사회관계망에서 효율적인 매개 중심도 계산을 위한 지역정보기반 알고리즘1)

신수진, 김용환, 김찬명, 한연희2)
한국기술교육대학교 컴퓨터공학부/첨단기술연구소
{soojin1116, cherish, cmdr, yhhan}@koreatech.ac.kr

Local Information-based Algorithm for Efficient Calculation of Betweenness Centrality in Social Networks

Soo-Jin Shin, Yong-hwan Kim, Chan-Myung Kim and Youn-Hee Han+
School of Computer Science and Engineering
/Advanced Technology Research Center,
Korea University of Technology and Education

요 약

사회관계망 분석에 있어서 매개중심도는 네트워크를 구성하는 노드들의 상대적인 중요도를 파악하기 위한 척도로서 오랫동안 사용되어 왔다. 그러나 이러한 매개중심도를 계산하기 위한 계산 복잡도가 높기 때문에 대규모 사회관계망에서는 매개중심도를 계산하기란 쉽지 않은 문제이다. 본 논문에서는 네트워크를 구성하는 각각의 노드들마다 자신의 지역정보를 활용하여 구성한 네트워크에서 매개중심도를 산출함으로써 시간복잡도를 줄이는 한편 지역정보 기반의 네트워크의 특징을 분석함으로써 매개중심도를 더 빠르게 산출할 수 있는 알고리즘을 제안한다. 그리고 실제 소셜 네트워크에서의 실험을 통하여 제안 알고리즘이 기존 알고리즘에 비해 매개중심도를 더 빠르게 산출함을 보인다.

1. 서론

사회관계망(Social Network)은 개인 또는 집단 사이의 상호 관계를 연결선으로 표현하여 구성한 논리적 네트워크이다 [1]. 이러한 사회관계망을 분석하면 네트워크에 참여하고 있는 각 사용자들 간의 관계와, 이들 관계의 패턴과 함의를 파악할 수 있다 [2].

특히 사회관계망 분석 기법 중 하나인 매개 중심성 분석은 네트워크를 구성하는 노드들 중에서 중요도가 높은 노드를 선별하기 위한 가장 중요한 척도 중 하나로서 많이 활용되고 있는데, 이는 사회관계망 내의 임의의 노드가 다른 노드들 사이의 경로 상에 위치되는 정도를 나타낸다 [3]. 그러나 매개중심도를 구하기 위해서는 각 노드 사이의 최단경로를 구하는 작업을 수행해야 하는데, 이의 계산은 시간복잡도가 높다고 알려져 있다. 따라서 대규모 사회관계망에서 각 노드의 매개중심도를 산출하는 작업은 매우 어렵다. 이와 같은 문제점을 해결하기 위해 사회관계망에서 빠른 시간 내에 모든 노드들에 대한 매개중심도를 추정해내는 방안에 관한 연구가 요구된다.

본 논문에서는 빠른 시간 내에 매개중심도를 구하기 위하여 사회관계망을 구성하는 각 노드마다 지역 정보, 즉 이웃 노드의 친구 관계 정보만을 활용하여 구성할 수 있는 "확장 자아 네트워크" 개념을 사용한다. 확장 자아 네트워크는 [5]에서 제안된 "자아 네트워크 (Ego Network)"를 기반으로 하여 본 연구팀에서 이전에 정의한 것으로서 [7], 포함하는 지역 정보의 양이 자아 네트워크보다 다소 많기 때문에 노드 중심성 분석의 정확도가 높은 편이다.

또한 본 논문에서는 확장 자아 네트워크의 특성을 분석하여 기존의 알고리즘보다 효율적으로 매개 중심성을 산출 할 수 있는 알고리즘을 제시하고 실제 페이스북 네트워크에서의 실험을 통하여 이의 효율성을 입증한다.

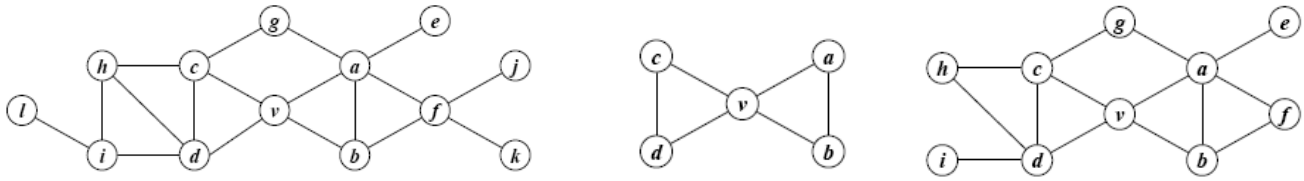
본 논문의 구성은 다음과 같다. 2장에서는 매개중심도와 확장 자아 네크워크에 대해 설명한다. 3장에서는 확장 자아네트워크의 특성을 활용하여 매개중심성을 계산하는 알고리즘을 제안한다. 4장에서는 실제 소셜 네트워크에서 제안하는 알고리즘과 기존의 알고리즘을 비교분석하여 제안하는 알고리즘의 효율성을 보인다. 마지막으로 5장에서는 본 논문의 결론 및 향후 연구에 대해 설명한다.

