

# 지연 감내 네트워크에서 커뮤니티 기반 영향력 측정 기법

김 찬 명\*, 김 용 환\*, 한 연 희<sup>o</sup>

## A Community-Based Influence Measuring Scheme in Delay-Tolerant Networks

Chan-Myung Kim\*, Yong-hwan Kim\*, Youn-Hee Han<sup>o</sup>

### 요 약

사회 관계망에서 영향력의 전파는 중요한 연구 이슈이다. 영향 전파는 임의의 노드로 부터 새로운 아이디어, 정보, 소문의 전파로 인해 다른 노드들의 상태나 성질이 변화하는 것을 뜻한다. 영향을 받은 노드는 자신과 통신하는 다른 노드에게도 영향을 주고 영향은 확산되어 네트워크 내에서 퍼져 나간다. 입소문 마케팅에 기반을 둔 영향력 전파 문제는 네트워크에 가장 영향력을 끼칠 수 있는 노드들을 찾아 전체 네트워크에 영향력을 최대화 하는 것이 목적이다. 본 논문에서는 Delay-Tolerant Networks에서 각 노드의 영향력을 측정하여 가장 영향력 있는 노드 집합을 선택하는 문제를 다룬다. 노드 간 연결성이 항상 보장되지 않는 Delay-Tolerant Networks 환경에서는 전체 네트워크 정보를 정확히 알 수 없기 때문에 노드의 영향력을 정확히 측정하는 것은 쉽지 않다. 본 논문에서는 Delay-Tolerant Networks 환경에서 분산 방식으로 각자 노드가  $k$ -clique 구조로 커뮤니티를 구성하여 한정된 지역 정보만을 활용하여 자신의 영향력을 추정하는 방법을 제시한다. 또한, 실험을 통해 제안 기법으로 산출한 영향력 있는 노드 정보가 전체 네트워크 관점에서 산출한 영향력 있는 노드 정보와 거의 일치함을 보인다.

**Key Words** : Delay-Tolerant Networks, Community Detection, Influence Propagation, Distributed Algorithms

### ABSTRACT

Influence propagation is an important research issue in social networks. Influence propagation means that the status or the disposition of nodes get changed by new idea, information and gossip propagated by other nodes. Influenced nodes also make other nodes influenced across the network. The influence propagation problem based on 'word of mouth' referral is to find most influential nodes set in networks to maximize influence. In this paper, we study the influence measuring and finding most influential nodes set in Delay-Tolerant Networks. It is difficult to measure exact influential power in Delay-Tolerant networks where network topology is not stable due to the nodal mobility. In this paper, we propose a distributed scheme that each node constructs  $k$ -clique community structure and estimates local influential power in Delay-Tolerant Networks. Simulation results show that the influential nodes information estimated by proposed scheme is in agreement with a global view of influential nodes information.

\* 본 논문은 2012년도 정부(교육과학기술부)의 재원으로 한국연구재단의 지원을 받아 수행된 연구 (2012-0002558)이며 또한 한국기술교육대학교 교육연구진흥비지원 프로그램의 지원에 의하여 수행된 결과입니다.

◆ 주저자 : 한국기술교육대학교 컴퓨터공학과 첨단기술연구소, cmdr@koreatech.ac.kr, 준회원

○ 교신저자 : 한국기술교육대학교 컴퓨터공학과 첨단기술연구소, yhhan@koreatech.ac.kr, 종신회원

\* 한국기술교육대학교 컴퓨터공학과 첨단기술연구소, cherish@koreatech.ac.kr, 준회원

논문번호 : KICS2012-10-491, 접수일자 : 2012년 10월 12일, 최종논문접수일자 : 2012년 12월 28일

## I. 서 론

최근 사회 관계망 (Social Networks)에서 영향의 확산 (Influence Propagation)에 대한 연구가 진행되고 있다<sup>[1-4]</sup>. 영향은 새로운 아이디어, 정보, 소문의 전파로 인해 개인의 상태나 성질이 변화하는 것을 뜻하며 영향을 받은 개인은 자신과 교류하는 다른 상대에게도 영향을 주고, 그러한 영향은 지속적으로 확산되어 네트워크 내에서 퍼져 나간다. 네트워크에서 영향력 최대화 문제는 입소문 마케팅 전략에 기반을 둔 것으로 전체 네트워크에 영향력을 최대화시킬 수 있는 초기 노드 집합을 찾는 문제이다<sup>[2]</sup>. 예를 들어 기업에서 새로운 제품이 나와 입소문 마케팅으로 홍보를 할 때 홍보할 제품의 샘플의 개수는 한정적일 경우 해당 샘플을 누구에게 나누어 주어 홍보효과를 극대화 할 것인가의 문제와 비슷하다. 이 때 개개인의 영향력을 측정하고 영향력을 크게 발휘할 수 있는 대상을 고르는 방식은 중요하다.

다수의 기존 연구들<sup>[1-4]</sup>은 주로 노드 간 연결성이 항시 보장되어 있고 네트워크 토폴로지가 변하지 않는 정적인 네트워크에서의 영향력 최대화 문제에 대하여 다루고 있다. 즉, 임의의 노드가 전체 네트워크 상황을 알고 있다고 가정하고 중앙 집중형 방식으로 네트워크에 영향력을 최대화시킬 수 있는 초기 노드 집합을 찾는 방식에 대한 연구가 진행되어 왔다. 영향력 최대화 문제에서 많이 이용되어온 영향 전파 모델<sup>[2]</sup>에 따라 각 노드는 활성상태와 비활성상태를 가질 수 있으며  $n$ 개의 초기 활성 상태 노드 집합으로부터 영향의 확산이 시작된다. 활성 상태의 노드는 다른 비활성 상태의 노드를 특정 조건에 따라 활성화(영향을 전파)시킬 수 있으며 한번 활성화시도에 실패한 노드에 대해서는 다시 활성화를 시도할 수 없다. 이 과정을 더 이상 영향력 확산이 발생하지 않을 때까지 반복한다. 하지만 노드들이 활발하게 이동하여 네트워크 토폴로지가 변하는 동적인 네트워크에서 기존의 중앙 집중형 방식으로 노드의 영향력을 측정하는 것은 어려운 문제이다. 임의의 노드에 대한 영향력 측정은 일반적으로 전체 네트워크 정보를 기반으로 측정해야 하기 때문에 동적인 네트워크의 특성상 중앙 집중형 노드에게 네트워크의 동적인 상황이 지속적으로 보고되어야 하며 이러한 상황은 전체 네트워크에 많은 오버헤드가 유발될 수 있다.

