

An Efficient Algorithm for Betweenness Centrality Estimation in Social Networks

Soo-Jin Shin[†] · Yong-hwan Kim^{**} · Chan-Myung 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. 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 our past study, we defined 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 defined a new measure, called the expanded ego betweenness centrality. In this paper, We propose algorithm that quickly computes expanded ego betweenness centrality by exploiting structural properties of expanded ego network. Through the experiment with virtual network used Barabási-Albert network model to represent the generic social network and facebook network to represent actual social network, We show that the node's importance rank based on the expanded ego betweenness centrality has high similarity with that the node's importance rank based on the existing betweenness centrality. We also show that the proposed algorithm computes the expanded ego betweenness centrality quickly than existing algorithm.

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

사회관계망에서 매개 중심도 추정을 위한 효율적인 알고리즘

신수진[†] · 김용환^{**} · 김찬명^{***} · 한연희^{****}

요약

사회관계망 분석에 있어서 매개 중심도(Betweenness Centrality)는 네트워크를 구성하는 노드들의 상대적인 중요도를 파악하기 위한 척도로서 주로 사용되어 왔다. 그러나 매개 중심도를 측정하기 위한 시간 복잡도가 높기 때문에 대규모의 온라인 사회관계망 서비스에서 각 노드의 매개 중심도를 산출하는 것은 쉽지 않은 문제이다. 그래서 본 연구팀에서는 과거에 네트워크를 구성하는 각각의 노드들마다 자신의 지역 정보를 활용하여 확장 자아 네트워크(Expanded Ego Network)를 정의하고 그 네트워크에서 확장 자아 매개 중심도(Expanded Ego Betweenness)를 산출하여 기존의 매개 중심도를 대체하려는 시도를 하였다. 본 논문에서는 지역정보 기반의 확장 자아 네트워크의 특징을 분석하여 확장 자아 매개 중심도를 빠르게 산출할 수 있는 알고리즘을 제안한다. 그리고 일반적인 사회관계망의 특징을 대표하는 Barabási-Albert 네트워크 모델을 사용한 가상 네트워크와 실제 사회관계망을 대표하는 페이스북 친구 관계 네트워크에서의 실험을 통하여 확장 자아 매개 중심도의 중요도 순위가 기존 매개 중심도의 중요도 순위와 거의 일치함을 보인다. 또한 제안하는 알고리즘이 기존 알고리즘에 비해 확장 자아 네트워크에서의 확장 자아 매개 중심도를 더 빠르게 산출함을 보인다.

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

1. 서론

사회관계망(Social Network)은 사회학에서 개인 또는 집단이 네트워크의 하나의 노드(node)를 이루며 이들 간의 상호적인 관계(tie)에 의해 만들어지는 논리적 네트워크를 말한다 [1]. 최근 페이스북(Facebook), 트위터(Twitter)와 같이 사회관계망을 기반으로 하는 웹서비스가 계속해서 증가하는 추세이며, 이러한 웹서비스들은 사용자들에게 다른 사용자들과 관계를 맺고, 관심사를 공유하고, 서로 소통할 수 있도록 도와줌으로써 사람들 간의 인맥형성과 정보교환을 위한

※ 이 논문은 2013년도 한국기술교육대학교 교수해외 파견연구비 지원에 의하여 연구되었으며, 또한 2013년도 정부(교육부)의 재원으로 한국연구재단의 지원을 받아 수행된 기초연구사업 (No. NRF-2013R1A1A2010050)에 의해 지원받았음.

[†] 준회원 : 한국기술교육대학교 컴퓨터공학부 석사과정

^{**} 정회원 : 한국기술교육대학교 컴퓨터공학부 박사과정

^{***} 준회원 : 한국기술교육대학교 컴퓨터공학부 박사과정

^{****} 종신회원 : 한국기술교육대학교 컴퓨터공학부 교수

Manuscript Received : October 14, 2014

First Revision : December 8, 2014

Accepted : December 8, 2014

* Corresponding Author : Youn-Hee Han(yhhan@koreatech.ac.kr)

새로운 플랫폼을 제시하고 있다 [2]. 이러한 새로운 플랫폼과 더불어 스마트 기기들이 보급화 되면서 사회관계망을 분석하여 의미 있는 정보를 추출하고 이를 다양한 방면에서 기업 마케팅 및 사회 현상 분석에 활용하는 등 이의 효용 가치 또한 늘어나고 있다 [3].

한편, 사회관계망 분석은 사람들 사이의 관계를 다방면에서 해석하기 위해 노드와 링크로 연결된 사회 그래프를 생성하고 이를 분석하는 데 활용되는 모든 분석 방법들을 통칭한다. 이를 통해 사회관계망 서비스에 참여하고 있는 각 사용자들 간의 관계(relationship)와, 이들 관계의 패턴과 함의를 파악할 수 있다 [4].

다양한 사회관계망 분석 기법 중에서 노드 중심도(Centrality) 분석은 네트워크의 존재하는 노드들 중에서 중요도가 높은 노드들을 선별하기 위하여 사용한다. 대표적으로 연결 중심도(Degree Centrality), 거리 중심도(Closeness Centrality), 매개 중심도(Betweenness Centrality), 고유벡터 중심도(Eigenvector Centrality) 분석 기법들이 존재한다.

특히 사회관계망 분석 기법 중 하나인 매개 중심도 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[4]. 그러나 매개 중심도를 구하기 위해서는 각 노드 사이의 최단경로를 구하는 작업을 수행해야 하는데, 이의 계산은 시간복잡도가 높다고 알려져 있다. 따라서 사회관계망 서비스를 사용하는 개인 및 집단의 증대에 따른 대규모 사회관계망에서 각 노드의 매개 중심도를 산출하는 작업은 매우 어렵다. 이와 같은 문제점을 해결하기 위해 사회관계망에서 빠른 시간 내에 모든 노드들에 대한 매개 중심도를 추정해내는 방법에 관한 연구가 요구된다.