2. 관련연구

2.1 매개중심도

본 논문에서 활용하는 사회관계망 분석 척도인 임의의 노드 v의 매개중심도 (Betweenness Centrality, $B(v)$)는 다음과 같이 정의된다 [2].

1) 본 논문은 정부(교육부)의 재원으로 한국연구재단의 지원을 받아 수행된 기초연구사업임 (No. NRF-2013R1A1A2010050)
2) 한연희: 교신저자



(a) 전체 네트워크 (b) 자아 네트워크 (c) 확장 자아 네트워크

(그림 1) 사회관계망 네트워크

$$B(v) = \frac{\sum_{s \neq v \neq t \in V} \frac{\sigma_{st}(v)}{\sigma_{st}}}{(|V|-1)(|V|-2)/2} \quad (1)$$

- V : 모든 노드의 집합
- σ_{st} : 노드 s와 t사의의 최단경로 의 개수
- $\sigma_{st}(v)$: 노드 s와 t사의의 최단경로 중에서 노드 v를 통과하는 경로의 개수

매개중심도 $B(v)$는 노드 s와 t사이의 노드 v를 통과하는 최단경로의 개수인 $\sigma_{st}(v)$를 노드 s와 t사이의 최단경로의 개수인 σ_{st}로 나눈 값을 v를 제외한 모든 쌍에 대하여 합하고 이를 v를 제외한 모든 쌍의 수로 나눈 값을 말한다. 여기서 $\sigma_{st}(v)/\sigma_{st}$는 s에서 t사이의 모든 경로들이 노드 v에게 얼마나 종속되어있는지 정도를 의미한다. 본 논문에서는 이를 s와 t의 v에 대한 의존성이라 명명하고 이를 다음과 같이 정의한다.

$$\delta_{st}(v) = \frac{\sigma_{st}(v)}{\sigma_{st}} \quad (2)$$

임의의 노드간의 최단경로를 구하기 위한 시간복잡도는 기본적으로 $O(n^2)$정도이기 때문에 노드의 개수가 많아지면 실용성이 매우 낮아진다. 그래서, Brandes은 매개중심도를 빠르게 계산하기 위해 [4]에서 복잡도가 $O(nm)$인 대수적 기법을 제안하였다. 한편, Everett와 Borgatti은 [5]에서 기존 사회관계망 분석에서 오랫동안 사용해 온 "자아 네트워크"를 각 노드마다 구성하고 그러한 각각의 자아 네트워크로부터 매개중심도를 산출하여 얻어낸 각 노드들의 상대적인 중요도 순위가 전통적인 방법으로 전체 네트워크에서 각 노드의 매개중심도를 산출하여 얻어낸 각 노드들의 중요도 순위와 매우 유사함을 실험을 통해 증명했다.

본 논문 이전의 연구에서는 [5]의 결과를 발전시켜서 자아 네트워크와 그 구성측면에서 구성의 용이성이 동일한 확장 자아 네트워크를 정의하고, 전체 네트워크 기반의 노드 중요도 순위와 유사성이 높은 중요도 순위를 산출할 수 있는 확장 자아 매개중심도를 제안하였다 [7].

본 논문은 이러한 확장 자아 네트워크의 특성을 분석하여 확장 자아 네트워크에서 기존의 매개중심도를 도출하는 Brandes알고리즘보다 보다 효율적으로 매개 중심성을 계산 할 수 있는 알고리즘을 제안한다.

2.2 자아 네트워크와 확장 자아 네트워크

주어진 사회관계망 그래프 $G(V,E)$와 그 네트워크에 속한 임의의 노드 v에 대하여 다음과 같은 기본적인 용어를 정의한다.