한편, 컴퓨팅 환경의 변화에 따라 전통적인 유선망 네트워크뿐만 아니라 이동성이 가미된 무선 모

바일 네트워크 또한 발전하고 있다. 이 중 Delay-Tolerant Network (DTNs)<sup>[5-10]</sup>는 유선망과 같이 노드와 노드가 항상 연결되어 있음이 보장되지 않는 네트워크로서 정보의 전달은 노드와 노드가 만날 때만 이루어진다. 이러한 DTN에서는 종단 간 연결성이 보장되지 않기 때문에 전체 네트워크로 임의의 메시지를 전파시켜야 할 때 메시지 전파 시간 (Message Dissemination Latency)이 예측할 수 없이 길어질 수 있다. 만약 이러한 DTN에서 각 노드의 영향력을 올바르게 파악할 수 있다면, 임의의 노드가 다른 노드들로 보내야 할 메시지가 있을 때 그 노드가 자신보다 높은 영향력을 지닌(높은 활동성으로 다수의 노드와 오랜 시간 접촉해온) 노드에게 해당 메시지를 전달하여 전체 메시지 전파 시간을 대폭 줄일 수 있다.

P. Hui 등은 [5] 임의의 DTN에서 노드와 노드 사이의 접촉기록에 따라 각 노드가 자기 자신의 국지적 커뮤니티 (Local Community)를 구성하는 3가지 알고리즘 (SIMPLE,  $k$ -CLIQUE, MODULARITY)을 제시했다. 여기서  $k$ -CLIQUE 알고리즘에서 활용하는  $k$ -clique 구조는  $k$ 개의 노드로 이루어진 무방향 그래프에서 각각의 노드가 서로서로 직접 연결되어 있는 구조이다. 즉, 크기가  $k$ 일 때 각 노드의 차수는  $k-1$ 을 유지해야 한다. 실제 사회에서 커뮤니티 혹은 그룹에 속해서 구성원들과 활발히 활동하는 사람이 있고 그렇지 않은 경우가 있듯이, 이동성이 고려된 DTN에서도 노드의 활동성에 따라 많은 노드와 만나고 교류하는 노드가 있는 반면 소수의 노드들과만 교류하는 노드가 존재한다. 노드는 자신과 전혀 교류가 없는 노드보다 자신이 속한 커뮤니티 내의 노드들에게 많은 정보와 영향력을 전파할 수 있으며, 완전그래프 형태로 응집력이 높은  $k$ -clique 구조에서는 임의의 노드가 자신의 커뮤니티 내부 노드들에게 많은 영향력을 발휘할 수 있다.

본 논문에서는 DTN 환경에서 P. Hui 등에 [5]의 의해 제시된 국지적  $k$ -CLIQUE 커뮤니티 구성 알고리즘을 사용하여 분산 방식으로 각각의 노드가 자신의 영향력을 추정하는 방법을 제시한다. DTN에서는 각 노드가 전체 네트워크 구조를 알 수 없기 때문에 자신의 국지적 커뮤니티에 미치는 영향력으로 전체 네트워크에 파급될 영향력을 추정한다. 제안 기법에서 각각의 노드는 접촉 기록에 따라 자신과 많이 교류하고 조건에 맞는 노드를 자신의 국지적 커뮤니티에 포함하게 된다. 자신의 커뮤니티에

노드를 추가할 때마다 가상으로 활성화(Pseudo-activation)를 시도하고 그러한 가상 활성화의 성공 여부 및 활성화 정도를 사용하여 자신의 영향력을 추정한다. 본 논문에서 제안하는 방식을 사용하면 여러 노드와 교류하고 높은 활동성을 지니며 커뮤니티 내 구성 노드들이 많은 노드들을 전체 네트워크 관점에서 영향력이 높은 노드로 선택할 수 있다. 또한, DTN에서 메시지 전파 효율을 높이기 위해 사회관계망 분석을 활용하는 다수의 기존 연구들이 활용해왔던 실험 데이터를 이용하여, 제안 기법으로 산출한 임의의 노드의 영향력이 전체 네트워크 관점에서 산출한 노드의 영향력에 근접함을 실험을 통해 증명한다. 본 논문이 기여하는 바를 요약하여 정리하면 다음과 같다.

- 1) 영향력 전파 모델에 활용되는 링크 가중치를 위하여 노드 간 접촉시간 정보를 활용한 모델 제시
- 2) 각 노드가 네트워크에 대한 선행 정보 없이 실시간적으로 국지적 커뮤니티를 구성하는 기법 제시
- 3) 이동성이 가미된 동적인 네트워크에서 분산 방식으로 각 노드의 영향력을 추정하는 방안 제시

본 논문의 구성은 다음과 같다. 2장에서는 본 논문에서 활용하는 DTN 모델과 Independent Cascade 영향력 전파 모델을 제시하고, 3장에서는 국지적 커뮤니티 구성을 활용한 노드 영향력 측정 기법을 제안한다. 4장에서는 성능 분석 및 결과를 보이고 5장에서 본 논문의 결론을 짓는다.

