 역할을 한다. 기업들은 이러한 SNS들을 통해 자기 기업에 이윤을 가져다 줄 수 있는 유용한 정보를 소셜 네트워크에 최대한 널리 퍼뜨리고자 한다. 이를 입소문 마케팅(*viral marketing*) 전략이라 하고, 이를 달성하기 위해, 소셜 네트워크에서 “입소문 효과”(word-of-mouth effect)를 최대화하는 영향력 있는 유저들을 찾는 연구들이 진행되어 오고 있다.

이 문제를 영향력 최대화(*influence maximization*) 문제라 하고, 소셜 네트워크에서 최대의 영향력을 갖는 k 개의 시드(*seed*) 노드로 이루어진 집합을 선출 하는 것으로 정의된다. Kemp et al.[1]은 영향력 최대화 문제의 최적의 해를 구하는 것이 NP-난해임을 보이고, k 단계에 걸쳐 각 단계마다 영향력이 최대인 노드를 하나씩 선출하여 최종적으로 k 개의 시드 노드로 이루어진 집합을 구성하는 SimpleGreedy를 제안하였다.

그러나, SimpleGreedy는 미시적, 거시적 측면에서 성능 문제가 존재한다. 먼저 미시적 측면에서 보면, 한 노드의 영향력을 계산하는 것 자체가 많은 시간이 소요된다는 점이다. 이는 한 노드의 영향력을 계산하기 위해 10,000번의 MC-simulation을 수행해야하기 때문이다. 그리고 거시적 측면에서는 이전 단계에서 시드 노드가 선출되면, 다음 단계의 시드 노드를 선출하기 위해 시드 집합을 제외한 모든 노드의 영향력을 새롭게 평가해야 한다는 점이다. 이는 이전 단계에서 선출된 시드 노드에 의해 다른 노드들의 영향력이 감소할 가능성이 존재하기 때문이다.

이러한 미시적, 거시적 측면의 문제를 개선하기 위해 다양한 연구들이 진행되어 오고 있다. 미시적 측면의 문제를 개선한 대표적인 연구로 PB-IM (*path-based influence maximization*)이 있다[2-7]. PB-IM은 영향력을 구하고자 하는 노드에서 도달 가능한 다른 노드들의 경로(path)들의 가중치의 합으로 노드의 영향력을 계산하는 방법이다. 이 방법은 10,000번의 MC-simulation을 해야 하는 SimpleGreedy와 달리, 노드로부터 시작하는 단 한번의 경로 탐색만으로 노드의 영향력을 계산할 수 있어 속도를 크게 개선하였다. 즉, SimpleGreedy의 미시적 측면의 문제를 성공적으로 해결하였다.

거시적 측면의 문제를 개선한 대표적인 연구로는 CB-IM (*community-based influence maximization*)이 있다[8,9]. CB-IM은 소셜 네트워크에서 커뮤니티들이 갖는 고유한 특성을 활용하였다. 소셜 네트워크의 커뮤니티들

은 갖는 특성은 다음과 같다[10]. (1) 동일한 커뮤니티에 속한 노드들은 영향을 활발히 주고받고 (2) 서로 다른 커뮤니티에 속한 노드들은 영향을 거의 주고받지 않는다. 즉, 노드의 커뮤니티 내 영향력과 전체 네트워크에 대한 영향력이 유사하다. 이러한 특성을 활용하여 새로운 시드 노드가 선출되었을 때, 다음 단계의 시드 노드를 선출하기 위해, 직전 단계의 시드 노드가 선출된 커뮤니티 내의 노드들만 영향력을 재계산하였다.

PB-IM과 CB-IM가 SimpleGreedy를 성능을 개선했지만, 다음과 같은 문제가 여전히 존재한다. PB-IM은 미시적 측면의 문제를 개선하였으나, 여전히 거시적 측면에서 성능 문제가 있고, 반대로 CB-IM는 거시적 측면의 문제를 개선했지만, 미시적 측면의 문제가 있다. 이는 PB-IM과 CB-IM 각 방법이 서로 독립적으로 하나의 문제만을 해결했음을 의미한다. 미시적, 거시적 문제를 모두 해결하기 위해 본 논문은 PB-IM과 CB-IM을 결합하는 새로운 영향력 최대화 방안을 제안한다. 나아가 CB-IM의 커뮤니티 탐지 단계에 존재하는 기술적 이슈를 파악하고 해당 이슈를 해결하는 경로 기반 커뮤니티 탐지 전략을 제안한다.