- $V_{v:1}$: 노드 v와 직접적으로 연결된 이웃 노드들의 집합 (즉, v에서부터 1-hop 떨어져 있는 노드들의 집합)
- $V_{v:2}$: v에서부터 2-hop 떨어져 있는 노드들의 집합)
- $E_{v:0:1}$: 노드 v와 v의 1-hop 노드들의 집합($V_{v:1}$)사이의 간선들의 집합
- $E_{v:1}$: $V_{v:1}$의 노드 간의 간선들의 집합
- $E_{v:1:2}$: 노드 v의 1-hop 노드들의 집합($V_{v:1}$)과 노드 v의 2-hop 노드들의 집합($V_{v:2}$)사이의 간선들의 집합

위의 용어를 사용하여 "자아 네트워크 (Ego Network, ε_v)와 확장 자아 네트워크 (Expanded Ego Network, X_v)"를 다음과 같이 정의할 수 있다 (그림1(b), 그림1(c) 참조).

- ε_v: $(\{v\} \cup V_{v:1},\ E_{v:0:1} \cup E_{v:1})$
- X_v: $(\{v\} \cup V_{v:1} \cup V_{v:2},\ E_{v:0:1} \cup E_{v:1} \cup E_{v:1:2})$

임의의 노드 v에 대한 자아 네트워크 ε_v와 확장 자아 네트워크 X_v는 사회관계망 분석에서 자주 사용되는 모델로서 흔히 v를 중심으로 한 정보를 얻어내기 위하여 사용하는 주어진 전체 네트워크의 부분 네트워크이다 [7, 9]. 이 때, 확장 자아 네트워크에서는 자아 네트워크에 비하여 더 많은 주변 정보를 활용하기 때문에 좀 더 의미 있는 데이터를 산출할 수 있게 된다. 예를 들어, 그림 1에서 주어진 전체 네트워크로부터 매개중심도를 산출하여 얻어낸 노드들의 중요도 순위와 자아 네트워크와 확장 자아 네트워크 각각으로부터 매개중심도를 산출하여 얻어낸 노드들의 중요도 순위들 간의 상관계수는 각각 0.79과 0.93으로 확장 자아 네트워크에서 보다 전제 네트워크와 유사한 결과를 도출해 낼 수 있음을 확인 할 수 있다.

3. 제안하는 알고리즘

지역 정보를 활용하여 구성되는 확장 자아 네트워크에서 매개중심도를 계산하는 작업은 전체 네트워크 때의 비해 상당히 적은 시간을 필요로 한다. 그러나 사회관계망에 속한 노드의 수가 상당히 많거나 노드들 간의 상호 관계가 복잡하여 각각의 노드들의 확장 자아 네트워크의 크기가 커질 경우, 매개중심도를 계산하기 위하여 적지 않은 시간을 필요로 하게 될 것이다.

식(1)에서 알 수 있듯이 임의의 노드 v의 매개중심도는 v를 제외한 노드의 모든 쌍의 의존성 $\delta_{st}(v)$(식 (2))을 계산함으로써 구해낼 수 있다. 이 때. 임의의 노드 쌍의 v에 대한 의존성이 없어 이가 노드 v의 매개 중심성에 아무런 영향을 미치지 않는다면 해당 노드 쌍의 최단경로를 계산하는 작업은 불필요한 일일 것이다. 제안 알고리즘에서는 확장 자아네트워크의 특성을 분석하여 이러한 불필요한 작업을 제거하는 한편 $\delta_{st}(v)$를 조금 더 간단하게 계산하는 방법을 제시함으로써 매개중심도를 산출하는 시간을 단축시키고자 한다. 제안하고자 하는 알고리즘은 다음과 같다.

Algorithm 1

```
1:  Input: X_v, P[1...u+w][1...u+w]
2:  output: B^χ(v)
3:  for s : 1 to u do
4:      for t : s+1 to u do
5:          if t ∈ V_{s:1} then
6:              P[s][t] ← 0.0
7:          else
8:              P[s][t] ← 1/|V_{s:1} ∩ V_{t:1}|
9:          total ← total+P[s][t]
10:     end for
11:     for t : u+1 to u+w do
12:         for each k∈ V_{t:1} do
13:             if t < k then
14:                 append P[s][k] to C
15:             else
16:                 append P[k][s] to C
17:             end if
18:         end for
19:         P[s][t] ← H(C)
20:         total ← total+P[s][t]
21:     end for
22: end for
23: for s : u+1 to u+2 do
24:     for t : s+1 to u+w do
25:         for each k∈ V_{t:1} do
26:             if s < k then
27:                 append P[s][k] to C
28:             else
29:                 append P[k][s] to C
30:             end if
31:         end for
32:         P[s][t] ← H(C)
33:         total ← total+P[s][t]
34:     end for
35: end for
36: return 2 • total / ((u+w)(u+w−1))
```

제안 알고리즘에서 D는 임의의 쌍에 대한 의존성 값에 대한 이차원 배열, C는 의존성을 계산하기 위한 임시 리스트, u는 v의 1-hop 노드의 수, w는 v의 2-hop 노드의 수이다.