## II. DTNs 및 영향 전파 모델

### 2.1. DTNs (Delay-Tolerant Networks) 모델

Delay-Tolerant Networks (DTNs)는 이중의 네트워크를 연결하기 위해 등장한 개념으로 중단 간 연결성이 보장되지 않는 상황에서도 노드 간 통신을 지원하는 네트워크이다. 본 논문에서는 이동성을 지닌 다수의 노드로 이루어져 네트워크 위상(Topology)이 동적으로 변하는 DTN을 가정한다.  $r$ 을 임의의 노드의 최대 통신 범위라고 할 때, 노드  $v_i$ 는 거리  $d \leq r$ 에 있는 노드에게 메시지 전달이 가능하다. 이 때 메시지간의 충돌은 발생하지 않으며 각 메시지는 성공적으로 한 번에 전달 가능하다고 가정한다. 노드  $v_i$ 와 노드  $v_j$ 의 거리가 가까워지고 노드  $v_j$ 가 브로드캐스트한

hello 메시지를  $v_i$ 가 수신했을 때 노드  $v_i$ 와 노드  $v_j$ 가 접촉(Contact)했다고 판단하며 향후 설명할 Independent Cascade 모델을 통하여 두 인접노드 사이에 영향력 전파가 가능하다고 가정한다. 또한, 노드  $v_i$ 가 노드  $v_j$ 의 통신범위 안에 있을 때  $v_i$ 는  $v_j$ 가 주기적으로 보내는 hello 메시지를 수신하며 미리 정해 놓은 수 (예를 들어, 3번)만큼의 hello 메시지를 수신하지 못할 경우 노드  $v_i$ 는 노드  $v_j$ 가 떠났다(Leave)고 판단한다. 만약  $\alpha$ 가 임의의 노드와 접촉한 시가지고  $\beta$ 가 접촉했던 노드가 떠난 시가지라면  $\beta - \alpha$ 을 접촉 시간 (Contact Duration)이라고 정의한다. 한편,  $cd_{ij}$ 는 노드  $v_i$ 가 임의의 노드  $v_j$ 와의 총 누적 접촉 시간 (Cumulative Contact Duration)을 의미한다. 즉, 노드  $v_i$ 는 임의의 노드  $v_j$ 와 접촉한 시각부터  $v_j$ 가 떠난 시각까지의 접촉 시간을 접촉과 떠남이 반복적으로 일어날 때 마다  $cd_{ij}$ 에 누적하여 기록한다.

### 2.2. Independent Cascade 모델

Independent Cascade 영향력 전파 모델<sup>[1]</sup>은 사회관계망 연구에서 많이 사용되는 동적 영향 전파 모델로서 노드와 노드가 연결된 각 링크에 정보가 확산될 확률을 정의하여 확산 패턴을 표현하는 모델이다. Independent Cascade 모델에서 각 노드는 두 가지 상태인 활성(Active)상태와 비활성(Inactive)상태를 가진다. 비활성 상태인 노드는 다른 활성화상태 노드로부터 활성화가 되지 않은 노드를 뜻한다. 입소문 마케팅에 기반한 이 모델은  $n$ 개의 초기 활성 상태 노드로부터 확산이 시작되며, 활성 상태의 노드는 다른 비활성 상태의 노드를 활성화(영향을 전파)시킬 수 있다. 비활성 상태에서 활성상태로 전환할 수는 있으나 반대로 활성상태에서 비활성 상태로의 전환은 불가하다.

활성 상태의 노드  $v_i$ 가 시각  $t$ 에 비활성 상태의 노드  $v_j$ 에 대해  $p(i, j)$ 의 확률로 활성화를 시도할 수 있고 활성화에 성공할 경우 시각  $t+1$ 에  $v_j$ 는 활성화 된다. 한번 활성화에 실패한 노드에 대해서는 다시 활성화를 시도할 수 없다. 이 과정을 더 이상 영향력 확산이 발생하지 않을 때까지 반복한다.

Fig. 1은 임의의 네트워크에서 Independent Cascade 영향력 전파 모델을 적용한 영향력 확산 예시를 나타낸다. 회색 노드는 활성 상태를 나타내고 흰색 노드는 비활성 상태를 나타낸다.  $t$ 시각에

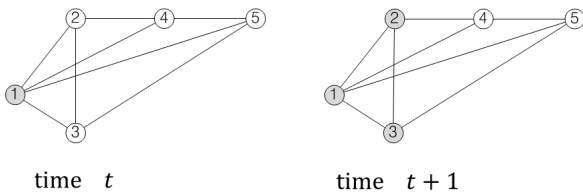


Fig. 1. Influence propagation in independent cascade model

활성상태에 있는 노드 1로부터 노드 2-5 까지 활성화를 각각 시도한 결과  $t + 1$  시각에 노드 2와 3에 대해서는 활성화를 성공하고 4와 5에 대해서는 활성화를 실패하였다. 이 때 초기 활성화노드 1은 네트워크에서 2

개의 노드를 활성화시킬 수 있을 만큼의 영향력을 가진다고 볼 수 있다.  $t + 2$  시각에는 노드 2와 3이 노드 4와 5에 대해 활성화 시도를 수행한다.

