

Local Information-based Betweenness Centrality to Identify Important Nodes in Social Networks

Jin Gon Shon[†] · Yong-hwan Kim^{**} · Youn-Hee Han^{***}

ABSTRACT

In traditional social network analysis, the betweenness centrality measure has been heavily used to identify the relative importance of nodes in terms of message delivery. Since the time complexity to calculate the betweenness centrality is very high, however, it is difficult to get it of each node in large-scale social network where there are so many nodes and edges. In this paper, we define a new type of network, called the *expanded ego network*, which is built only with each node's local information, i.e., neighbor information of the node's neighbor nodes, and also define a new measure, called the *expanded ego betweenness centrality*. Through the intensive experiment with Barabási-Albert network model to generate the scale-free networks which most social networks have as their embedded feature, we also show that the nodes' importance rank based on the expanded ego betweenness centrality has high similarity with that based on the traditional betweenness centrality.

Keywords : Social Network Analysis, Betweenness Centrality, Local Information, Expanded Ego Network

사회관계망에서 중요 노드 식별을 위한 지역정보 기반 매개 중심도

손진곤[†] · 김용환^{**} · 한연희^{***}

요약

전통적인 사회관계망 분석에 있어서 각 노드의 매개 중심도는 메시지 전달 측면에서의 각 노드들의 상대적인 중요도를 파악하는 척도로 오랫동안 사용되어 왔다. 하지만, 매개 중심도를 산출하기 위한 계산 복잡도가 높기 때문에 노드의 수와 간선의 수가 매우 많은 대규모 사회관계망에서는 각 노드의 매개 중심도를 산출하기가 어렵다. 본 논문에서는 각 노드들마다 자신의 지역정보, 즉 이웃노드들이 지닌 각각의 이웃노드 정보를 활용하여 구성가능한 확장 자아 네트워크(Expanded Ego Network)를 새롭게 정의하고 이러한 네트워크를 기반으로 확장 자아 매개 중심도(Expanded Ego Betweenness Centrality)를 정의한다. 일반적인 사회관계망의 특성인 척도 없는 네트워크(Scale-free Network)를 생성할 수 있는 Barabási-Albert 네트워크 모델을 사용한 실험을 통하여 제안한 확장 자아 매개 중심도의 각 노드별 순위는 기존의 전통적인 방식으로 산출한 매개 중심도의 각 노드별 순위와 거의 일치함을 보인다.

키워드 : 사회관계망 분석, 매개 중심도, 지역 정보, 확장 자아 네트워크

1. 서론

사회관계망(Social Network)은 개인과 개인 사이의 사회적 관계를 연결선으로 표현하여 구성된 논리적 네트워크를 지칭한다[1]. 이러한 사회관계망을 기반으로 하는 웹 서비스들이 현재 계속해서 증가하고 있으며, 이러한 웹 서비스들은 사용자

들에게 다른 사용자들과 관계를 맺고, 관심사를 공유 하고, 서로 소통할 수 있도록 도와줌으로써 사람들 간의 인맥형성과 정보교환을 위한 새로운 플랫폼을 제시하고 있다[2]. 국내에서는 SK커뮤니케이션즈의 싸이월드가 최대 규모의 사회관계망 기반 웹서비스를 시작하였지만 최근에는 트위터(Twitter), 페이스북(Facebook), 링크드인(LinkedIn) 등의 외국계 서비스가 대표적인 사회관계망 기반 웹서비스로서 자리를 잡았다. 또한, 이러한 웹 서비스들은 스마트폰의 결합으로 사람 사이의 관계형성에 많은 변화를 가져다주고 있다[1, 3].

이러한 사회관계망 웹 서비스들은 자신들이 관리하는 사회관계망 데이터를 Open API를 통해 여러 IT 개발자, 데이터

[†] 중신회원: 한국방송통신대학교 컴퓨터학과 교수
^{**} 정회원: 한국기술교육대학교 컴퓨터공학부 박사과정
^{***} 중신회원: 한국기술교육대학교 컴퓨터공학부 부교수
 논문접수: 2013년 3월 19일
 수정일: 1차 2013년 4월 16일
 심사완료: 2013년 4월 16일
 * Corresponding Author: Youn-Hee Han(yhhan@koreatech.ac.kr)

분석가, 사회과학자 등에게 공개하고 있다. 그러한 Open API를 이용하여 전통적인 통계사회학(Statistical Sociology)에서 활용되어왔던 사회관계망 분석 기법을 이용하여 사회관계망 서비스에 참여하고 있는 각 사용자들 간의 관계(relationship)와, 이들 관계의 패턴과 함의를 파악할 수 있다[3].

사회관계망 분석 기법 중에는 노드 중심성(Centrality) 분석, 노드 유사성(Similarity) 분석, 약한 간선(Weak Tie)/강한 간선(Strong Tie) 분석, 커뮤니티 검출(Community Detection) 등 다양한 기법들이 존재한다[3, 4]. 그 중에서 노드 중심성은 네트워크에 존재하는 노드들 중에서 중요도가 높은 노드를 선별하기 위하여 사용하며 대표적으로 연결 중심도(Degree Centrality), 거리 중심도(Closeness Centrality), 매개 중심도(Betweenness Centrality), 고유벡터 중심도(Eigenvector Centrality) 측정들이 존재한다.

본 논문은 이 4 가지 중심도 측정들 중에서 매개 중심도에 관한 연구 결과를 소개한다. 각 노드들의 매개 중심도를 구하기 위해서는 각 노드 사이의 최단경로를 구하는 작업을 수행해야 하며 이와 같은 계산은 시간 복잡도(Time Complexity)가 높다고 알려져 있다[3]. 따라서, 노드의 수와 간선의 수가 매우 많은 대규모 사회관계망에서 각 노드의 매개 중심도를 산출하거나 각 노드들에 대한 매개 중심도 측면의 상대적인 중요도 순위를 얻어내는 작업은 매우 어렵다. 본 논문에서는 이와 같은 문제점을 해결하기 위해 사회관계망에서 빠른 시간 내에 모든 노드들에 대한 매개 중심도 기반 중요도 순위를 추정해내는 새로운 방안을 제안한다.