제안하는 방법의 우수성을 규명하기 위해서 네 가지의 실세계 데이터를 이용하여 실험을 진행하였다. 실험을 통해 제안하는 PB-CD로 커뮤니티를 탐지하고 시드 집합을 선출했을 때, CB-IM의 기존 커뮤니티 탐지 방법 이용했을 때 보다 28% 더 큰 영향력을 갖는 시드 집합을 선출하였다. 최종적으로 제안하는 PB-IM과 CB-IM를 결합하고 PB-CD를 적용한 방법이 CB-IM에 비해 28 % 더 큰 영향력을 갖는 시드 집합을 100배 빠르게 선출했고, PB-IM으로 선출한 시드 집합의 영향력의 평균 97%의 영향력을 갖는 시드 집합을 최대 43배 빠르게 선출하였다.

2. 제안하는 방법

2.1 개요

서론에서 언급했듯이, PB-IM과 CB-IM은 미시적, 거시적 문제를 독립적으로 해결하였을 뿐, 두 문제 모두 해결하지는 못하였다. 본 연구는 미시적, 거시적 문제를 모두 해결하기 위해 PB-IM과 CB-IM을 결합하는 방법을 제안한다. 제안하는 방법은 다음 두 가지 단계로 이루어진다.

(1) *커뮤니티 탐지 단계*. 본 단계는 CB-IM과 같이 거시적 측면의 문제를 개선하기 위해 전체 네트워크를 여러 개의 커뮤니티로 분할한다. 커뮤니티 탐지 단계의 상세 과정에 대해서는 다음 장에서 다룬다.

(2) *시드 선출 단계*. 본 단계는 단계 (1)에서 탐지된 커뮤니티들로 이루어진 네트워크를 대상으로 k -시드 집

합을 선출한다. 이때, 노드의 영향력은 PB-IM을 이용하여 계산함으로써 미시적 측면의 문제를 개선한다.

결과적으로 제안하는 방법은 PB-IM과 CB-IM을 결합하여 미시적, 거시적 측면의 문제를 동시에 해결하고, 커뮤니티 탐지 단계를 더욱 정확하게 수행하기 위해 CB-IM의 커뮤니티 탐지 단계의 기술적 이슈를 파악하여 그를 해결하기 위한 새로운 커뮤니티 탐지 기법도 제안한다.

2.2 경로 기반 커뮤니티 탐지(path-based community detection(PB-CD))

노드의 커뮤니티내에서의 영향력은 전체 네트워크에 대한 영향력과 매우 유사하다는 특징이 있다[10]. 그러나 이러한 특징이 명확히 나타나려면 소셜 네트워크로부터 커뮤니티가 올바르게 탐지 되어야 한다는 전제가 필요하다. 만약 서로 다른 커뮤니티에 속한 노드들간 영향 교류의 양이 커지도록 커뮤니티가 탐지된다면 노드의 커뮤니티내 영향력과 네트워크에 대한 영향력의 차이가 커지게 된다. 즉, 노드의 커뮤니티내 영향력을 노드의 영향력으로 사용할 경우, 부정확하게 계산된 영향력을 이용하여 시드 노드를 선출하기 때문에 최종 시드 집합의 영향력에 악영향을 주게 된다. 그러므로, 커뮤니티 탐지 단계가 올바르게 선행되어야 양질의 시드 집합을 선출할 수 있다.

커뮤니티간 노드들이 서로 주고받는 영향력을 정확하게 계산하는 방법은 MC-simulation을 이용하는 것이다. 그러나, MC-simulation을 이용하여 해당 영향력을 계산하게 되면 커뮤니티 탐지 단계에 너무 오랜 시간이 걸린다. 본디 커뮤니티 탐지 단계는 거시적 측면에서 성능을 개선하기 위함인데, 커뮤니티 탐지 단계에 오랜 시간이 걸린다면 그 의미가 퇴색되게 된다. 이러한 이유로 CB-IM은 커뮤니티 탐지 단계를 빠르게 수행하기 위해 특정 임계값 이하의 가중치를 갖는 엣지들은 그래프에서 제거하고 남아 있는 엣지(live edge)들만 이용하여 커뮤니티간 교류되는 영향력을 계산한다.