DTN에서 서로 다른 노드들이 많은 양의 메시지들을 주고받는다면 그러한 메시지들에 대한 전달 성공률은 높아야 하고 전달 비용은 반대로 적어야 한다. 이 때, 각 노드들은 활발하게 이동을 하면서 서로 간에 간헐적으로 접촉을 하기 때문에 메시지 전달 효율 측면에서 영향력이 높은 노드는 다른 다수의 노드와 활발하게 접촉하는 노드라고 간주할 수 있다. 그러므로, 본 논문에서는 각각의 노드  $v_i$ 가 노드  $v_j$ 를 활성화할 확률  $p(i, j)$ 를 다음과 같이 우선 정의한다.

$$p(i, j) = \frac{cd_{ij}}{cd_{max}(i)} \quad (1)$$

위 식에서  $cd_{max}(i)$ 는 노드  $v_i$  입장에서 그동안 접촉해 온 모든 노드들 중 가장 높은 누적 접촉 시간을 뜻한다 ( $cd_{max}(i) = \{\forall j, max(cd_{ij})\}$ ). 즉,  $p(i, j)$ 는 그러한 누적 접촉시간에 대하여  $v_j$ 와 의 누적 접촉시간 비율을 의미한다. 만약 노드  $v_i$ 가  $v_j$ 와 많은 접촉 시간을 기록하고 있다면  $p(i, j)$  값은 높아질 것이며, 그 만큼  $v_i$  입장에서  $v_j$ 에 대한 활성화 성공률이 높아진다. 결국  $v_i$ 가 다수의 노드들과 많은 접촉 시간을 기록한다면 영향력이 높은 노드가 된다.

### III. 제안 기법

본 장에서는 노드 이동성을 고려하는 DTN에서 각

각의 노드가 자신의 국지적 커뮤니티를 구성하며 영향력을 추정하는 기법을 제시한다. 제안 기법은 P. Hui 등에 [5] 의해 제시된  $k$ -CLIQUE 분산 방식 알고리즘을 토대로 커뮤니티를 구성하며 각각의 노드가 독립적으로 수행하게 된다.  $k$ -CLIQUE 알고리즘에서 활용하는  $k$ -clique 커뮤니티 구조는 형성하는 기준이 엄격하고 커뮤니티 내의 노드끼리 응집력이 높은 특성을 지닌다. 커뮤니티 내의 노드끼리는 접촉을 많이 하고 그로 인해 커뮤니티 내의 영향력이 최대화될 수 있음을 예상할 수 있다. 반면에 커뮤니티 밖의 노드와는 접촉이 적고 영향력 또한 발휘되기 어렵다. 이런 특성으로 인해  $k$ -CLIQUE 알고리즘으로 커뮤니티를 구성하면서 각 노드의 국지적인 영향력을 측정하고 전체 네트워크 관점에서의 영향력을 추정할 수 있다. 제안 알고리즘에 사용되는 기호는 Table. 1과 같다.

한편, 임의의 노드  $v_i$ 가 다른 노드  $v_j$ 와 접촉을 하게 되면  $v_j$ 의 Familiar Set인  $\widetilde{F}_{ij}$ 와  $v_j$ 의 커뮤니티인  $C_j$ 에 속한 각 노드들의 Familiar Set을 원소로 지니는 집합인  $\Pi_{ij}$ 를  $v_j$ 로부터 얻는다. 임의의 노드  $v_i$  자신은 항상 자신의 Familiar Set을 정확하게 파악이 가능하지만, 이웃 노드  $v_j$ 나  $C_j$ 에 속한 노드  $v_k$ 들로부터 얻는 Familiar Set 정보는 해당 정보를 얻는 시점 직후부터 변경이 될 수 있고  $v_i$  입장에서 그러한 변경 사항을 정확하게 파악하기 힘들기 때문에 "추정 Familiar Set"이라고 일컫는다.

Table. 1. Notations in Algorithm

$F_i$	The Familiar Set of a node $v_i$
$\widetilde{F}_{ij}$	The estimated Familiar Set of a node $v_j$ which is the neighbor of $v_i$
$\gamma$	The Familiar Set threshold
$k$	The size of clique
$C_i$	The local community of a node $v_i$
$\Pi_{ij}$	The set including the estimated Familiar Set of each node $v_k$ in the local community $C_j$ of a node $v_j$ which is the neighbor of $v_i$ (e.g., $\Pi_{ij} = \{\widetilde{F}_{jk} \mid v_k \in C_j\}$ )
$N_i^C$	The number of nodes pseudo-activated by a node $v_i$ in the local community $C_i$

한편, 제안 알고리즘에서  $k$  파라미터는 커뮤니티를 구성하는 기본 단위인  $k$ -clique 구조의 크기를 나타내며 각 노드마다 동일한  $k$ 값이 미리 설정되어 있다고 가정한다.

임의의 노드  $v_i$ 는 자신의 국지적 커뮤니티  $C_i$ 에 새로운 노드  $v_j$ 를 추가할 때마다 의사 활성화(pseudo-activation)를 시도한다. 의사활성화는 Independent Cascade 모델을 따르며 새로 추가되는  $v_j$ 의 각 인접노드들 중 활성화상태에 있는 노드  $v_l$ 은  $p(l, j)$ 의 확률로  $v_j$ 에 의사활성화를 시도한다. 각 노드는 다른 노드에 대해 의사활성화 시도 여부를 기록하고 있다고 가정하고 한번 의사활성화를 시도한 노드에 다시 의사활성화를 시도할 수 없다. 노드  $v_j$ 에 대해 의사활성화가 성공하였다는 것을 알게 된  $v_i$ 는 자신의  $N_i^C$ 를 업데이트한다. 임의의 노드  $v_i$ 의 커뮤니티  $C_i$ 에 새롭게 추가된  $v_j$ 에 대한 의사 활성화 절차인 **Pseudo-Activation**은 다음과 같이 기술된다.

**Pseudo-Activation ( $v_j$ )**

1.  $v_j$ 의 인접노드들 중 활성화상태인 노드  $v_l$ 은  $p(l, j)$ 의 확률로  $v_j$ 에 대해 의사 활성화를 시도한다.
2. 노드  $v_j$ 는 활성화 성공 여부를  $v_i$ 에 알려준다.
3.  $v_j$ 가 의사 활성화되었다는 것을 알게 되면  $v_i$ 는  $N_i^C$ 를 업데이트 한다 (즉,  $N_i^C = N_i^C + 1$ ).