본 논문은 빠른 시간 내에 매개 중심도를 추정해 내기 위하여 사회관계망을 구성하는 각 노드의 이웃 노드의 친구 관계 정보만을 활용하여 구성할 수 있는 “확장 자아 네트워크(Expanded Ego Network)” 개념을 사용한다. 확장 자아 네트워크는 [5]에서 제안된 “자아 네트워크(Ego Network)”를 기반으로 하여 본 연구팀의 선행 연구 [6]에서 정의한 것으로서, 포함하는 지역 정보의 양이 자아 네트워크보다 다소 많기 때문에 노드 중심도 분석의 정확도가 높은 편이다. 또한 [6]에서는 확장 자아 네트워크에서의 자아 노드(Ego Node)가 지니는 매개 중심도를 “확장 자아 매개 중심도(Expanded Ego Betweenness Centrality)”라 정의하였다. 본 논문에서는 확장 자아 네트워크의 특성을 면밀하게 분석하여 기존의 알고리즘보다 효율적으로 확장 자아 매개 중심도를 산출할 수 있는 알고리즘을 제시한다. 그리고 일반적인 사회관계망의 특성인 척도 없는 네트워크(Scale-free Network)를 생성할 수 있는 Barabási-Albert 네트워크 모델을 사용한 가상 네트워크와 현재 상용 중인 사회관계망 서비스인 페이스북의 친구 관계 데이터에 기반한 실제 네트워크에서의 실험을 통하여 확장 자아 매개 중심도의 효율성을 입증한다. 마지막으로, 기존에 잘 알려진 매개 중심도를 산출하는 Brandes 알고리즘에 비해 제안 알고리즘이 확장 자아 네트워크에서의 확장 자아 매개 중심도를 더 빠르게 산출함을 보인다.

본 논문의 구성은 다음과 같다. 2절에서는 매개 중심도에 대한 설명 및 매개 중심도를 빠르게 측정하기 위한 기존 연구들에 대해 소개하고 자아 네트워크와 본 연구팀의 선행 연구에서 정의한 확장 자아 네트워크에 대해 설명한다. 3절에서는 자아 매개 중심도와 확장 자아 매개 중심도에 대하여 자세히 설명한다. 4절에서는 확장 자아 네트워크의 특성을 활용하여 매개 중심도를 측정하는 알고리즘을 제안한다. 5절에서는 실험을 통해 확장 자아 매개 중심도와 기존의 매개 중심도의 유사성을 보이고, 제안하는 알고리즘과 기존의 알고리즘을 비교분석한다. 마지막으로 6절에서는 본 논문의 결론 및 향후 연구에 대해 설명한다.

2. 관련 연구

2.1 매개 중심도

본 논문에서 활용하는 사회관계망 분석 척도인 임의의 노드 v 의 매개 중심도 (Betweenness Centrality, $B(v)$)는 다음과 같이 정의된다 [4].

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

- V : 모든 노드의 집합
- σ_{st} : 노드 s 와 t 사이의 최단경로의 개수
- $\sigma_{st}(v)$: 노드 s 와 t 사이의 최단경로 중에서 노드 v 를 통과하는 경로의 개수

임의의 노드 간의 최단경로를 구하기 위한 시간 복잡도는 기본적으로 $O(n^2)$ 정도이기 때문에 노드의 개수가 많아지면 시간이 매우 길어지고 실용성이 매우 낮아진다. 그래서, Brandes은 매개 중심도를 빠르게 측정하기 위해 [7]에서 노드의 개수가 n 이고 간선의 개수가 m 일 때 복잡도가 $O(nm)$ 인 대수적 기법을 제안하였다. 즉, 간선의 개수가 노드의 개수보다 작은 희소 네트워크(Sparse Network)인 경우 좀 더 빠르게 각 노드의 매개 중심도를 측정할 수 있다. 또한 Baglioni는 [8]에서 제안한 방법을 더욱 발전시켜 2배 정도 복잡도가 개선된 매개 중심도를 측정 방안을 제시했다.

한편, Everett와 Borgatti은 [5]에서 기존 사회관계망 분석에서 오랫동안 사용해 온 “자아 네트워크”를 각 노드마다 구성하고 그러한 각각의 자아 네트워크로부터 매개 중심도를 측정하여 얻어낸 각 노드들의 상대적인 중요도 순위가 전통적인 방법으로 전체 네트워크에서 각 노드의 매개 중심도를 측정하여 얻어낸 각 노드들의 중요도 순위와 매우 유사함을 실험을 통해 증명했다. 임의의 노드에 대한 자아 네트워크는 그 노드와 변경 1-hop 내에 존재하는 노드들, 그리고 노드 간의 간선들로 이루어진 네트워크로서 구성이 매우 쉬우면서 기존 네트워크의 특징을 잘 반영하여 통계사회학에서 자주 활용되어왔다.

본 논문의 선행 연구에서는 [5]의 결과를 발전시켜서 자아 네트워크와 그 구성측면에서 구성의 용이성이 동일한 확장 자아 네트워크를 정의하고, 확장 자아 네트워크에서 매개 중심도를 측정한 확장 자아 매개 중심도를 제안하였다 [6].