임계값 이하의 가중치를 갖는 엣지가 제거되어 단순화된 그래프를 대상으로 커뮤니티를 탐지하여 수행 시간은 단축시켰지만, 원래 그래프의 많은 양의 정보가 손실되어 다음과 같은 문제가 발생한다. (1) 너무 많은 수의 엣지들이 무시된다. CB-IM의 커뮤니티 탐지 방법[7]을 이용했을 때, 평균 93%의 엣지가 제거되었다. (2) 커뮤니티 탐지에 이용되는 살아남은 엣지(live edge)들도 그들의 가중치는 완전히 무시되고, 모든 엣지들이 동일하게 취급된다. 이러한 두가지 문제에 의해 커뮤니티간 교류되는 영향력이 올바르게 계산되지 못하고 결과적으로 커뮤니티 탐지 결과의 질이 떨어지게 된다.

이러한 문제들을 해결하기 위해, 본 연구는 임계값 이상의 가중치를 갖는 일부 엣지만을 이용하는 CB-IM

보다 많은 수의 엣지들과 그 가중치를 모두 이용하는 경로 기반 커뮤니티 탐지(path-based community detection(PB-CD)) 전략을 제안한다. PB-CD는 CB-IM보다 더 많은 정보를 이용하여 커뮤니티간 교류되는 영향력을 정확하게 예측한다. 또한, PB-CD는 경로 기반으로 영향력을 계산하기 때문에 더 많은 엣지를 이용함에도 불구하고 합리적인 시간안에 커뮤니티 탐지를 효율적으로 수행한다.

PB-CD는 서로 다른 두 커뮤니티에 대해, 한 커뮤니티에서 다른 커뮤니티로 넘어가는 영향력을 기반으로 두 커뮤니티간 병합 여부를 결정한다. 만약 두 커뮤니티간 넘어가는 영향력이 정해진 값보다 작을 경우, 병합하지 않는다. 이는 커뮤니티가 이미 잘 나누어져 있어 병합이 필요없음을 의미한다. 반대로 넘어가는 영향력이 클 경우, 아직 커뮤니티들이 제대로 나누어져 있지 않아 커뮤니티간 넘어가는 영향력이 큰 것이므로 두 커뮤니티를 결합한다. 두 커뮤니티 간 넘어가는 영향력은 다음과 같이 정의한다.

$$IOF(C_m, C_l) = \max_{v \in C_m, u \in (N(v) \cap C_l)} \frac{w(v, u) \cdot \bar{\sigma}_m(v)}{\sigma_m(v)}$$

위 식에서 $w(v, j)$ 는 노드 v 와 j 를 연결하는 엣지의 가중치를, $N(v)$ 는 v 의 이웃 노드의 집합을, $\sigma_m(v)$ 와 $\bar{\sigma}_m(v)$ 는 각각 노드 v 의 커뮤니티 C_m 내에서의 영향력과 커뮤니티 C_m 밖에서의 영향력을 의미한다. 두 값을 계산하는 방법을 설명하기에 앞서, 일반적으로 노드 v 의 영향력을 구하는 방법을 먼저 설명한다.

$$\sigma(v) = \sum_{i=1} W(p(v, u_i)), \quad W(p) = \prod_{i=1}^{m-1} w(v_i, v_{i+1})$$

$p(v, u)$ 는 노드 v 에서 u 로 이어지는 경로를 의미하고, $W(p(v, u))$ 는 노드 v 에서 u 로 이어지는 경로의 가중치를 의미한다. 결과적으로, $\sigma(v)$ 는 노드 v 에서 출발하는 경로들의 가중치의 합으로 계산된다. 이때, $\sigma(v)$ 은 노드 v 에서 출발하는 경로 중 커뮤니티 m 을 벗어나는 것들을 무시하여 위 식을 적용하면 계산할 수 있다. 그리고, $\bar{\sigma}_m(v)$ 는 $\sigma(v) - \sigma_m(v)$ 으로 계산할 수 있다.

그러나, 모든 경로를 찾는 것은 #P-hard[6]이므로 미리 정의한 임계 값 이하의 경로들은 무시하여 노드의 영향력을 계산한다. 이는 영향력 파급의 대부분이 가까운 이웃들 사이에서 일어나, 경로가 길어질수록 영향력이 매우 작아지고 그러한 경로들은 무시하더라도 노드의 영향력 계산의 결과에는 큰 영향을 주지 않기 때문이다.