Fig. 2는  $v_i$ 의 국지적 커뮤니티  $C_i$ 에 새로 추가된 노드  $v_j$ 에 대해 인접 노드들이 의사활성화를 시도하여 성공한 상황을 보여준다. 이 그림에서 회색 노드들은 활성화 노드를 뜻하며 흰색 노드들은 비 활성화 노드를 뜻한다. 한편, 각각의 노드  $v_i$ 는 독립적으로  $F_i$ 와  $C_i$ 를 계속하여 유지 관리하는 다음과 같은 제안 알고리즘을 수행한다. 임의의 노드  $v_i$ 가 DTN에 참여할 때에 다음과 같은 초기화 절차 **Procedure 0**을 수행한다.

**Procedure 0**

$C_i = \{v_i\}, F_i = \emptyset, \Pi_i = \{\emptyset\}$

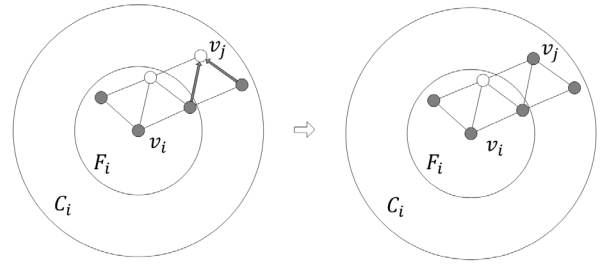


Fig. 2. Adjacent nodes of  $v_j$ 's Pseudo-Activation to a  $v_j$  when  $v_i$  adds  $v_j$  to its local community  $C_i$

다음 **Procedure 1**은 임의의  $v_i$ 가 다른 노드  $v_j$ 와 접촉이 되었을 때 그 노드를 자신의  $C_i$ 에 추가할지 판단하여 만약  $v_j$ 가  $C_i$ 에 추가된다면 의사 활성화를 시도하는 절차이다.

**Procedure 1**

접촉한 노드  $v_j$ 에 대해서

1.  $v_j$ 의  $\widetilde{F}_{ij}, \Pi_{ij}$  정보를 획득한다.
2. 만약  $v_j$ 가 자신의  $C_i$ 에 존재하지 않으면 (즉,  $v_j \notin C_i$ )
  - 2-1. 자신의  $C_i$ 와 노드  $v_j$ 의  $F_j$ 가 공통으로 최소  $k-1$ 개의 노드를 공유하고 (즉,  $|C_i \cap F_j| \geq k-1$ ) 있다면 노드  $v_j$ 를  $C_i$ 에 추가한다.
  - 2-2.  $v_j$ 에 대해 **Pseudo-Activation ( $v_j$ )** 을 수행한다.
  - 2-3. **Procedure 3**을 수행한다.

Fig. 3은 **Procedure 1**에서 노드  $v_i$ 의 국지적 커뮤니티  $C_i$ 에 노드  $v_j$ 를 추가를 시도하는 모습을 나타낸다.  $k=3$ 일 때  $v_j$ 의  $F_j$ 와  $C_i$ 가 노드 2개를 공유하므로  $v_i$ 는  $v_j$ 를  $C_i$ 에 추가한다. 이렇게  $C_i$ 에 새로운 노드  $v_j$ 가 추가되면 그 노드에 대해 **Pseudo-Activation**과 이후 설명할 **Procedure 3**도 수행한다. 여러 노드와 활발히 접촉하는 노드일수록 **Procedure 1**을 수행함으로써 자신의 국지적 커뮤니티에 노드를 많이 추가할 수 있고 그에 따라 새로운 노드에 대해 의사활성화를 시도하고 성공할 확률 또한 높아진다.

임의의  $v_i$ 가 다른 노드  $v_j$ 와 접촉이 되었을 때  $v_j$ 가  $C_i$ 에 추가되지 않더라도 현재의 접촉 상황에

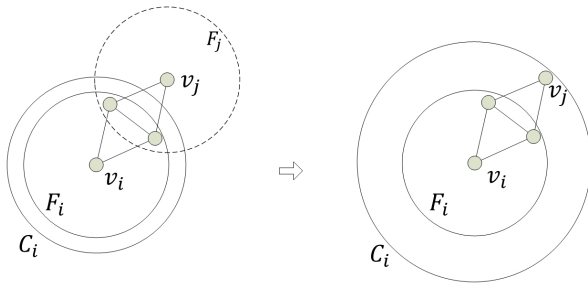


Fig. 3. Adding  $v_j$  to  $C_i$  of  $v_i$  by Procedure 1 ( $k=3$ )

의하여  $v_j$ 가 떠나는 시점에  $v_j$ 를  $C_i$ 에 추가시킬 수 있다. 다음 Procedure 2는 임의의  $v_i$ 가 다른 노드  $v_j$ 와 접촉 이후  $v_j$ 가 떠날 때 마다 그 노드를 자신의  $C_i$ 에 추가할지 판단하여 만약  $v_j$ 가  $C_i$ 에 추가되면 그 노드에 대해 Pseudo-Activation과 이후 설명할 Procedure 3을 수행하는 절차이다.

**Procedure 2**

접촉 이후 떠난 노드  $v_j$ 에 대해서

1.  $cd_{ij}$ 와  $cd_{max}(i)$ 를 업데이트 한다.
2.  $p(i, j)$ 를 업데이트 한다.
3. 만약  $p(i, j) \geq \gamma$ 이면
  - 3-1.  $v_i$ 는  $v_j$ 를 자신의  $F_i$ 에 추가한다.
  - 3-2. 만약  $v_j$ 가  $C_i$ 에 존재하지 않으면 (즉,  $v_j \notin C_i$ )  $v_j$ 를  $C_i$ 에 추가한다.
  - 3-3.  $v_j$ 에 대해 Pseudo-Activation ( $v_j$ ) 을 수행한다.
  - 3-4. Procedure 3을 수행한다.
4. 만약  $p(i, j) < \gamma$ 이면  $v_i$ 는  $v_j$ 를 자신의  $F_i$ 에서 삭제한다 ( $C_i$ 는 그대로 유지함).