이러한 제안을 위하여 각 노드마다 지역 정보 (Local Information), 즉 이웃 노드의 친구 관계 정보만을 활용하여 구성할 수 있는 “확장 자아 네트워크(Expanded Ego Network)”를 새롭게 정의한다. 확장 자아 네트워크(Expanded Ego Network)는 [7]에서 제안된 “자아 네트워크(Ego Network)”를 기반으로 하여 정의된 것으로서, 확장 자아 네트워크는 기존 자아 네트워크와 구성의 용이성은 비슷하지만 포함하는 지역 정보의 양이 다소 많기 때문에 확장 자아 네트워크에서 노드 중심성 분석의 정확도가 높을 수 있다. 한편, 본 논문에서는 각 노드마다 확장 자아 네트워크를 구성하고 이 네트워크에서 각각 독립적으로 “확장 자아 매개 중심도(Expanded Ego Betweenness Centrality)”를 산출하여 이들 간의 상대적 중요도 순위를 정한다. 이렇게 정한 순위는 원래 주어진 전체 네트워크에서 전통적인 매개 중심도 산출 기법을 통하여 얻은 각 노드들의 중요도 순위와 매우 유사함을 다양한 실험을 통하여 증명한다.

본 논문의 구성은 다음과 같다. 2장에서는 매개 중심도에 대한 설명 및 매개 중심도를 빠르게 얻어내기 위한 여러 기존 연구들에 대해 소개한다. 3장에서는 본 논문에서 제안하는 확장 자아 네트워크 및 확장 자아 매개 중심도에 대해 자세히 설명한다. 4장에서는 제안하는 확장 자아 매개 중심도가 기존의 매개 중심도와 비교하여 노드의 중요도 순위를 거의 유사하게 산출할 수 있음을 실험을 통해 증명한다. 마지막으로 5장에서는 본 논문의 결론 및 향후 연구에 대해 설명한다.

2. 관련 연구

본 논문에서 활용하는 사회관계망 분석 척도인 임의의 노드 v_i 의 매개 중심도 (Betweenness Centrality, $C_B(i)$)는 다음과 같이 정의된다[3].

$$C_B(i) = \frac{\sum_{v_s \neq v_i \neq v_t \in V, s < t} \frac{\sigma_{st}(i)}{\sigma_{st}}}{(n-1)(n-2)/2} \quad (1)$$

위 식 (1)에서 V 는 모든 노드의 집합이며, n 은 노드의 총 개수이다. 한편, σ_{st} 는 노드 v_s 와 v_t 사이의 최단경로(shortest path)의 개수이며, $\sigma_{st}(i)$ 는 노드 v_s 와 v_t 사이의 최단경로 중에서 v_i 를 통과하는 것들의 개수이다. 한편, 식 (1)에서 $(n-1)(n-2)/2$ 는 $C_B(i)$ 를 0과 1사이로 정규화(Normalization)하기 위한 식이다.

식 (1)에서 보여주듯이 임의의 노드 v_i 의 매개 중심도 $C_B(i)$ 값은 네트워크에 존재하는 임의의 두 노드 사이의 최단 경로가 노드 v_i 를 많이 지나갈수록 높아진다. 즉, 다른 노드들끼리 메시지를 교환할 때 그 메시지는 높은 매개 중심도를 지닌 노드를 지나갈 확률이 높아지며, 역으로 네트워크에 존재하는 모든 노드들에게 전달할 메시지가 있을 때 매개 중심도가 높은 노드에게 먼저 전달해주면 그 메시지가 전체 노드들에게 전파되어지는 속도가 빨라질 수 있다.

이렇듯, 매개 중심도는 사회관계망 연구에 있어서 가장 중요한 척도중 하나로서 많이 활용되어 왔지만, 노드들 사이의 최단경로를 구하기 위한 Dijkstra 알고리즘의 복잡도는 기본적으로 $O(n^2)$ 정도이기 때문에 네트워크에 속한 노드의 개수가 많아지면 매개 중심도를 구하기 위한 시간이 매우 길어지고 실용성이 낮아진다.

그래서, Brandes은 [5]에서 매개 중심도를 빠르게 계산하기 위한 새로운 대수적 기법을 제안하고 있으며 이 기법의 복잡도는 노드의 개수가 n 이고 간선의 개수가 m 일 때 $O(nm)$ 이다. 즉, 간선의 개수가 노드의 개수 보다 작은 희소 네트워크 (Sparse Network)인 경우 본 기법을 통해 좀 더 빠르게 각 노드의 매개 중심도를 구할 수 있다. 또한, 최근에 Baglioni 등은 [6]에서 [5]에서 제안한 방법을 더욱 발전시켜 2배 정도 복잡도가 개선된 매개 중심도를 산출 방안을 제시했다.

한편, Everett와 Borgatti은 [7]에서 기존의 방법과는 다른 방법을 사용하여, 기존 사회관계망 분석에서 오랫동안 사용해 온 “자아 네트워크 (Ego Network)”를 각 노드마다 구성하고 그러한 각각의 자아 네트워크로부터 매개 중심도를 산출하여 얻어낸 각 노드들의 상대적인 중요도 순위가 전통적인 방법으로 전체 네트워크에서 각 노드의 매개 중심도를 산출하여 얻어낸 각 노드들의 중요도 순위와 매우 유사함을 실험을 통해 증명했다. 임의의 노드에 대한 자아 네트워크는 그 노드와 반경 1-hop 내에 존재하는 노드들과 간선으로