다시 $IOF(C_m, C_l)$ 로 돌아오면, $IOF(C_m, C_l)$ 값이 클수록 C_m 에서 C_l 로 넘어가는 영향력이 많아진다는 것을 의미한다. 이러한 직관에 기반하여 커뮤니티 병합 단계는 다음 과정과 같이 동작한다.

- (1) 모든 커뮤니티 쌍 C_m 과 C_l 에 대해,
 - 1) $IOF(C_m, C_l)$ 를 계산한다.
 - 2) $IOF(C_m, C_l)$ 가 미리 정의한 특정 값보다 클 경우, C_m 과 C_l 을 결합한다.
- (2) $IOF(C_m, C_l)$ 가 특정 값보다 큰 쌍이 나오지 않을 때까지 이 과정을 반복한다.

3. 실험 결과

3.1 실험 환경

데이터. 제안하는 방법의 효용성을 입증하기 위해 우리는 4개의 실 세계 데이터를 이용하여 실험을 진행하였다. 사용한 데이터 중, NetHEPT, NetPHY, 그리고 DBLP는 공동 저자들간의 네트워크 데이터로 네트워크의 노드는 논문의 저자들, 엣지는 저자간 공동 연구 유무를 의미한다. Stanford는 웹 문서간 네트워크 데이터로 각 노드는 웹 문서를, 엣지는 두 웹 문서를 연결하는 하이퍼링크를 의미한다.

파급 모델. 노드의 영향력을 측정하기 위해서는 영향력이 어떻게 파급되는 지를 설명하는 모델이 필요하다. 모든 실험은 이 분야에서 가장 많이 이용되는 모델 중 하나인 *weighted cascade* (WC) 모델을 이용하였다 [9]. WC 모델에서 엣지 (u, v) 의 가중치는 $1/degree(v)$ 로 부여한다. $degree(v)$ 는 노드 v 의 인엣지의 개수를

의미한다.

파라미터 설정. PB-IM에서 더욱 효과적으로 노드의 영향력을 계산하기 위해, 경로 가중치의 프루닝(pruning) 임계치를 설정할 필요가 있다[2,6,7]. 프루닝 임계치란 경로 가중치의 하한선으로, 만약 프루닝 임계치보다 작은 가중치를 갖는 경로는 노드의 영향력 계산에서 제외하고 임계치보다 큰 가중치를 갖는 경로들만 노드의 영향력 계산에 이용한다. 각 데이터에 대해, 프루닝 임계치에 따른 시드 집합의 영향력과 선출 시간을 비교하여 최적의 프루닝 임계치를 설정하였다. 최종적으로 NetHEPT, NetPHY, Stanford, DBLP 데이터에 1/320, 1/640, 1/160, 1/320의 프루닝 임계치 값을 설정하였다.

3.2 PB-CD의 커뮤니티 탐지 결과 비교 실험

본 실험을 통해 PB-CD가 CB-IM의 커뮤니티 탐지 방법보다 커뮤니티간 교류되는 영향력을 얼마나 정확하게 계산하여 커뮤니티를 탐지하는지를 보이고자 한다. 각 비교 방법으로 커뮤니티를 탐지하고 동일한 방법으로 시드 집합을 선출하여 선출된 시드 집합의 영향력을 비교하였다. 이때, 공평한 성능 비교를 위해 CB-IM에서 이용한 임계 값과 동일한 값 (0.3)을 이용하여 실험을 진행하였다.

그림 1은 각 비교 방법으로 커뮤니티를 탐지하고 선출한 시드 집합의 영향력을 비교한 결과이다. 그림 1에서 x 축은 커뮤니티 탐지 방법을, y 축은 시드 집합의 영

그림 1 커뮤니티 탐지 방법간 비교(시드 집합의 영향력)

Fig. 1 Comparison of CD methods (influence spread)

향력을 나타낸다. 모든 데이터에서 PB-CD가 기존의 방법을 앞서는 결과를 보였다. 특히, PB-CD는 CB-IM's CD보다 평균 5%의 더 큰 영향력을 갖는 시드 집합을 선출하였다. 이를 통해 제안하는 PB-CD가 기존 방법보다 커뮤니티 간 교류되는 영향력을 올바르게 계산하여 커뮤니티를 탐지함을 보였다.

3.3 제안하는 방법

3.3.1 제안하는 방법의 수행 속도

해당 실험은 제안하는 방법이 다른 시드 집합 선출 방법들과 비교하여 수행 속도 측면에서 얼마나 효율적인 지를 보이고자 한다. 다음 비교 방법으로 시드를 1,000개 선출하면서 하나의 시드가 선출될 때마다 시간을 측정하였다.