본 논문에서는 이러한 확장 자아 매개 중심도의 중요도

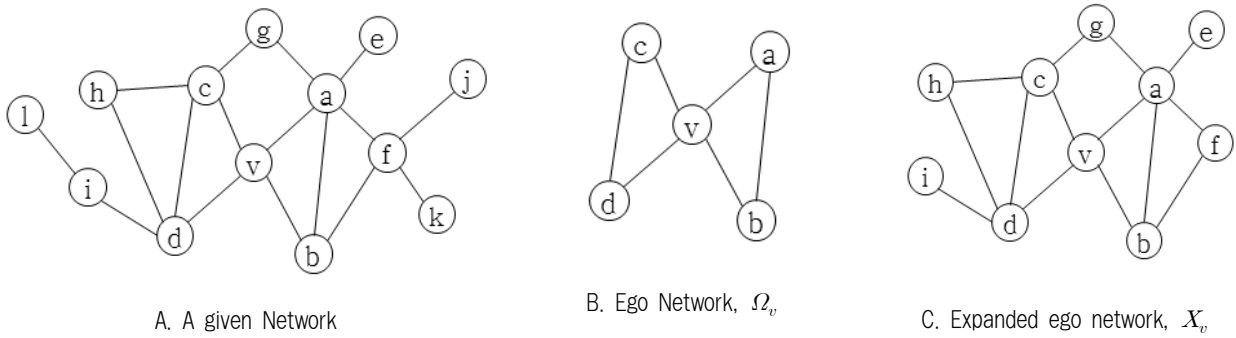


Fig. 1. A given network and the node v 's ego and expanded ego network

순위와 자아 매개 중심도의 중요도 순위를 산출하여 각각 전체 네트워크 매개 중심도의 중요도 순위와 얼마나 유사한 지 스피어만 상관계수(Spearman Coefficient)를 계산하여 상관관계를 비교 분석한다. 또한 확장 자아 네트워크의 특성을 분석하여 확장 자아 네트워크에서 기존의 매개 중심도를 도출하는 Brandes 알고리즘보다 보다 빠르게 매개 중심도를 계산할 수 있는 알고리즘을 제안한다.

2.2 자아 네트워크

자아 네트워크를 정의하기 위하여 주어진 사회관계망 그래프 $G(V, E)$ 와 그 네트워크에 속한 임의의 노드 v 에 대하여 다음과 같은 기본적인 용어를 우선 정의한다.

- $V_{v:1}$: 노드 v 와 직접적으로 연결된 이웃 노드들의 집합 (즉, v 에서부터 1-hop 떨어져 있는 노드들의 집합)
- $E_{v:0:1}$: 노드 v 와 v 의 1-hop 노드들의 집합($V_{v:1}$) 사이의 간선들의 집합
- $E_{v:1}$: $V_{v:1}$ 의 노드 간의 간선들의 집합

위의 용어를 사용하여 “자아 네트워크(Ego Network, Ω_v)”를 다음과 같이 정의할 수 있다(Fig. 1(B) 참조).

[정의 1] 자아 네트워크 (Ego Network, Ω_v): 노드 v 와 $V_{v:1}$ 의 각 노드들을 노드 집합으로 지니며, 노드 v 와 $V_{v:1}$ 의 각 노드 사이의 간선들($E_{v:0:1}$) 및 $V_{v:1}$ 에 속한 노드들끼리의 간선들($E_{v:1}$)을 간선 집합으로 지니는 네트워크

- $\Omega_v : (\{v\} \cup V_{v:1}, E_{v:0:1} \cup E_{v:1})$

임의의 노드 v 에 대한 자아 네트워크 Ω_v 는 사회관계망 분석에서 자주 사용되는 모델로서 흔히 v 를 중심으로 한 정보를 얻어내기 위하여 사용하는 주어진 전체 네트워크의 부분 네트워크이다 [5].

3. 확장 자아 네트워크와 확장 자아 매개 중심도

확장 자아 네트워크를 정의하기 위하여 주어진 사회관계망 그래프 $G(V, E)$ 와 그 네트워크에 속한 임의의 노드 v 에

대하여 추가적으로 다음과 같은 용어를 정의한다.

- $V_{v:2}$: v 에서부터 2-hop 떨어져 있는 노드들의 집합
- $E_{v:1:2}$: 노드 v 의 1-hop 노드들의 집합($V_{v:1}$)과 노드 v 의 2-hop 노드들의 집합($V_{v:2}$)사이의 간선들의 집합

확장 자아 네트워크 X_v 에서는 자아 네트워크 Ω_v 에 비하여 더 많은 주변 정보를 활용하기 때문에 좀 더 의미 있는 데이터를 산출할 수 있다.

위의 용어를 사용하여 “확장 자아 네트워크 (Expanded Ego Network, X_v)”를 다음과 같이 정의 할 수 있다. (Fig. 1(C) 참조).

[정의 2] 확장 자아 네트워크 (Expanded Ego Network, X_v): 노드 v , $V_{v:1}$, $V_{v:2}$ 의 각 노드들을 노드 집합으로 지니며, 노드 v 와 $V_{v:1}$ 의 각 노드 사이의 간선들($E_{v:0:1}$) 및 $V_{v:1}$ 에 속한 노드들끼리의 간선들($E_{v:1}$), $V_{v:1}$ 에 속한 노드와 $V_{v:2}$ 에 속한 노드들 사이의 간선들을 간선 집합으로 지니는 네트워크

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

주어진 전체 네트워크에서 노드의 개수가 n 개일 때 자아 네트워크와 확장 자아 네트워크는 각 노드 개수만큼 생성될 수 있으므로 총 n 개가 도출된다. 이때, n 개의 노드 중 임의의 노드 v 를 선택하여 자아 네트워크와 확장 자아 네트워크를 도출하고, 도출된 네트워크 n 개에서 각각 임의의 노드 v 에 대한 매개 중심도를 측정할 수 있다. 본 논문에서는 임의의 노드 v 의 매개 중심도에 대하여 다음과 같은 용어를 정의한다.