이루어진 네트워크로서 구성이 매우 용이한 장점이 있으면서 해당 노드를 중심으로 하는 중요한 정보를 많이 담고 있어서 통계사회학에서 자주 활용되어 왔다[8]. Everett와 Borgatti는 그들의 실험에서 두 노드 사이의 간선의 유무가 미리 정해진 확률 p 에 따라 정해지는 Bernoulli 네트워크를 조사대상 네트워크로 정하고 노드의 수를 25개부터 500개까지, 확률 p 는 0.3에서 0.6까지 변화시켜가며 실험할 결과 자아 네트워크를 사용하여 얻어낸 노드 중요도 순위와 전체 네트워크를 사용하여 얻어낸 노드 중요도 순위의 유사성이 최소 85.4%에서 최대 98.1%까지 나타남을 실험을 통하여 입증했다. 즉, 사회관계망의 특성과 약간 거리가 있는 Bernoulli 네트워크에서도 자아 네트워크를 사용한 노드 중요도 측정값이 매우 의미가 있음을 밝힌 것이다. 아울러, 실제 사회관계망에서 도출한 8개의 샘플 네트워크 데이터를 사용하여 실험한 결과 그 유사도는 평균적으로 95% 정도를 나타냄을 보였다. 이렇듯 매개 중심도 분석에 있어서 비교적 구성이 용이한 자아 네트워크의 사용이 노드들 간의 상대적인 중요도를 파악함에 있어서 그 정확성이 매우 높음을 알 수 있다. 또한, 이와 같은 사실은 최근에 이동 네트워크(Mobile Networks)에서 이동 노드들 사이의 메시지 전달 효과를 높이기 위하여 사회관계망 분석 기법을 적용하는 연구들에서도 활용되고 있다[9, 10].

본 논문은 [7]의 결과를 더욱 발전시켜서 자아 네트워크와 그 구성측면에서 구성의 용이성이 동일한 확장 자아 네트워크를 정의하고, 전체 네트워크 기반의 노드 중요도 순위와 더욱 유사성이 높은 중요도 순위를 산출할 수 있는 확장 자아 매개 중심도 (Expanded Ego Betweenness Centrality)를 제안한다.

3. 제안하는 확장 자아 네트워크 및 확장 자아 매개 중심도

3.1 이웃 네트워크와 자아 네트워크

임의의 주어진 네트워크(연결 그래프, Fig. 1A 참조)와 그 네트워크에 속한 임의의 노드 v_i 에 대하여 다음과 같은 기본적인 용어를 우선 정의한다.

- N_i : 노드 v_i 와 직접적으로 연결된 이웃 노드들의 집합
- \dot{E}_{N_i} : 노드 v_i 와 N_i 간의 간선 집합
- \ddot{E}_{N_i} : N_i 에 속한 노드들끼리 연결된 간선 집합

위 세 개의 정의를 사용하여 다음과 같이 “이웃 네트워크 (Neighbor Network, Π_i)”와 “자아 네트워크 (Ego Network, Ω_i)”를 정의할 수 있다(Fig. 1B와 Fig. 1C 참조).

- Π_i : 노드 v_i 와 N_i 의 각 노드들을 노드 집합으로 지니며, 노드 v_i 와 N_i 의 각 노드 사이의 간선들을 간선 집합으로

지니는 네트워크 (즉, $\{v_i\} \cup N_i$ 와 \dot{E}_{N_i} 를 각각 노드 집합 및 간선 집합으로 지니는 네트워크)

- Ω_i : 노드 v_i 와 N_i 의 각 노드들을 노드 집합으로 지니며, 노드 v_i 와 N_i 의 각 노드 사이의 간선들 및 N_i 에 속한 노드들끼리의 간선들을 간선 집합으로 지니는 네트워크 (즉, $\{v_i\} \cup N_i$ 와 $\dot{E}_{N_i} \cup \ddot{E}_{N_i}$ 를 각각 노드 집합 및 간선 집합으로 지니는 네트워크)

임의의 노드 v_i 에 대한 자아 네트워크 Ω_i 는 사회관계망 분석에서 자주 사용되는 모델로서 주어진 네트워크 내에서 v_i 로부터 반경(radius) 1-hop 내에 존재하는 노드들 및 그들 간에 연결된 간선들로 이루어진 네트워크로서 흔히 v_i 를 중심으로 한 정보를 얻어내기 위하여 사용하는 주어진 전체 네트워크의 부분 네트워크이다[7, 9]. 네트워크를 구성하는 노드가 n 개라면 그 네트워크에 대한 서로 다른 자아 네트워크는 총 n 개가 존재한다.

3.2 2차 자아 네트워크

한편, 임의의 노드 v_i 에 대하여 다음과 같은 노드 집합 및 간선 집합을 추가로 정의할 수 있다.

- N_i^2 : 노드 v_i 의 N_i 에 속한 각 노드들이 직접적으로 연결된 이웃 노드들의 집합. 이 때, v_i 는 N_i^2 에서 제외한다 (즉, v_i 에서부터 2-hop 떨어져 있는 노드들의 집합).
- $\dot{E}_{N_i^2}$: N_i 에 속한 각 노드들과 N_i^2 에 속한 노드들 간의 간선 집합
- $\ddot{E}_{N_i^2}$: N_i^2 에 속한 노드들끼리 연결된 간선 집합

위 N_i^2 , $\dot{E}_{N_i^2}$ 및 $\ddot{E}_{N_i^2}$ 정의를 사용하면 자아 네트워크를 구성하는 노드로서 v_i 로부터 반경 1-hop 내에 존재하는 노드들 뿐 아니라 반경 2-hop 내에 존재하는 노드들까지 포함할 수 있으며 간선은 그러한 모든 노드들 사이에 존재하는 간선들을 포함할 수 있다. 이러한 네트워크는 “2차 자아 네트워크 (2-Order Ego Network, Ω_i^2)”라고 칭하며 다음과 같이 정의할 수 있다(Fig. 1D 참조).