비교방법. *SDD* (*single degree discount*)는 노드의 엣지 수가 많은 순으로 시드를 선출하되, 시드 노드가 선출되면 선출된 시드의 이웃 노드들의 엣지 수를 1만큼 감소시키는 방법이다. CB-IM과 PB-IM은 서론에서 설명한 방법이고, 마지막으로 *Proposed Method*는 본 논문에서 제안하는 PB-IM과 CB-IM을 결합하고, 커뮤니티 탐지 단계에 PB-CD를 적용한 방법이다. 그림 2는 각 방법들이 시드를 선출하는데 소요되는 총 수행 시간

을 보여준다. *x*축은 선출된 시드 노드의 수를, *y*축은 수행 시간(로그 스케일)을 나타낸다. CB-IM과 제안하는 방법의 경우, 비교 방법들간 형평성을 위해 커뮤니티 탐지 단계에 소요되는 시간도 전체 수행 시간에 포함시켰다.

제안하는 방법은 모든 데이터에서 나머지 방법들보다 빠르게 시드 집합을 선출한다. 제안하는 방법, PB-IM 그리고 SDD 순으로 빠른 수행 속도를 보인다. 마지막으로 CB-IM이 가장 느린 수행 속도를 보였다. CB-IM은 Stanford와 DBLP 데이터에서 합리적인 시간안에 1,000개의 시드 노드를 선출하지 못했기 때문에 그림 2의 (c), (d)에는 제외되었다.

제안하는 방법은 NetHEPT와 NetPHY 데이터에서 CB-IM보다 약 100배 빠르게 시드 집합을 선출하였다. 그 이유는 제안하는 방법이 PB-IM을 이용하여 미시적 측면에서 성능을 개선하기 때문이다. 그리고 제안하는 방법은 모든 데이터에서 PB-IM보다도 빠르게 시드 집합을 선출한다. PB-IM은 커뮤니티 탐지 단계가 없기 때문에 시드 선출 초기에 제안하는 방법보다 다소 빠르게 시드를 선출하지만, 시드 집합의 크기가 증가할수록 제안하는 방법과 PB-IM의 차이가 줄어들고, 시드 집합의 크기가 일정 값을 넘어가면 제안하는 방법이 PB-IM

(a) NetHEPT

(b) NetPHY

(c) Stanford

(d) DBLP

그림 2 비교 방법들의 시드 선출 시간

Fig. 2 Comparison of IM methods (running time)

(a) NetHEPT

(b) NetPHY

(c) Stanford

(d) DBLP

그림 3 비교 방법들로 선출된 시드 집합의 영향력
Fig. 3 Comparison of IM methods (influence spread)

을 앞선다. 결과적으로 제안하는 방법은 PB-IM과 CB-IM을 결합하여 PB-IM과 CB-IM 모두를 앞서는 결과를 얻었다.

3.3.2 제안하는 방법의 영향력

해당 실험에서는 제안하는 방법이 선출한 시드 집합의 영향력이 다른 비교 방법들로 선출한 시드 집합의 영향력과 비교하여 얼마나 더 큰 지를 확인하고자 한다. 각 비교 방법으로 시드 1,000개를 선출하고 MC-simulation을 이용하여 시드 집합의 영향력을 평가하였다. 그림 3은 각 방법으로 선출한 시드 집합의 영향력을 보여준다. x축은 선출된 시드 노드의 수를, y축은 시드 집합의 영향력을 나타낸다.

제안하는 방법은 CB-IM보다 평균 28% 더 큰 영향력을 갖는 시드 집합을 선출하였다. 이는 CB-IM이 특정 값 이상의 가중치를 갖는 엣지로만 구성된 그래프를 이용하여 올바르게 커뮤니티를 탐지하지 못하기 때문이다. 이 결과를 통해 제안하는 방법의 PB-CD 전략이 유의미함을 다시 한 번 확인할 수 있다.

제안하는 방법은 PB-IM과 비교하여 평균 3.73% 낮

은 영향력을 갖는 시드 집합을 선출하였다. 이러한 작은 손실은 제안하는 방법이 커뮤니티를 이용하기 때문에 발생한다. 그럼에도, 제안하는 방법은 PB-IM보다 최대 43배 빠르게 시드 집합을 선출하였다.