위에 기술한 Procedure 2는 임의의 노드  $v_i$ 가  $p(i, j)$ 를 기반으로 자신의 Familiar Set인  $F_i$ 를 최신 정보로 유지하는 목적을 지닌다. 제안 기법에서는 Procedure 1과 Procedure 3에서 임의의 노드가 자신의 로컬 커뮤니티에 노드를 추가할 때  $k$ -clique 구조를 만족하기 위하여 추가되는 노드의 Familiar Set 정보를 이용하므로 특정 시점(노드가 접촉 이후 떠날 때)마다 Familiar Set 정보를 업데이트할 필요가 있다. 또한,  $v_i$ 의 커뮤니티  $C_i$ 는 항상  $F_i$ 에 속한 노드를 포함하도록 유지하며 한번이라도  $C_i$ 에

속했던 노드는  $C_i$ 에 계속하여 유지된다.

또한, 임의의 노드  $v_i$ 가 새로운  $v_j$ 를 자신의  $C_i$ 에 포함시키면  $v_j$ 가 알려주는  $\Pi_{ij}$  정보를 이용하여  $v_j$ 의 커뮤니티인  $C_j$ 안의 각 노드  $v_k$ 들을  $C_i$ 에 추가할지도 추가적으로 판단하며 이에 대한 절차는 Procedure 3에 기술된다.

Fig. 4는 Procedure 1 또는 Procedure 2에 의하여 노드  $v_i$ 의 커뮤니티  $C_i$ 에 노드  $v_j$ 를 추가한 이후, Procedure 3에서 제시된 절차에 의해  $\widetilde{F}_{jk}$ 과  $C_i$ 가 노드 2개를 공유하므로 부가적으로  $C_j$ 안의  $v_k$ 을  $C_i$ 에 추가하는 모습을 나타낸다. DTN에 참여하는 각 노드  $v_i$ 는 독립적으로 제안한 절차들을 수행하면서 자신의 국지적 커뮤니티  $C_i$ 를 구축할 수 있으며,  $C_i$ 에 계속하여 추가되는 노드들에 대한 의사 활성화 시도에 성공한 노드 수인  $N_i^C$  또한 증가시킬 수 있다.

**Procedure 3**

1.  $C_i$ 에 새롭게 추가된  $v_j$ 의 커뮤니티인  $C_j$ 안의 각각의 노드  $v_k$ 에 대해서 만약  $\widetilde{F}_{jk}$ 와  $C_i$ 가 공통으로 최소  $k-1$ 개의 노드들을 공유 (즉,  $|\widetilde{F}_{jk} \cap C_i| \geq k-1$ ) 한다면  $v_k$ 를  $C_i$ 에 추가한다.
2.  $v_k$ 에 대해 Pseudo-Activation ( $v_k$ ) 을 수행한다.

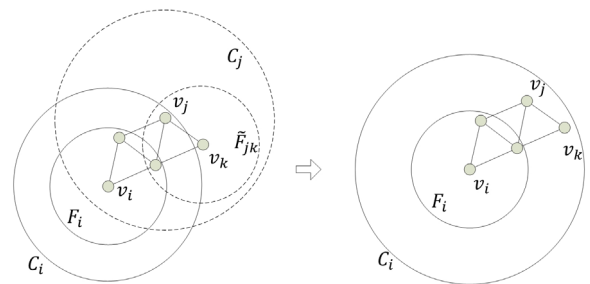


Fig. 4. Adding  $v_k$  to  $C_i$  of  $v_i$  by Procedure 3 ( $k=3$ )

이러한  $N_i^C$  값에 기반하여 각 노드는 전체 네트워크에서 자신의 영향력을 추정할 수 있다.

#### IV. 실험 및 성능 분석

본 장에서는 실제 노드의 접촉 기록이 있는 Hagle 프로젝트 데이터 집합[11]을 사용하여 실험을 수행한다. 본 데이터 집합들의 특징은 이전의 DTN 관련 연구들[5-9]에 의해 심도있게 분석 및 활용 되어왔다. 본 논문에서는 Hagle 프로젝트 데이터 집합 중 ‘Intel’과 ‘Infocomm05’ 데이터 집합으로 실험하였으며 각 데이터 집합의 정보는 Table. 2와 같다.

본 실험에서는 주어진 네트워크 데이터 집합에 기반하여 제안 기법으로 각각의 노드가 국지적 커뮤니티를 구성한 후 주어진 시간 주기마다 각 노드  $v_i$ 에 대하여  $N_i^C$ 를 검출한다. 또한, 각 노드  $v_i$ 가 제안 알고리즘과 상관없이 전체 네트워크 내의 인접노드들을 따라 의사활성화할 수 있는 노드 수인  $N_i$ 를 측정하여 각각을 비오름차순으로 정렬한다.

즉, 본 실험의 의도는 DTN에서 각 노드가 전체 네트워크 정보를 알 수 없으므로 본 논문에서 제안하는 방법으로 각 DTN 노드마다 독립적으로  $N_i^C$ 를 구하고, 전체 네트워크 정보를 알 수 있는 가상의 중앙 집중 노드를 만든 후 이 노드에서 모든 DTN 노드마다 전체 네트워크 정보를 대상으로  $N_i$ 를 구한 후  $N_i^C$ 와  $N_i$ 의 유사성을 비교하여 제안하는 분산 알고리즘의 우수성을 입증하는 의도를 지닌다. 평가 방법으로는  $N_i$ 가 높은 순으로 선택한  $n$ 개의 노드들이 전체 네트워크에서 활성화하는 노드 수를  $\alpha$ 라고 하고, 마찬가지로  $N_i^C$ 가 높은 순으로 선택한  $n$ 개의 노드들이 전체 네트워크에서 활성화하는 노드 수를  $\beta$ 라고 할 때,  $\alpha$ 에 대한  $\beta$ 의 비율인  $\rho_1 = \beta/\alpha$ 를 매 단위 시간마다 측정한다.