- Ω_i^2 : 노드 v_i , N_i , N_i^2 의 각 노드들을 노드 집합으로 지니며, 노드 v_i 와 N_i 에 속한 각 노드 사이의 간선들, N_i 에 속한 노드들끼리의 간선들, N_i 에 속한 노드와 N_i^2 에 속한 노드들 사이의 간선들 및 N_i^2 에 속한 노드들끼리의 간선들을 간선 집합으로 지니는 네트워크 (즉, $\{v_i\} \cup N_i \cup N_i^2$ 와 $\dot{E}_{N_i} \cup \ddot{E}_{N_i} \cup \dot{E}_{N_i^2} \cup \ddot{E}_{N_i^2}$ 를 각각 노드 집합 및 간선 집합으로 지니는 네트워크)

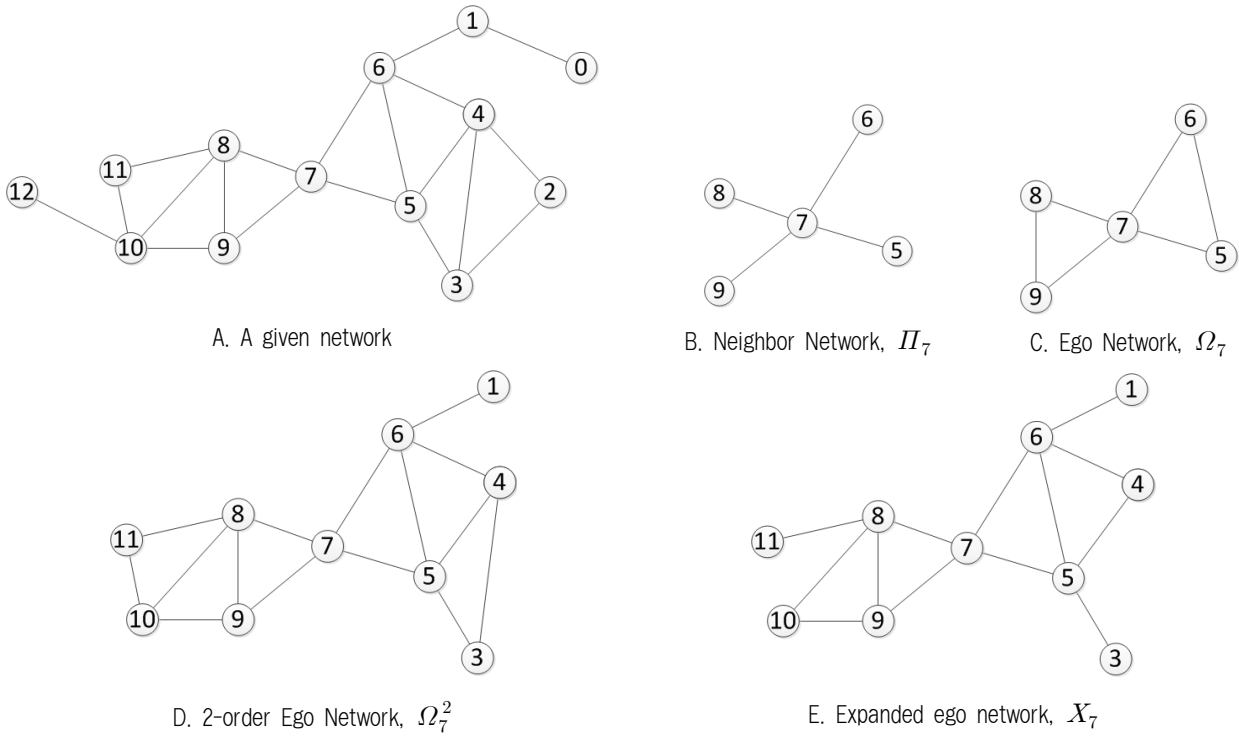


Fig. 1. A given network and the node v_7 's neighbor, ego, 2-order ego, and expanded ego networks

2-Order Ego Network인 Ω_i^2 또한 사회관계망 분석 분야에서 정의된 것으로 임의의 노드 v_i 를 중심으로 더 많은 주변 정보를 모아서 좀 더 의미있는 데이터를 산출하기 위해 사용되어 진다.

3.3 확장 자아 네트워크

본 논문에서 제안하는 임의의 노드 v_i 에 대한 “확장 자아 네트워크 (Expanded Ego Network, X_i)”를 다음과 같이 정의한다(Fig. 1E 참조).

- X_i : 노드 v_i , N_i , N_i^2 의 각 노드들을 노드 집합으로 지니며, 노드 v_i 와 N_i 에 속한 각 노드 사이의 간선들, N_i 에 속한 노드들끼리의 간선들, N_i 에 속한 노드와 N_i^2 에 속한 노드들 사이의 간선들을 간선 집합으로 지니는 네트워크 (즉, $\{v_i\} \cup N_i \cup N_i^2$ 와 $\dot{E}_{N_i} \cup \ddot{E}_{N_i} \cup \dot{E}_{N_i^2}$ 를 각각 노드 집합 및 간선 집합으로 지니는 네트워크)

확장 자아 네트워크인 X_i 와 2차 자아 네트워크인 Ω_i^2 의 차이는 간선집합 내에 N_i^2 에 속한 노드들끼리 연결된 간선 집합, 즉 $\ddot{E}_{N_i^2}$ 의 포함유무이다. 2차 자아 네트워크를 구성하기 위하여 $\ddot{E}_{N_i^2}$ 정보를 얻어오기 위해서는 노드 v_i 입장에서 자신의 이웃의 이웃들이 지닌 간선 정보를 활용해야 하기

때문에 2차 자아 네트워크의 구성은 확장 자아 네트워크 보다 어렵다. 반면, X_i 는 임의의 노드 v_i 에 대해 다음 식 (2)와 같이 정의할 수 있다.

$$X_i = \bigcup_{v_k \in N_i} \Pi_k \tag{2}$$

즉, X_i 를 구성하기 위해 단지 이웃 노드들에 대한 이웃 네트워크를 모두 합치면 된다. 이러한 특성은 2차 자아 네트워크인 Ω_i^2 와 비교하여 X_i 의 구성방법이 쉽다는 것을 설명하여 그 활용성을 높여준다.