- $B(v)$: 전체 네트워크로부터 임의의 노드 v 에 대해 측정된 매개 중심도
 - $B_{\Omega}(v)$: 전체 네트워크로부터 임의의 노드 v 에 대한 자아 네트워크에서 측정된 “자아 매개 중심도”
 - $B_X(v)$: 전체 네트워크로부터 임의의 노드 v 에 대한 확장 자아 네트워크에서 측정된 “확장 자아매개 중심도”
- 자아 매개 중심도와 확장 자아 매개 중심도는 두 가지

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

Nodes		v	a	b	c	d	e	f	g	h	i	j	k	l
$B(v)$	Value	0.405	0.383	0.124	0.131	0.286	0.000	0.318	0.049	0.030	0.167	0.000	0.000	0.000
	Rank	1	2	7	6	4	10	3	8	9	5	10	10	10
$B_{\Omega}(v)$	Value	0.667	0.750	0.167	0.583	0.333	0.000	0.833	1.000	0.167	0.667	0.000	0.000	0.000
	Rank	4	3	8	6	7	10	2	1	8	4	10	10	10
$B_X(v)$	Value	0.383	0.500	0.125	0.269	0.339	0.000	0.524	0.214	0.133	0.400	0.000	0.000	0.000
	Rank	4	2	9	6	5	10	1	7	8	3	10	10	10

모두 2.2절에서 제시한 식 (1)을 이용하기 때문에 계산 복잡도가 전체 네트워크의 매개 중심도와 차이가 없을 것 같지만 다음과 같은 측면에서 계산 복잡도가 다르다.

- 전체 네트워크의 노드의 수 및 간선의 수가 매우 많아도 자아 네트워크 및 확장 자아 네트워크의 노드의 개수 및 간선의 수는 상대적으로 작다.
- 각 노드마다 자아 매개 중심도($B_{\Omega}(v)$)와 확장 자아 매개 중심도($B_X(v)$)의 계산 과정은 서로 독립적이기 때문에 임의의 노드 v 의 자아 매개 중심도 및 확장 자아 매개 중심도는 동시에 계산이 가능하다.

즉, 노드의 매개 중심도를 측정하는데 있어서 자아 매개 중심도 $B_{\Omega}(v)$ 및 확장 자아 매개 중심도 $B_X(v)$ 가 기존의 방식에 의하여 전체 네트워크에서 산출한 매개 중심도 $B(v)$ 와 유사하다면 기존의 매개 중심도에 비해 활용 가치가 높다고 할 수 있을 것이다. 따라서 임의의 노드 v 의 대한 매개 중심도가 자아 매개 중심도 및 확장 자아 매개 중심도와 유사한지 비교해보기 위하여 2장에서 제시한 Fig. 1의 전체 네트워크 모형을 가지고 각 노드마다 자아 매개 중심도와 확장 자아 매개 중심도를 측정하여 측정값에 따른 순위(Rank)를 산출해 보았다 (Table. 1 참조). 매개 중심도에 대한 값 자체에는 차이가 있지만 위에서 산출한 중요도 순위를 가지고 유사성을 측정한 결과 $B(v)$ 와 $B_{\Omega}(v)$ 와 $B(v)$ 와 $B_X(v)$ 간의 스피어만 상관계수가 각각 0.79와 0.93으로 양의 상관관계를 보임을 알 수 있었다. 특히 $B(v)$ 와 $B_X(v)$ 간의 상관계수 값이 0.93으로 $B(v)$ 와 $B_{\Omega}(v)$ 의 상관관계에 비하여 더 높은 양의 상관관계를 가짐을 확인 할 수 있었다.

4. 제안하는 알고리즘

식 (1)에 따르면, 매개 중심도의 계산 복잡도는 네트워크를 구성하는 모든 노드 쌍의 최단경로의 계산에 비례하기 때문에 이웃 노드의 친구 관계 정보만을 활용하여 구성된 확장 자아 네트워크에서 확장 자아 매개 중심도를 계산할 경우 이의 측정시간을 단축시킬 수 있다. 그러나 사회관계망에 속한 노드의 수가 상당히 많거나 노드들 간의 상호 관계가 복잡하여 각각의 노드들의 확장 자아 네트워크의 크기가 커지는 경우를 고려하여, 본 논문에서는 매개 중심도의 계산 복잡도의 요인인 최단경로를 계산하는 작업을 최적화 및 전정함으로써 빠르게 확장 매개 중심도를 구하고자 한다.

식 (1)에 따르면 매개 중심도는 노드 s 와 t 사이의 노드 v 를 통과하는 최단경로의 개수인 $\sigma_{st}(v)$ 를 노드 s 와 t 사이의 최단경로의 개수인 σ_{st} 로 나눈 값을 v 를 제외한 모든 쌍에 대하여 합하고 이를 v 를 제외한 모든 쌍의 수로 나눈 값을 말한다. 여기서 $\sigma_{st}(v)/\sigma_{st}$ 는 s 에서 t 사이의 모든 경로들이 노드 v 에게 얼마나 중속되어있는지 정도를 의미한다. 본 논문에서는 이를 s 와 t 의 v 에 대한 의존성이라 명명하고 이를 다음과 같이 정의한다.