3-10번째 줄에서는 임의의 노드 쌍 (s,t)이 v의 1-hop 노드인 경우, 의존성 $\delta_{st}(v)$을 계산한다. 여기서 s와 t 가 이웃이라면 최단경로는 $s-t$로 v를 거치는 경로가 발생할 수 없기 때문에 $\sigma_{st}(v)$가 0되며 $\delta_{st}(v)$ 또한 0이다. 따라서, 이의 경우는 의존성의 계산을 하지 않는다. 그렇지 않을 경우에는 s와 t사이의 최단 경로는 두 노드 사이의 공통된 이웃의 수와 같고 $\sigma_{st}(v)$는 1이므로 $\delta_{st}(v)$는 $1/|V_{s:1} \cap V_{t:1}|$ 과 같다.

11-21번째 줄에서는 노드 s는 v의 1-hop노드이고 노드 t는 v의 2-hop 노드인 경우, 의존성 $\delta_{st}(v)$을 계산한다. 이의 경우, s와 t사이의 최단 경로는 s와 t의 1-hop노드들의 사이의 최단 경로를 거쳐 t까지 가게 되므로 $\delta_{st}(v)$는 s와 t의 1-hop노드 쌍의 의존성 값들의 조화평균(H(C))함으로써 구할 수 있다. 이 때, s와 t의 1-hop노드 쌍이 v의 1-hop노드들이 아니라면, 6번째 줄에 초기화된 값인 0을 사용하므로 $\delta_{st}(v)$에는 아무런 영향을 미치지 않게 된다.

23-34번째 줄에서는 임의의 노드 쌍 (s,t)이 v의 2-hop 노드인 경우, 의존성 $\delta_{st}(v)$을 계산한다. 이의 방법은 노드 s는 v의 1-hop노드이고 노드 t는 v의 2-hop 노드인 경우에 의존성 $\delta_{st}(v)$을 계산하는 방법과 동일하다.

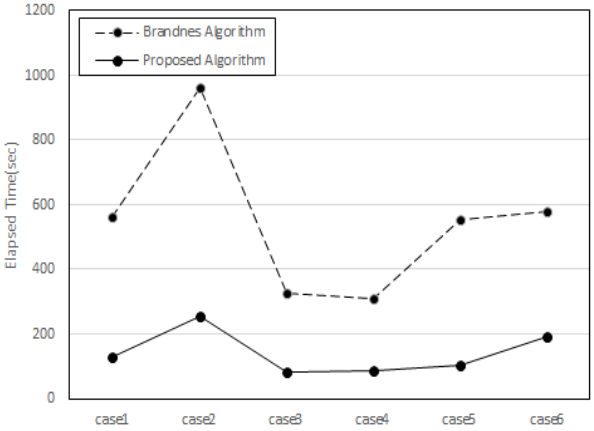
36번째 줄에서는 계산되어진 모든 노드 쌍 (s,t)의 의존성 $\delta_{st}(v)$을 합하여 $total$을 구하고 식 (1)과 같이 전체노드에 대하여 생성되는 모든 간선의 수로 나누어 정규화시켜서 확장 자아 네트워크의 매개 중심도 값을 반환한다.

즉, 제안 알고리즘은 확장 자아네트워크의 특성을 분석하여 불필요한 의존성 계산 작업을 제거하고 $\delta_{st}(v)$의 계산을 위한 최단 경로의 수를 구하기 위하여 Dijkstra 알고리즘과 같은 최단 경로 알고리즘이 아닌 노드간의 관계 정보만으로 최단 경로의 수를 계산함으로써 매개중심도를 산출하는 시간을 상당부분 단축시킨다.

4. 실험결과 및 분석

본 장에서는 실제 소셜 네트워크 데이터를 활용하여 기존의 Brandes 알고리즘과 제안하는 알고리즘이 확장 자아 네트워크에서의 매개중심도를 계산하는데 소요시간을 비교 분석한다.

해당 네트워크는 SNAP에서 제공하는 페이스북 데이터를 이용하여 총 6개의 전체 네트워크를 구축하였고 네트워크를 구성하는 각각의 노드마다 확장 자아 네트워크를 생성하여 두 알고리즘을 비교하였다.



(그림 2) 확장 자아 매개중심도 계산 소요시간

(그림 2)는 매개중심도를 구하기 위한 두 알고리즘의 수행시간을 비교한 그래프로써 여러 네트워크 환경에서 실험을 수행하였다. (그림 2)에서 볼 수 있듯이, 제안한 알고리즘을 사용해 확장 자아 매개중심도를 도출했을 때 걸린 시간이 Brandes알고리즘을 사용한 했을 때 걸린 시간보다 더 짧은 시간이 소요됨을 확인할 수 있다. 제안한 알고리즘을 사용했을 때 6개의 전체 네트워크 모두 예외 없이 소요시간이 단축되었고, 그 차이가 확연하였다.