예를 들어, 페이스북은 현재 Open API 보안 정책으로서 각 계정 정보를 기반으로 해당 계정 사용자 v_i 에 대한 친구 정보인 Π_i 네트워크만을 그 계정 사용자에게 공개한다. 즉, 임의의 사용자가 다른 사용자에 대한 친구관계 정보 (즉, 식 (2)에서 N_i 에 속한 v_k 의 이웃 네트워크인 Π_k)를 Open API로 획득하기 위해서는 그 다른 사용자들에 대한 허가가 필요하다. 이와 같은 페이스북의 보안정책은 임의의 노드 v_i 에 대한 자아 네트워크 구성을 다소 어렵게 만들며, 2차 자아 네트워크 구성을 위해서는 임의의 사용자를 대표하는 노드 v_i 의 N_i^2 에 속한 노드들끼리 연결된 간선 집합인 $\ddot{E}_{N_i^2}$ 정보가 필요하며 이는 다시 해당 사용자의 친구들에 대한 친구들까지도 그 사용자에게 정보 열람 허가를 해주어야 하기 때문에 2차 자아 네트워크 구성은 더욱더 힘들게 만

들다. 여기서 주목할 점은 만약 임의의 사용자의 친구들 모두가 자신의 정보 공개를 그 사용자에게 허용한다면 그 사용자를 중심으로 한 자아 네트워크 구성이 가능하며 또한 본 논문에서 제안하는 확장 자아 네트워크인 X_i 도 쉽게 구성할 수 있다는 점이다. 즉, 자아 네트워크와 확장 자아 네트워크는 구성의 용이성이 동일하다.

또한, 무선 이동 네트워크 분야에서 이동 애드혹 네트워크(Mobile Ad-hoc Network)나 지연 감내 네트워크(Delay-Tolerant Network, DTN)에 사회관계망 분석을 활용하여 메시지 전달 효율을 높이려는 시도가 최근 연구 중이다[9-12]. 이러한 연구들에서는 네트워크를 구성하는 각 이동 노드들 사이의 직접적인 무선 연결 정보를 시간 순서대로 관리하여 연결을 자주 확립했던 노드들 사이에 사회적 관계(Social Relation)를 부여 한다. 이후 사회관계망 분석기법을 활용하여 네트워크 내에 중요한 역할을 하는 이동 노드를 선별하는 작업이나 동질성(Common features)을 나타내는 이동 노드들을 하나의 커뮤니티(Community)로 구성하는 작업을 시도할 수 있다. 이 때 지리적으로 분산되어 있는 이동 노드들의 연결 관계 정보를 하나의 중앙 집중화된 서버에 기록하기가 매우 어렵고 설령 가능하더라도 중앙에서 분석된 결과를 분산되어 이동하고 있는 노드들에게 전파시키는 작업도 어렵기 때문에 대개 각 노드들이 자신의 지역 정보, 즉 자신과 현재 통신이 가능한 이웃 노드들의 정보만을 활용하여 자율적으로 자아 네트워크나 확장 자아 네트워크를 구성하고 이를 바탕으로 의미있는 분석 정보를 산출하는 시도를 하고 있다. 예를 들어, 임의의 노드들이 모두 자신의 식별자(Identity)를 포함한 Hello 메시지 내에 자신의 이웃 노드 정보를 넣어 주기적으로 방송(Broadcast)한다고 하면, 다른 노드들은 무선 연결이 가능한 이웃 노드들로부터 그러한 Hello 메시지를 받아서 그 안에 있는 이웃 노드들의 정보를 활용하여 자아 네트워크를 구성하던지 아니면 식 (2)를 활용하여 확장 자아 네트워크를 쉽게 구성할 수 있다. 하지만, 이러한 경우에 Hello 메시지 내에 이웃 노드의 이웃 노드 정보까지 넣어 보내지 않는 이상 2차 자아 네트워크 구성은 매우 어렵다.

즉, 본 논문에서 제안하는 확장 자아 네트워크 X_i 는 기존의 자아 네트워크 Ω_i 와 비교하여 동일한 구성의 용이성

을 가지면서도 자아 네트워크 Ω_i 보다 더욱 많은 지역 정보를 가지고 있기 때문에 사회 관계망 분석에 대한 정확성을 좀 더 높여줄 수 있다.

3.4 자아 매개 중심도 및 확장 자아 매개 중심도

주어진 전체 네트워크의 노드 개수가 n 개 일 때 3.1절에서 정의한 자아 네트워크 Ω_i 는 각 노드 당 생성될 수 있으므로 총 n 개가 도출되며, 3.3절에서 정의한 확장 자아 네트워크도 같은 개념으로 총 n 개가 도출된다. 이 때 임의의 노드 v_i 의 자아 매개 중심도 (Ego Betweenness Centrality, $C_B^\Omega(i)$)와 확장 자아 매개 중심도 (Expanded Ego Betweenness Centrality, $C_B^X(i)$)는 각각 다음과 같이 정의할 수 있다.

$$C_B^\Omega(i) = \frac{\sum_{v_s \neq v_i, v_i \in V_\Omega, s < t} \frac{\sigma_{st}(i)}{\sigma_{st}}}{(n_\Omega - 1)(n_\Omega - 2)/2} \tag{3}$$

$$C_B^X(i) = \frac{\sum_{v_s \neq v_i, v_i \in V_X, s < t} \frac{\sigma_{st}(i)}{\sigma_{st}}}{(n_X - 1)(n_X - 2)/2} \tag{4}$$

위 식에서 V_Ω 및 n_Ω 는 v_i 의 자아 네트워크 Ω_i 의 노드 집합과 노드의 개수를 각각 나타내며, V_X 및 n_X 는 v_i 의 확장 자아 네트워크 X_i 의 노드 집합과 노드의 개수를 각각 나타낸다.