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

식 (1)에서 알 수 있듯이 임의의 노드 v 의 매개 중심도는 v 를 제외한 노드의 모든 쌍의 의존성 $\delta_{st}(v)$ (식 (2))을 계산함으로써 측정할 수 있다. 이때, 임의의 노드 쌍이 v 에 대한 의존성이 없어 노드 v 의 매개 중심도에 아무런 영향을 미치지 않는다면 해당 노드 쌍의 최단경로를 계산하는 작업은 불필요한 일일 것이다. 예를 들면 노드 s 와 t 가 Fig. 1(C)의 a 와 b 로 가정하여 의존성을 계산하면 v 를 거치는 최단경로인 $\sigma_{ab}(v)$ 가 0이므로 노드 v 의 매개 중심도에 아무런 영향을 미치지 않는다. 본 절에서는 확장 자아 네트워크의 특성을 분석하여 위에서 언급한 것과 같은 불필요한 작업을 제거하는 한편 의존성 $\delta_{st}(v)$ 을 조금 더 간단하게 계산하는 방법을 제안알고리즘에 반영함으로써 확장 자아 매개 중심도를 측정하는 시간을 단축시키고자 한다. 제안하고자 하는 알고리즘은 다음과 같다.

Algorithm 1 XegoBetweenness(X_v, V)

```

1: Input:  $X_v, V$ 
2: output:  $\frac{2 \cdot total}{(u+w)(u+w-1)}$ 
3: create an Array  $P[1...u+w][1...u+w]$ 
4: for  $s : 1$  to  $u$  do
5:   for  $t : s+1$  to  $u$  do
6:     if  $t \in V_{s,1}$  then
7:        $P[s][t] \leftarrow 0.0$ 
8:     else
9:        $P[s][t] \leftarrow 1/|V_{s,1} \cap V_{t,1}|$ 
10:    total  $\leftarrow total + P[s][t]$ 

```

```

11: end for
12: for t : u+1 to u+w do
13:   for each k ∈ Vt:1 do
14:     if t < k then
15:       append P[s][k] to C
16:     else
17:       append P[k][s] to C
18:     end if
19:   end for
20:   P[s][t] ← H(C)
21:   total ← total + P[s][t]
22: end for
23: end for
24: for s : u+1 to u+2 do
25:   for t : s+1 to u+w do
26:     for each k ∈ Vt:1 do
27:       if s < k then
28:         append P[s][k] to C
29:       else
30:         append P[k][s] to C
31:       end if
32:     end for
33:     P[s][t] ← H(C)
34:     total ← total + P[s][t]
35:   end for
36: end for
37: return  $\frac{2 \cdot total}{(u+w)(u+w-1)}$ 

```

제안하는 알고리즘은 확장 자아네트워크의 전체 노드를 순환하며 네트워크의 특성에 따라 임의의 노드 쌍 (s, t) 의 의존성 $\delta_{st}(v)$ 을 계산하고 이를 합하여 식 (1)과 같이 전체 노드에 대하여 생성되는 모든 간선의 수로 나누어 정규화함으로써 확장 자아 네트워크의 매개 중심도 값을 반환하는 의사코드이다.

Algorithm 1에서 X_v 는 임의의 노드 v 에 대한 확장 자아 네트워크이며, V 는 X_v 에 속한 노드들의 1-hop 이웃 노드, P 는 네트워크 X_v 에서 전체 노드 쌍의 의존성 값을 담은 이차원 배열, C 는 의존성을 계산하기 위한 임시 리스트, $total$ 은 계산되어진 모든 노드 쌍 (s, t) 의 의존성 $\delta_{st}(v)$ 의 합, u 는 v 의 1-hop 노드의 수, w 는 v 의 2-hop 노드의 수이다.

4-11번째 줄에서는 임의의 노드 쌍 (s, t) 이 v 의 1-hop노드인 경우, 의존성 $\delta_{st}(v)$ 을 계산한다. 여기서 s 와 t 가 이웃인 경우에는 최단경로는 $s-t$ 로 v 를 거치는 경로가 발생할 수 없기 때문에 $\sigma_{st}(v)$ 가 0되며 $\delta_{st}(v)$ 또한 0이다. 예를 들어, 노드 s 와 t 가 Fig. 1(C)의 노드 a 와 b 로 가정하여 의존성을 계산하면 노드 a 와 b 는 이웃이기 때문에 v 를 거치는 최단경로인 $\sigma_{ab}(v)$ 가 0이므로 노드 v 의 매개 중심도에 아무

런 영향을 미치지 않아 의존성 $\delta_{ab}(v)$ 는 0이다. 따라서 이의 경우는 의존성의 계산을 하지 않는다.

한편 s 와 t 가 이웃이 아닌 경우에는 s 와 t 사이의 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. 예를 들면, 노드 s 와 t 가 Fig. 1(C)의 a 와 c 로 가정하여 계산하면 노드 a 와 c 는 서로 이웃이 아니기 때문에 의존성 $\delta_{ac}(v)$ 는 $1/|V_{a:1} \cap V_{c:1}|$ 이다. 두 노드 사이의 공통된 이웃은 v 노드와 g 노드이며 최단 경로는 두 가지이다. 따라서 의존성 $\delta_{ac}(v)$ 는 $1/2$ 이다.

12-22번째 줄에서 노드 s 는 v 의 1-hop노드이고 노드 t 는 v 의 2-hop 노드인 경우, 의존성 $\delta_{st}(v)$ 을 계산한다. 이의 경우, s 와 t 사이의 최단 경로는 s 와 t 의 1-hop노드들의 사이의 최단 경로를 거쳐 t 까지 가게 되므로 $\delta_{st}(v)$ 는 s 와 t 의 1-hop노드 쌍의 의존성 값들의 조화평균($H(C)$)함으로써 구할 수 있다. 예를 들면 노드 s 와 t 가 Fig. 1(C)의 a 와 h 로 가정하여 계산하면 h 의 1-hop노드인 c 와 d 의 최단경로를 거쳐서 h 까지 도달하므로 앞서 계산한 $\delta_{ac}(v)$ 값인 $1/2$ 과 $\delta_{ad}(v)$ 값인 1 을 조화평균 하여 $\delta_{ah}(v)$ 값인 $2/3$ 를 계산할 수 있다. 이때, s 와 t 의 1-hop노드 쌍이 v 의 1-hop노드들이 아니라면, 6번째 줄에 초기화된 값인 0 을 사용하므로 $\delta_{st}(v)$ 에는 아무런 영향을 미치지 않게 된다.