(표 1) 페이스북 네트워크 특성

(Scor. = 스피어만 상관계수, G : 전체 네트워크)

	노드 수	간선 수	밀도	Scor.(G.X_v)	Scor.(G,ε_v)
case1	449	3381	0.033	0.854	0.711
case2	544	5025	0.034	0.949	0.58
case3	534	4892	0.034	0.883	0.653
case4	660	7479	0.034	0.903	0.641
case5	447	5438	0.054	0.924	0.4
case6	392	4959	0.064	0.928	0.604

소요 시간 외에도 확장 자아 매개중심도의 정확도를 확인하기 위하여 (표 1)에 총 6개의 페이스북 네트워크 특성을 요약하여 제시하였다. 여기서 네트워크의 밀도는 전체 네트워크에서 실제 간선의 개수를 노드 개수 별 생성 가능한 전체 간선의 개수로 나눈 수치를 의미하며 스피어만 상관계수는 전체네트워크, 확장자아네트워크 그리고 자아네트워크에서 매개중심도의 높은 값부터 순위(Rank)를 매기고 그 순위에 따른 유사성을 측정한 값이다. (표 1)에서 확인 할 수 있듯이 전체 네트워크의 매개 중심도와 확장자아 네트워크의 매개중심도 간의 유사도가 대부분 1에 가까운 수치를 보여줌으로써 전체 네트워크를 기준으로 한 확장 자아 네트워크의 정확도가 높음을 확인할 수 있었다.

5. 결론

본 논문은 네트워크를 구성하는 각각의 노드들마다 자신의 지역정보를 활용하여 구성한 네트워크에서 매개중심도를 산출함으로써 시간복잡도를 줄이는 한편 사회관계망에 속한 노드의 수가 상당히 많거나 노드들 간의 상호 관계가 복잡하여 각각의 노드들의 확장 자아 네트워크의 크기가 커질 경우를 고려해야했다. 그 결과 확장 자아 네트워크의 특성을 분석하여 매개중심도를 더 빠르게 산출할 수 있는 알고리즘을 제안하였고 실제 소셜 네트워크에서의 실험을 통하여 제안 알고리즘이 기존 알고리즘에 비해 매개중심도를 더 빠르게 산출함을 증명하였다.

본 논문에서 제안한 알고리즘은 대규모 사회 관계망에서 확장 자아 매개중심도 계산할 때 소요 시간을 단축하는데 많은 도움이 될 것으로 기대된다. 특히 대중적으로 잘 알려진 페이스북은 임의의 대상 한명 한명마다 친구간의 관계가 복잡한데 이런 경우에 좀 더 빠른 시간 내에 매개중심도를 측정하도록 활용 될 수 있다. 본 논문의 연구결과를 더욱 확장하여 향후에는 매개중심도 뿐만 아니라 다양한 사회관계망 분석 기법에 대해서 시간복잡도를 감소시킬 수 있는 방안과 적용할 수 있는 알고리즘에 대해 연구할 예정이다.

참고문헌

[1] H. Kwak, C. Lee, H. Park, and S. Moon, “What is Twitter, a social network or a new media?,” in Proceedings of the 19th International World Wide Web (WWW) Conference, pp. 591-600, April 26-30, 2010.

[2] Lei Tang and Huan Liu, “Community Detection and Mining in Social Media,” Synthesis Lectures on Data Mining and Knowledge Discovery, Vol. 2, No. 1, 2010.

[3] S. Wasserman and K. Faust. “Social Network Analysis: Methods and Applications,” Cambridge University Press, 1994.

[4] U. Brandes. “A Faster Algorithm for Betweenness Centrality,” Journal of Mathematical Sociology, Vol. 25, No. 2, pp.163-177, 2001.

[5] M. Everett, S. P. Borgatti, “Ego Network Betweenness,” Social Networks, Vol. 27, No. 1, pp. 31-38, Jan. 2005.

[6] F. Odella, “Using Ego-networks in Surveys: Methodological and Research Issues,” In Proceedings of International Conference on Network Science, May 2006.

[7] J,-G. Son, Y.-h. Kim and Y.-H. Han “Local Information-based Betweenness Centrality to Identify Important Nodes in Social Networks” The KIPS Transactions on Computer and Communication Systems, Vol. 2, No. 5, pp. 209-216, 2013