식 (3)과 식 (4)에서 볼 수 있듯이 자아 매개 중심도와 확장 자아 매개 중심도는 식 자체의 구성이 기존의 매개 중심도와 유사하여 그 계산 복잡도가 동일한 것처럼 보이지만 사실 다음과 같은 두 가지 측면에서 계산 복잡도가 매우 낮다:

- 전체 네트워크의 노드의 수 및 간선의 수가 매우 많아도 자아 네트워크 Ω_i 및 확장 자아 네트워크 X_i 의 노드의 수 및 간선의 수는 상대적으로 작다.

Table 1. Comparison of 1) Betweenness Centrality, 2) Ego Betweenness Centrality, and 3) Expanded Ego Betweenness Centrality for the given network of Fig. 1

		0	1	2	3	4	5	6	7	8	9	10	11	12
Betweenness Centrality	Value	0.00	0.16	0.00	0.03	0.13	0.21	0.37	0.53	0.25	0.12	0.17	0.00	0.00
	Rank	10	6	10	9	7	4	2	1	3	8	5	10	10
Ego Betweenness Centrality	Value	0.00	1.00	0.00	0.16	0.33	0.33	0.58	0.66	0.33	0.16	0.58	0.00	0.00
	Rank	9	1	9	8	5	5	3	2	5	8	3	9	9
Expanded Ego Betweenness Centrality	Value	0.00	0.40	0.00	0.10	0.27	0.26	0.47	0.55	0.30	0.14	0.45	0.00	0.00
	Rank	10	4	10	9	6	7	2	1	5	8	3	10	10

- 각 노드마다의 $C_B^O(i)$ 와 $C_B^X(i)$ 계산 과정은 서로 독립적이기 때문에 각 노드의 자아 매개 중심도 및 확장 자아 매개 중심도는 병렬적으로 동시 계산이 가능하다.

위와 같은 특징 때문에 노드의 중요도를 체크하는 데 있어서 자아 매개 중심도 및 확장 자아 매개 중심도가 전통적인 방식에 의하여 전체 네트워크를 기준으로 산출한 매개 중심도와 거의 유사하다면 그러한 전통적인 매개 중심도에 비해 활용 가치가 높다고 할 수 있다.

Table 1은 Fig. 1에서 주어진 전체 네트워크에서 각 노드들에 대한 1) 매개 중심도, 2) 자아 매개 중심도, 3) 확장 자아 매개 중심도 및 상대적인 순위(Rank)를 비교하여 보여준다. 각 매개 중심도들에 대한 값 자체는 차이가 있지만 상대적인 순위를 비교해 보면 본 논문에서 제안하는 확장 자아 매개 중심도가 자아 매개 중심도보다 기존 매개 중심도와 더욱 유사함을 알 수 있다.

4. 확장 자아 매개 중심도의 비교 분석

본 장에서는 주어진 전체 네트워크에 대하여 다음과 같은 세 개의 데이터 리스트에 대한 순위 측면에서의 유사성 비교 분석을 수행한다.

- 전체 네트워크에서 산출한 매개 중심도의 각 노드별 $C_B(i)$ 값 리스트
- 자아 네트워크에서 산출한 자아 매개 중심도의 각 노드별 $C_B^O(i)$ 값 리스트
- 확장 자아 네트워크에서 산출한 확장 자아 매개 중심도의 각 노드별 $C_B^X(i)$ 값 리스트

이 때, $C_B(i)$ 값 리스트가 전체 네트워크의 모든 노드 및 모든 간선에 대해 종합적인 분석을 통해 얻은 값이기 때문에 가장 정확한 데이터라고 할 수 있으며, 결국 1) 각 노드별 $C_B(i)$ 값과 $C_B^O(i)$ 값의 순위 유사성 및 2) 각 노드별 $C_B(i)$ 값과 $C_B^X(i)$ 값의 순위 유사성을 비교하여 확장 자아 매개 중심도가 기존의 자아 매개 중심도보다 더 정확성이 높음을 밝힌다. 사회관계망 분석에서는 각 노드의 중심도를 활용할 때 대부분의 경우 매개 중심도의 절대값 자체보다는, 상대적인 크기 비교에 더 주안점을 둔다. 즉, 네트워크 내에 임의의 한 노드가 다른 노드들 보다 중심도가 높은지 정도를 판단하여 높은 중심도를 지닌 노드를 메시지 전달의 매개체로 활용하는 경우가 대부분이다.

본 실험을 위하여 Python 2.7과 NetworkX, Scipy, Numpy 등의 Python 모듈을 활용하였으며 iMAC (3.4Ghz Intel Core i7, 16GB 1600Mhz DDR3) OS X (Version 10.8.2)에서 본 실험을 수행하였다.

4.1 조사 대상 네트워크 및 유사성 분석 방법

확장 자아 매개 중심도의 정확성 실험의 조사 대상 네트워크를 생성하기 위하여 본 논문에서는 Barabási-Albert 네트워크 [13]라고 알려진 척도 없는(Scale-free) 네트워크 생성 방법에 노드간의 클러스터링(Clustering) 특성을 부여한 네트워크 생성 모델 [14]을 사용한다. Barabási-Albert의 척도 없는 네트워크를 생성하기 위해서는 우선 처음 1개 또는 그 이상의 노드를 랜덤하게 생성하고, 새로운 노드를 추가할 때 기존에 이미 생성되어 있는 노드들과 λ 개의 간선을 생성한다. 이 때, 기존에 이미 생성되어 있는 어떤 노드와 간선이 생성될 지는 기존 노드들이 지닌 이웃 수에 비례한다. 이렇게 간선이 생성되는 것을 “선호적 연결(preferential attachment)”이라고 지칭하며, 이와 같은 선호적 연결은 노드들의 이웃노드 수 분포가 멱급수(Power-law) 함수를 따르도록 만들어 주어 매우 많은 이웃 노드 수를 지닌 허브 역할을 담당하는 노드가 등장할 수 있고, 대다수의 노드는 이웃 노드의 수가 1이나 2가 된다. 또한, 네트워크에 참여되는 노드의 수가 매우 많이 늘어나도 임의의 두 노드 사이의 경로 길이가 매우 짧다. 이러한 특징은 좁은 세상(Small world)이라고 지칭한다[13]. 한편, 선호적 연결 과정을 통하여 새로운 노드 v 가 기존 노드 w 와 간선을 지니게 된다면, 확률 p_t 에 따라서 노드 v 는 추가적으로 기존 노드 w 의 이웃 노드 중 한 노드와 간선을 생성한다. 이와 같은 추가적인 간선은 결국 세 개의 노드들끼리 Triad를 형성하여 여러 Triad가 인접하게 되면 결국 몇몇의 노드들끼리 상호간에 많은 연결성을 지닌 클러스터를 형성하도록 만들어준다[14].