24-35번째 줄에서는 임의의 노드 쌍 (s, t) 이 v 의 2-hop 노드인 경우, 의존성 $\delta_{st}(v)$ 을 계산한다. 이의 방법은 노드 s 는 v 의 1-hop노드이고 노드 t 는 v 의 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 정규화 하여 확장 자아 네트워크의 매개 중심도 값을 반환한다.

즉, 제안 알고리즘은 확장 자아 네트워크의 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 노드 간의 관계 정보만으로 최단 경로의 수를 계산함으로써 확장 자아 매개 중심도를 측정하는 시간을 상당부분 단축시킨다.

5. 실험결과 및 분석

5.1 조사대상 네트워크

조사 대상 네트워크 모형을 생성하기 위하여 총 9개의 네트워크 모형을 구축하였다. 이때 가상 네트워크와 실제 네트워크, 두 가지 환경에서 모두 유사한 실험결과가 나오는지 확인하기 위하여 가상 네트워크 모형 5개와 실제 사회관계망 모형 4개로 나누어 비교 분석하였다.

가상 네트워크 모형을 구축하기 위해 Barabási-Albert 네트워크 [9]라고 알려진 척도 없는(Scale-free) 네트워크 생성 방법에 노드 간의 클러스터링(Clustering) 특성을 부여한 네트워크 생성 모델 [10]을 사용한다. Barabási-Albert의 척도 없는 네트워크를 생성하기 위해서는 처음 1개 또는 그 이상의 노드를 랜덤하게 생성하고, 새로운 노드를 추가할 때 기존에 이미 생성되어 있는 노드들과 λ 개의 간선을 생성한다.

이때, 기존에 이미 생성되어 있는 어떤 노드와 간선이 생성 될지는 기존 노드들이 지닌 이웃 수에 비례한다. 이렇게 간선이 생성되는 것을 “선호적 연결(preferential attachment)”이라고 지칭하며, 이와 같은 선호적 연결은 노드들의 이웃 노드 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 [9].

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 [10].

이와 같은 특징들 때문에 본 논문에서 조사대상으로서 생성되는 네트워크들은 사회관계망과 매우 유사한 형태를 지닌다. 즉, 트위터, 페이스북, 링크드인의 친구관계를 분석해보면 노드들의 이웃노드 수 분포가 멱급수 함수를 따르는 것이 이미 여러 문헌에서 보고되었다 [4, 10]. 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 Python의 NetworkX 모듈에서 제공해주는 `barabasi_Albert_graph(n, m)` 함수를 사용하여 랜덤하게 생성하였으며 노드의 수 n 을 200으로 고정된 상태에서 m (11, 23, 37, 55, 100)을 증가시켜 네트워크의 밀도를 0.1부터 0.5까지 조정하면서 실험하였다.

한편, 실제 사회관계망 모형을 구축하기 위해 본 논문에서 사용한 데이터는 스탠포드 대학교의 SNAP(Stanford Network Analysis Platform)에서 제공하는 실제 페이스북 데이터이다 [11]. 이 데이터에서 제공하는 10개의 자아 네트워크 중에서 랜덤으로 3개의 네트워크를 선택하여 총 4개의 케이스로 나누어 네트워크 샘플을 구축하였다. 구축된 네트워크의 밀도가 증가 되는 순으로 나열하였고 각 네트워크의 특성은 Table 2에 제시하였다.

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.

Table 2. Facebook network characteristic

Network	Number of Node	Number of Edge	Density
#1	449	3381	0.033
#2	544	5025	0.034
#3	447	5438	0.054
#4	392	4959	0.064

5.2 확장 자아매개 중심도의 비교 분석

본 연구는 5.1절에서 구축한 총 9개의 전체 네트워크에 대하여 다음과 같은 데이터 리스트에 중요도 순위를 산출하여 유사성 비교 분석을 수행한다.

- 전체 네트워크에서 측정된 매개 중심도의 각 노드별 $B(v)$ 값 리스트
- 자아 네트워크에서 측정된 자아 매개 중심도의 각 노드별 $B_{\Omega}(v)$ 값 리스트
- 확장 자아 네트워크에서 산출한 확장 자아 매개 중심도의 각 노드별 $B_X(v)$ 값 리스트

이때, 전체 네트워크에서 임의의 노드 v 를 선택하여 측정된 매개 중심도 $B(v)$ 리스트는 모든 노드 및 모든 간선에 대해 종합적인 분석을 통해 얻은 값이기 때문에 가장 정확한 데이터라고 할 수 있다.

따라서 Fig. 2과 Fig. 3은 조사대상 네트워크에서 확장 자아 매개 중심도 리스트의 중요도 순위와 자아 매개 중심도 리스트의 중요도 순위를 산출하여 각각 전체 네트워크 매개 중심도 리스트의 중요도 순위와 얼마나 유사한지를 스피어만 상관계수를 통하여 보여준다.

스피어만 상관계수는 데이터 자료 자체의 값 대신 순위(Rank)에 대한 데이터를 가지고 상관계수를 구할 때 자주 사용되는 지표로서, 두 개의 서로 다른 데이터 집합에 대하여 데이터 값을 내림차순으로 순위를 매겨 서열 순서를 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.