4. 결론

본 연구는 SimpleGreedy의 미시적, 거시적 측면의 성능 문제를 모두 개선하기 위해 PB-IM과 CB-IM을 결합하는 효율적인 영향력 최대화 방안을 제안하였다. 나아가 CB-IM의 커뮤니티 탐지 단계의 기술적 이슈를 파악하고 그를 해결하는 경로 기반 커뮤니티 탐지(PB-CD) 전략도 제안한다. 네 가지의 실 세계 데이터를 이용한 실험을 통해, 제안하는 PB-CD가 기존의 커뮤니티 탐지보다 우수함을 보였고, PB-IM과 CB-IM을 결합함으로써 미시적, 거시적 측면의 문제가 동시에 해결될 또한 보였다.

References

- [1] D. Kempe, J. Kleinberg, and E. Tardos, "Maximizing

- the spread of influence through a social network," *Proc. of the ninth ACM SIGKDD*, pp. 137-146, 2003.
- [2] W. Chen, C. Wang, and Y. Wang, "Scalable influence maximization for prevalent viral marketing in large-scale social networks," *Proc. of the 16th ACM SIGKDD*, pp. 1029-1038, 2010.
- [3] W. Chen, Y. Wang, and S. Yang, "Efficient influence maximization in social networks," *Proc. of the 15th ACM SIGKDD*, pp. 199-208, 2009.
- [4] W. Chen, Y. Yuan, and L. Zhang, "Scalable influence maximization in social networks under the linear threshold model," *Proc. of the 10th IEEE ICDM*, pp. 88-97, 2010.
- [5] A. Goyal, W. Lu, and L. V. Lakshmanan, "Simpath: An efficient algorithm for influence maximization under the linear threshold model," *Proc. of the 11th IEEE ICDM*, pp. 211-220, 2011.
- [6] J. Kim, S. Kim, and H. Yu, "Scalable and parallelizable processing of influence maximization for large-scale social networks?," *Proc. of the 29th IEEE ICDE*, pp. 266-277, 2013.
- [7] Y. Ko, D. Chae, and S. Kim, "Accurate path-based methods for influence maximization in social networks," *Proc. of the 25th IW3C2 WWW*, pp. 59-60, 2016.
- [8] G. Song et al., "Influence maximization on large-scale mobile social network: a divide-and-conquer method," *Journal of IEEE Transactions on Parallel and Distributed Systems*, Vol. 26, No. 5, pp. 1379-1392, May. 2015.
- [9] Y. Wang et al., "Community-based greedy algorithm for mining top-k influential nodes in mobile social networks," *Proc. of the 16th ACM SIGKDD*, pp. 1039-1048, 2010.
- [10] J. Leskovec et al., "Community structure in large networks: Natural cluster sizes and the absence of large well-defined clusters," *Journal of Internet Mathematics*, Vol. 6, No. 1, pp. 29-123, Jul. 2010.

조 경 재

2016년 한양대학교 컴퓨터전공 졸업 (학사). 2016년~현재 한양대학교 컴퓨터 소프트웨어과 석사 과정. 관심분야는 사회 연결망 분석, 데이터 마이닝

김 상 옥

1985년~1989년 서울대학교 컴퓨터공학과 졸업(학사). 1989년~1991년 한국과학기술 전산학과 졸업(석사). 1991년~1994년 한국과학기술원 전산학과 졸업 (박사). 1994년~1995년 KAIST 정보전자연구소 위촉 연구원. 1999년~2000년

IBM T.J. Watson Research Center, Post-Doc. (invited by Dr. Philip Yu). 1995년~2003년 강원대학교 정보통신공학과. 2003년~2006년 한양대학교 정보통신대학 정보통신학부 부교수. 2006년~현재 한양대학교 정보통신대학 정보통신학부 교수. 관심분야는 데이터베이스 시스템, 저장 시스템, 트랜잭션 관리, 데이터 마이닝, 멀티미디어 정보 검색, 공간 데이터베이스/GIS, 주기억장치 데이터베이스, 이동 객체 데이터베이스/텔레매틱스, 사회연결망 분석, 웹 데이터 분석

고 윤 용

2013년 한양대학교 컴퓨터·소프트웨어학과 졸업(학사). 2013년~현재 한양대학교 컴퓨터·소프트웨어학과 박사 과정. 관심분야는 사회 연결망 분석, 데이터 마이닝