이와 같은 특징들 때문에 본 논문에서 조사대상으로서 생성되는 네트워크들은 사회관계망과 매우 유사한 형태를 지닌다. 즉, 트위터, 페이스북, 링크드인의 친구관계를 분석해 보면 노드들의 이웃노드 수 분포가 멱급수 함수를 따름이 이미 여러 문헌에서 보고되었다[3, 15]. 그래서, 다수의 사용자들은 친구관계를 매우 적게 형성하고 있지만 때로는 매우 많은 수의 친구관계를 형성하여 친구관계의 허브 역할을 담당하는 사용자도 발견할 수 있다. 또한, 몇몇의 사용자들끼리는 커뮤니티를 만들어서 커뮤니티 내부 사용자들끼리는 친구관계를 상호간에 형성하게 된다. 본 논문에서는 이와 같이 실제 사회관계망을 모델링하는 네트워크를 랜덤하게 생성하였으며 노드의 수 n 은 50, 200, 500, 1000개로 나누어 실험하였고, 새로운 노드를 추가할 때 기존에 이미 생성되어 있는 노드들과 생성하는 간선의 수 λ 는 1로 설정하였다. 마지막으로 Triad를 형성하는 추가 간선이 생성되는 확률인 p_t 는 0.1, 0.5, 0.9로 나누어 설정하였다. Fig. 2는 조사 대상 네트워크들 중 $n = 50, \lambda = 1, p_t = 0.1$ 일 때 만들어진 네트워크를 보여준다.

한편, 각 노드별 $C_B(i)$ 값과 $C_B^O(i)$ 값의 순위 유사성 및 $C_B(i)$ 값과 $C_B^X(i)$ 값의 순위 유사성을 얻기 위하여 스피어만 상관계수(Spearman correlation coefficient)를 사용하였다. 스피어만 상관계수는 데이터 자료 자체의 값 대신

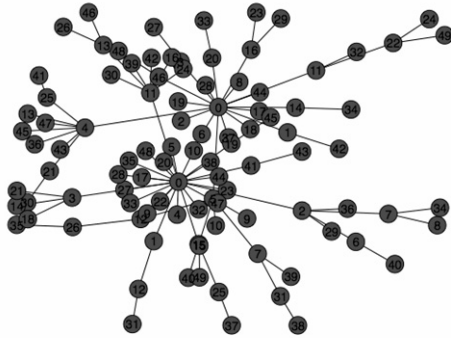


Fig. 2. An inspected network sample (Barabási-Albert network with clustering, $n = 50, \lambda = 1, p_t = 0.1$)

순위(rank)에 대한 데이터 상관계수를 구할 때 흔히 사용되는 것으로서, 두 개의 서로 다른 데이터 집합에 대해 데이터 값이 작은 것부터 차례로 순위를 매겨 서열 순서로 바꾼 뒤 순위를 이용해 두 데이터 집합 사이의 상관계수를 구한다. 스피어만 상관계수는 -1.0과 1.0 사이의 값을 가지는데 두 데이터 집합안의 데이터 순위가 완전히 일치하면 1.0이고, 그 순위가 완전히 반대이면 -1.0이 된다.

4.2 실험 결과

Table 2는 각 파라미터별로 100번의 실험을 수행한 결과 얻어낸 스피어만 상관계수의 평균값(Avg. Corr.)과 표준편차(S.D.)값을 보여준다. 노드의 수 n 이나 추가 간선이 생성 확률 p_t 의 변화에 따라 특별한 성향이 보이진 않고 모든 경우에 있어서 $C_B(i)$ 값과 $C_B^Q(i)$ 값의 순위 유사성과 $C_B(i)$ 값과 $C_B^X(i)$ 값의 순위 유사성이 매우 높으며, 표

Table 2. Experimental results (Avg. Corr. = Average Correlation, S.D. = Standard Deviation)

		n	50	200	500	1000
$p_t = 0.1$	$C_B(i) \leftrightarrow C_B^Q(i)$	Avg. Corr.	0.972	0.974	0.973	0.972
		S.D.	0.008	0.003	0.002	0.002
	$C_B(i) \leftrightarrow C_B^X(i)$	Avg. Corr.	0.983	0.982	0.985	0.983
		S.D.	0.012	0.004	0.004	0.004
$p_t = 0.5$	$C_B(i) \leftrightarrow C_B^Q(i)$	Avg. Corr.	0.978	0.973	0.974	0.973
		S.D.	0.012	0.003	0.001	0.001
	$C_B(i) \leftrightarrow C_B^X(i)$	Avg. Corr.	0.987	0.983	0.982	0.983
		S.D.	0.005	0.004	0.001	0.004
$p_t = 0.9$	$C_B(i) \leftrightarrow C_B^Q(i)$	Avg. Corr.	0.974	0.973	0.973	0.974
		S.D.	0.008	0.003	0.002	0.005
	$C_B(i) \leftrightarrow C_B^X(i)$	Avg. Corr.	0.984	0.985	0.985	0.982
		S.D.	0.006	0.007	0.005	0.003

준편차 값도 매우 작다는 것을 알 수 있다. 게다가, $C_B(i)$ 값과 $C_B^X(i)$ 값의 순위 유사성이 $C_B(i)$ 값과 $C_B^Q(i)$ 값의 순위 유사성보다 약 1% 정도 높다. 즉, 확장 자아 매개 중심도가 자아 매개 중심도 보다 좀 더 기존의 매개 중심도에 가깝다고 할 수 있다. 1% 정도의 차이는 작을 수 있어도 그 두 개의 매개 중심도를 구하기 위한 자아 네트워크 및 확장 자아 네트워크에 대한 구성의 용이성이 비슷하기 때문에 가능하면 확장 자아 매개 중심도를 사용하는 것이 각 노드의 보다 정확한 중심도를 파악하는 데 도움이 된다.