한편, Fig. 2와 Fig. 3에서 네트워크 모형마다 집단화 계수(Clustering Coefficient)를 제시하였다. 집단화 계수는 특정노드와 이웃한 노드들이 서로 연결되어 있을 확률이며 각각의 노드들이 서로 뭉치는 경향을 알 수 있는 한 척도로 사용한다. 집단화 계수의 값이 클수록 연결의 밀도가 높으며, 사회관계망과 같이 실제로 존재하는 네트워크는 노드들 사이의 연결이 랜덤하게 생성된 네트워크에 비해서 상대적으로 집단화 계수가 높다고 알려져 있다. 따라서 집단화 계수의 값을 실험 대상 네트워크로써 적합하지에 대한 척도로 사용하는 한편, 값을 달리 함으로써 다양한 네트워크 환경에서 실험해 보았다.

Fig. 2에서 확인할 수 있듯이 $B(v)$ 와 $B_X(v)$ 의 순위 유사성은 모두 1에 가까운 양의 상관관계를 보였고, 반면에 $B(v)$ 와 $B_{\Omega}(v)$ 의 순위 유사성은 하나의 네트워크를 제외하고는 모두 0 이하의 음의 상관관계의 값을 보여주었다.

Fig. 3에서는 $B(v)$ 와 $B_X(v)$ 의 순위유사성과 $B(v)$ 와 $B_{\Omega}(v)$ 의 순위유사성 모두 양의 상관관계를 보였지만, 그 수치가 $B(v)$ 와 $B_X(v)$ 의 순위유사성일 때 더욱 1에 가깝다. 즉, 확장 자아 매개 중심도가 자아 매개 중심도보다 좀더 기존의 매개 중심도와 유사하다고 할 수 있다. 자아 네트워크 및 확장 자아 네트워크를 구성하는 방법의 용이성이 비슷하기 때문에 가능하면 확장 자아 매개 중심도를 사용하는 것이 각 노드의 보다 정확한 매개 중심도를 파악하는 데 도움이 된다.

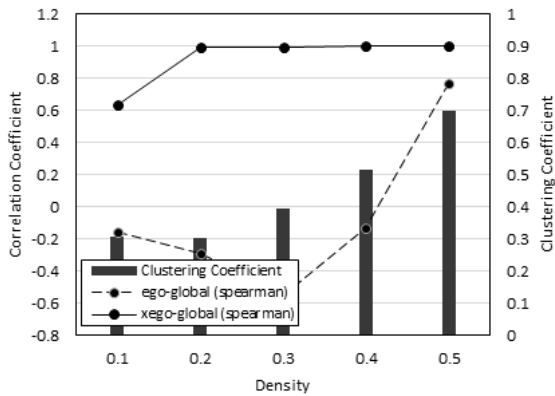


Fig. 2. Correlation Comparison in Barabási-Albert Network

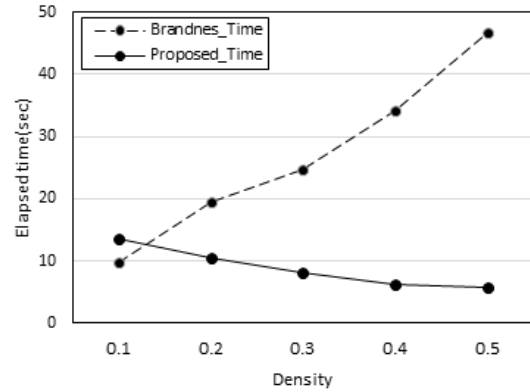


Fig. 4. Algorithm Performance Comparison in Barabási-Albert Network

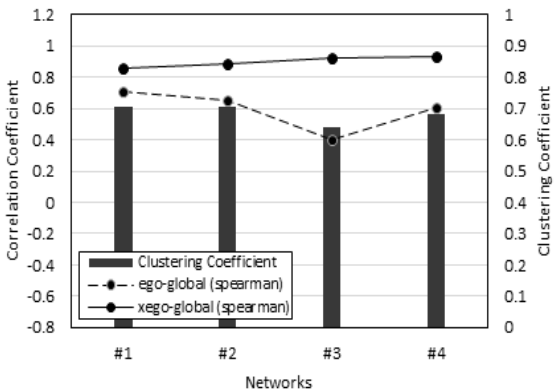


Fig. 3. Correlation Comparison in Facebook Network

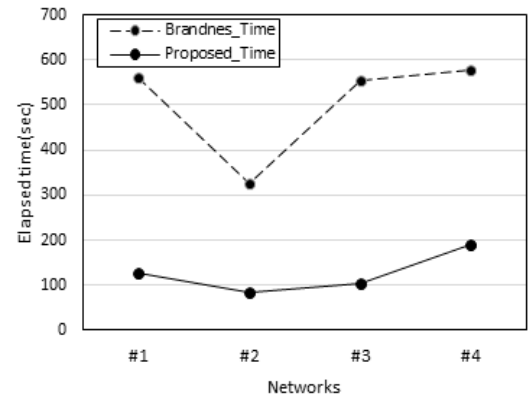


Fig. 5. Algorithm Performance Comparison in Facebook Network

5.3 제안하는 알고리즘 소요시간 분석

본 절에서는 5.1절에서 제시한 9개의 네트워크를 이용하여 자아 네트워크와 확장 자아 네트워크를 구성하고, 기존의 Brandes 알고리즘과 제안하는 알고리즘이 확장 자아 매개 중심도를 계산하기 위한 소요시간을 비교 분석한다.