Table. 2. Experimental Datasets

Dataset	Intel	Infocomm05
Device	iMote	iMote
Protocol	Bluetooth	Bluetooth
Duration (days)	5	3
Number of Nodes	20	41
Number of Contacts	560	22459
Average Number of Contacts/pair/day	0.084	4.6

만약  $\rho_1$ 값이 1과 비슷하다면 전체 네트워크 정보 없이 국지적 커뮤니티 정보만을 활용하는 제안 방법이 각 노드의 영향력을 평가하는데 유용하게 사용될 수 있음을 나타낸다고 볼 수 있다. 전체 실험에서  $\gamma = 0.3$ 과  $k = 3$ 을 사용한다.

Fig. 5는 Intel 데이터 집합에서 단위 시간마다의  $\rho_1$ 의 변화를 보여준다. 영향도 측면에서 선별하는 상위 노드 개수인  $n$ 은 1개 또는 2개의 노드를 선택하였다. 단위 시간  $t=0$ 에서  $t=103542$ 까지는 제안 기법의 성능 지표인  $\rho_1$ 값의 변동 폭이 크지만 최종 단위 시간  $t=359312$ 에는  $n = 1$ 일 때  $\rho_1$ 값은 0.88이고  $n = 2$ 일 때  $\rho_1$ 값은 0.95로 수렴하는 모습을 나타낸다. 즉,  $N_i^C$ 와  $N_i$  기준으로 선별한 상위 노드 개수인  $n$ 이 클 때  $\rho_1$ 값 또한 높음을 알 수 있다. 따라서, 제안 기법으로 국지적 커뮤니티를 구성하여 선택된 노드들이 전체 네트워크 관점에서 선택된 노드들과 비교하여 전체 네트워크 노드들에게 행사할 수 있는 영향력이 거의 유사함을 알 수 있다.

Fig. 6은 Intel 데이터 집합에 존재하는 모든 노드 수  $N$ 에 대하여 각 단위 시간  $t$ 마다 각각  $N_i^C$ 와  $N_i$ 가 높은  $n$ 개의 노드들로부터 네트워크 내에서 활성화되는 노드 수  $N_A$ 의 비율  $\rho_2 = N_A/N$ 의 변화를 보여준다. 본 실험은 제안 기법에 의하여 국지적으로 커뮤니티를 구성하며 추정된 영향력이 높은 노드들이 전체 네트워크 내에서 영향력을 확산할 수 있음에 대한 검증을 목적으로 한다. 이전 실험에서 제시했던 결과대로 제안 기법은 높은  $\rho_1$ 값을 보이지만, 이 뿐만 아니라 제안 기법으로 추정된 영향력이 높은 노드들이 전체 네트워크 내의 노드들을 충분히 활성화시킬 수 있어야 제안 기법이 신뢰할 수 있는 성능을 나타낸다고 볼 수 있다.

Fig. 6 (a)와 6 (b)는 각각  $n = 1$ 일 때와  $n = 2$ 일 때의  $\rho_2$ 의 변화를 보여주며, Local은 제안 기법에 의하여 국지적으로 커뮤니티를 구성하면서 얻은  $\rho_2$ 의 변화를 나타내고, Global은 전체 네트워크 관점에서 얻은  $\rho_2$ 의 변화를 나타낸다.

Fig. 6 (a)에 볼 수 있듯이  $n = 1$ 일 때 Local의 경우 최종 시간에 전체 노드의 80%를 활성화시켰고 Global의 경우 최종 시간에 전체 노드의 90%를 활성화시켰다. 즉, 제안 기법으로 국지적 영향력이

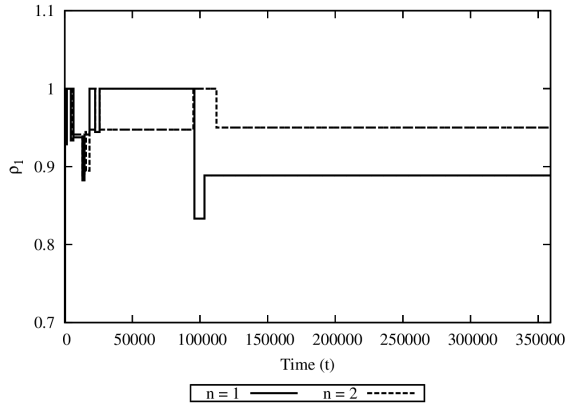


Fig. 5. Influence ratio of nodes selected by the proposed scheme (Intel dataset,  $N=20$ )