5. 결론

본 논문은 사회관계망 분석에 있어서 전통적으로 많이 활용되어 온 각 노드의 매개 중심도 값을 보다 빠르게 산출하기 위하여 각 노드에 대한 지역 정보를 활용한 확장 자아 네트워크를 새롭게 정의하고 그 확장 자아 네트워크를 활용한 확장 자아 매개 중심도를 산출하는 방법을 제안하였다. 또한 확장 자아 매개 중심도의 노드별 상대적 순위가 기존의 전통적인 방법으로 구한 매개 중심도의 노드별 상대적 순위와 거의 차이가 없음을 실험을 통하여 증명하였다. 실험 결과의 가치를 높이기 위하여 일반적인 사회관계망이 지닌 특성인 척도 없는 네트워크를 생성할 수 있는 Barabási-Albert 네트워크 모델을 활용하였다. 확장 자아 네트워크의 구성과 이에 대한 확장 자아 매개 중심도 산출은 각 노드별로 독립적으로 수행이 가능하기 때문에, 네트워크의 각 노드별로 병렬적인 동시 계산이 가능하다. 그러므로, 본 논문에서 제안한 확장 자아 매개 중심도는 대규모 사회관계망에서 상대적인 노드의 중요도를 파악하는데 많은 도움이 될 것으로 기대된다. 또한, 무선 이동 네트워크 분야에서 사회관계망 분석기법을 도입할 때 지리적으로 분산되어 있는 이동 노드들의 연결 관계 정보를 하나의 중앙 집중화된 서버에 기록하기가 어렵기 때문에 각 노드들이 자율적으로 확장 자아 네트워크를 구성하고 각자 자신의 확장 자아 매개 중심도를 파악하여 인접한 다른 노드들과 중요도를 상대적으로 비교하는 데에도 활용될 수 있다. 본 논문의 연구 결과를 더욱 확장하여 향후에는 각 노드의 지역 정보만을 활용하여 자신과 사회적 관계를 자주 형성한 노드들과의 커뮤니티(Community) 구성방법에 대해 연구할 예정이다.

참고 문헌

[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] R. Pittinger, "Linkbar and Forensik: Two Systems for Interactive Visualization of Online Social Networks," University of Applied Sciences Augsburg, 2007.

[3] 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.

[4] S. Wasserman and K. Faust. "Social Network Analysis: Methods and Applications," Cambridge University Press, 1994.

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

[6] M. Baglioni, F. Geraci, M. Pellegrini and E. Lastres, "Fast Exact Computation of Betweenness Centrality in Social Networks," in Proceedings of IEEE/ACM International Conference on Advances in Social Networks Analysis and Mining, pp.450-456, 2012.

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

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

[9] E. Daly and M. Haahr, "Social Network Analysis for Information Flow in Disconnected Delay-Tolerant MANETs," IEEE Trans. on Mobile Computing, Vol.8, No.5, pp.606 - 621, May, 2009.

[10] Y.-h. Kim, C.-M. Kim, Y.-H. Han, Y.-S. Jeong and D.-S. Park, "Betweenness of Expanded Ego Networks in Sociality-Aware Delay Tolerant Networks," Ubiquitous Information Technologies and Applications (LNEE, Proc. of CUTE 2012), Springer, pp.499-505, Dec., 2012.

[11] A. Mtibaa, M. May, M. Ammar, and C. Diot. "PeopleRank: combining social and contact information for opportunistic forwarding," In Proceedings of INFOCOM, 2010.

[12] F. Fabbri and R. Verdone, "A sociability-based routing scheme for delay-tolerant networks," EURASIP Journal on Wireless Communications and Networking, Jan., 2011.

[13] A. L. Barabási and R. Albert "Emergence of scaling in random networks," Science 286, pp.509-512, 1999.

[14] P. Holme and B. J. Kim, "Growing scale-free networks with tunable clustering," Physics Review. E, 65, 026107, 2002.

[15] B. A. Huberman, D. M. Romero, and F. Wu. "Social networks that matter: Twitter under the microscope," First Monday, Vol.14, No.1, 2009.



손진곤

e-mail : jgshon@knou.ac.kr
 1991년 고려대학교 전산학전공(이학박사)
 1991년~현 재 한국방송통신대학교
 컴퓨터학과 교수
 1997년~1998년 State University of
 New York (Stony Brook)
 Visiting Professor

2000년~현 재 ISO/IEC JTC1/SC36 Korea Delegate
 2010년 한국정보처리학회 부회장
 2009년~현 재 이러닝학회 부회장
 관심분야: 컴퓨터통신망, 분산시스템, 그리드 컴퓨팅,
 e-Learning, 정보기술 표준화



김용환

e-mail : cherish@koreatech.ac.kr
 2010년 한국기술교육대학교 인터넷미디어
 공학부(공학석사, 학석사 통합과정)
 2010년~현 재 한국기술교육대학교
 컴퓨터공학부 박사과정
 관심분야: 이동 컴퓨팅, 센서 네트워크,
 사회관계망 분석



한연희

e-mail : yhhan@koreatech.ac.kr
 2002년 고려대학교 컴퓨터학과(이학박사)
 2002년~2006년 삼성중합기술원 전문연구원
 2006년~현 재 한국기술교육대학교
 컴퓨터공학부 부교수
 2005년~현 재 한국정보통신기술협회
 (TTA) IT 국제 표준화 전문가

관심분야: 이동 컴퓨팅, 센서 네트워크, 사회관계망 분석