Fig. 4와 Fig. 5는 확장 자아 매개 중심도를 구하기 위한 두 알고리즘의 수행시간을 비교한 그래프로 서로 다른 네트워크 밀도와 서로 다른 집단화 계수를 가진 여러 네트워크 환경에서 실험을 수행하였다. 이때, 네트워크의 밀도는 전체 네트워크에서 실제 간선의 개수를 노드 개수별 생성 가능한 전체 간선의 개수로 나눈 수치를 의미한다. Fig. 4는 5개의 가상 네트워크 모형이며 이 그래프에서 볼 수 있듯이, 네트워크 #1의 경우만을 제외하고 제안하는 알고리즘을 사용해 확장 자아 매개 중심도를 도출했을 때 걸린 시간이 Brandes 알고리즘을 사용했을 때 걸린 시간보다 더 짧은 시간이 소요됨을 확인할 수 있었다. 또한 밀도가 증가하고 집단화 계수가 높아질수록 제안하는 알고리즘과 Brandes 알고리즘의 소요시간 차이가 더 확연해졌다. Fig. 5는 페이스북 데이터를 이용한 4개의 실제 사회관계망 모형이며 제안한 알고리즘을 사용했을 때 더 짧은 시간이 소요되었다. 가상 네트워크 모형실험에서와 같이 밀도와 집단화 계수가 증가하면 할수록 확연한 소요시간 차이를 보인 것은 아니지만 제안한 알고리즘을 사용했을 때 4개의 전체 네트워크 모두 예외 없이 소요 시간이 단축되었고 그 차이가 확연히 드러났다.

6. 결론

본 논문은 네트워크를 구성하는 각각의 노드들마다 자신의 지역정보를 활용하여 구성된 네트워크의 매개 중심도를 산출함으로써 시간 복잡도를 줄이는 한편, 전체 네트워크에서 자아 네트워크와 확장 자아 네트워크를 구성하여 자아 매개 중심도와 확장 자아 매개 중심도를 측정하고 기존의 매개 중심도와와의 순위 유사성을 가상 네트워크 모형과 실제 사회관계망 모형에서 비교 분석하였을 때, 모든 경우에서 확장 자아 매개 중심도가 자아 매개 중심도보다 좀 더 기존의 매개 중심도와 유사함을 실험을 통해 증명하였다. 자아 네트워크 및 확장 자아 네트워크를 구성하는 방법의 용이성이 비슷하기 때문에 가능하면 확장 자아 매개 중심도를 사용하는 것이 각 노드의 보다 정확한 매개 중심도를 파악하는 데 도움이 될 것이다.

한편, 사회관계망에 속한 노드의 수가 상당히 많거나 노드들 간의 상호 관계가 복잡하여 각각의 노드들의 확장 자아 네트워크의 크기가 커질 경우를 고려해야 했다. 그 결과 확장 자아 네트워크의 특성을 분석하여 매개 중심도를 더 빠르게 측정할 수 있는 알고리즘을 제안하였고 9개의 가상 네트워크 모형과 실제 사회관계망 모형의 실험을 통하여 제안 알고리즘이 기존 알고리즘에 비해 매개 중심도를 더 빠르게 측정함을 증명하였다.

본 논문에서 제안한 알고리즘은 대규모 사회 관계망에서 확장 자아 매개 중심도 계산할 때 소요 시간을 단축하는 데 많은 도움이 될 것으로 기대된다. 특히 대중적으로 잘 알려진 페이스북은 임의의 대상 한 명 한 명마다 친구 간의 관계가 복잡한데 이런 경우에 좀 더 빠른 시간 내에 매개 중심도를 측정하도록 활용될 수 있다. 본 논문의 연구결과를 더욱 확장하여 향후에는 매개 중심도뿐만 아니라 다양한 사회관계망 분석 기법에 대해서 시간 복잡도를 감소시킬 수 있는 방안과 적용할 수 있는 알고리즘에 대해 연구할 예정이다.

References

[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] Sohn, J. S., S. W. Cho, K. L. Kwon, and I. J. Chung, "Improved Social Network Analysis Method in SNS," *Journal of Intelligence and Information Systems*, Vol.18, No.4(2012), pp.117-127.

[4] 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.

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

[6] 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

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

[8] 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.

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

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

[11] SNAP Stanford, <http://snap.stanford.edu/>.

[12] 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.

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

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

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



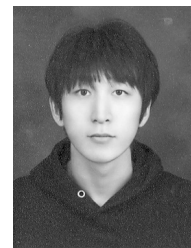
신수진

e-mail : soojin1116@koreatech.ac.kr

2013년 한국기술교육대학교 컴퓨터공학부 (학사)

2013년~현재 한국기술교육대학교 컴퓨터공학부 석사과정

관심분야: 사회관계망 분석



김용환

e-mail : cherish@koreatech.ac.kr

2010년 한국기술교육대학교 인터넷미디어 공학부(공학석사, 학석사 통합과정)

2010년~현재 한국기술교육대학교 컴퓨터공학부 박사과정

관심분야: 이동 컴퓨팅, 센서 네트워크, 사회관계망 분석



김찬명

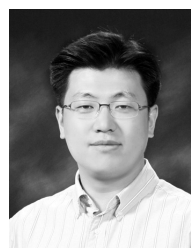
e-mail : cmdr@koreatech.ac.kr

2009년 한국기술교육대학교 멀티미디어 공학과(학사)

2011년 한국기술교육대학교 컴퓨터공학과 (석사)

2011년~현재 한국기술교육대학교 컴퓨터공학부 박사과정

관심분야: 사회관계망 분석, 지연 감내 네트워크



한연희

e-mail : yghan@koreatech.ac.kr

2002년 고려대학교 컴퓨터학과(이학박사)

2002년~2006년 삼성종합기술원 전문연구원

2006년~현재 한국기술교육대학교 컴퓨터공학부 교수

2005년~현재 한국정보통신기술협회(TTA) IT 국제 표준화 전문가

관심분야: 이동 컴퓨팅, 센서 네트워크, 사회관계망 분석