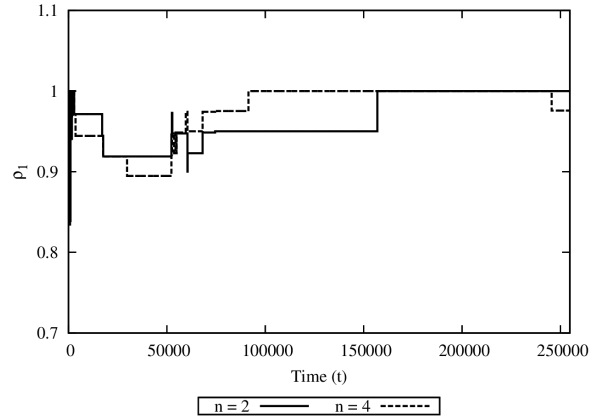
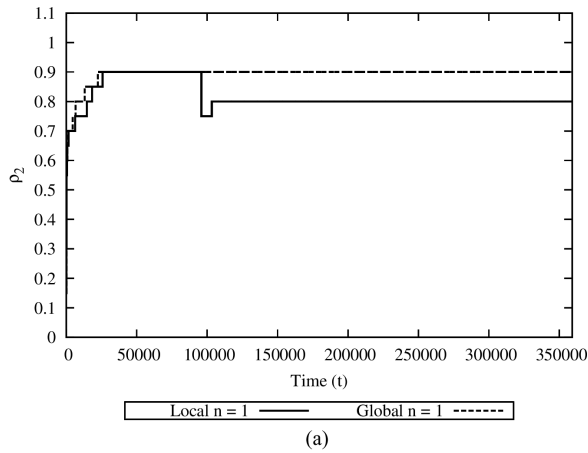
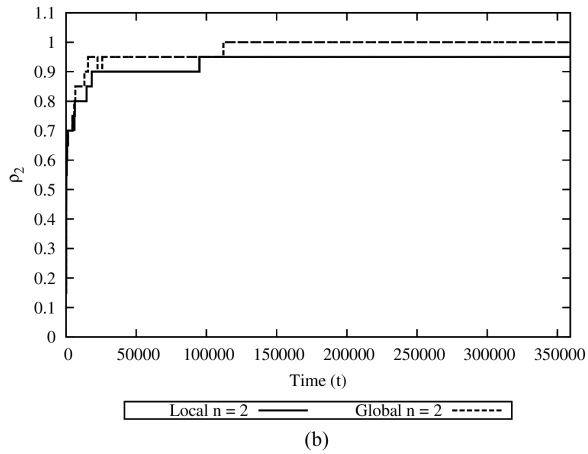


Fig. 7. Influence ratio of nodes selected by the proposed scheme (Infocomm05 dataset,  $N=41$ )



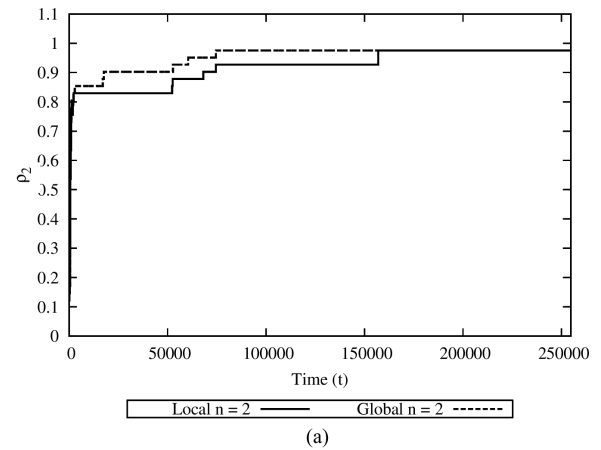
(a)



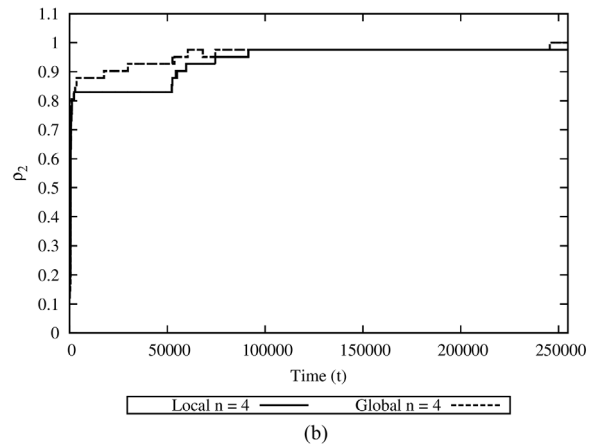
(b)

Fig. 6. Influence ratio of the nodes selected by the proposed scheme and the whole nodes in the network (Intel dataset)

높은 노드만 선별할 경우 실제 네트워크 내에서 활성화되는 노드 수는 낮아질 수 있다. 하지만,  $n=2$ 인 경우에는 Fig. 6 (b)에서 볼 수 있듯이 Local의 경우와 Global의 경우에 최종 시간에 활성화되는 노드 수에 대한 격차는 더 줄어든다 (Local의 경우 최종 시간



(a)



(b)

Fig. 8. Influence ratio of the nodes selected by the proposed scheme and the whole nodes in the network (Infocom05 dataset)

에 95%, Global의 경우 최종 시간에 100%). 결과적으로, 본 실험을 통해 제안 기법으로 추정된 영향력이 높은 노드들이 전체 네트워크 내에서 충분한 영향력을 행사할 수 있다.



Fig. 7은 Infocomm05 데이터 집합에서 단위 시간  $t$ 마다의  $\rho_1$ 의 변화를 보여준다. 영향도 측면에서 선별하는 상위 노드 개수인  $n$ 은 2개와 4개의 노드를 선택하였다. 최종 단위 시간  $t=275909$ 에서  $n=2$ 일 때는  $\rho_1$  값이 1.0이고  $n=4$ 일 때는  $\rho_1$  값이 0.98을 나타냈다. 즉, Infocomm05 데이터 집합에서의 실험 결과에서도 Intel 데이터 집합에서의 실험 결과와 유사하게 제안 기법으로 국지적 커뮤니티를 구성하여 선택된 노드들이 전체 네트워크 관점에서 선택된 노드들과 비교하여 전체 네트워크 노드들에게 행사할 수 있는 영향력이 거의 유사함을 알 수 있다.

Fig. 8 (a)와 8 (b)는 Infocomm05 데이터 집합에서 각 단위 시간  $t$ 마다 각각  $n=2$  일 때와  $n=4$ 일 때의  $\rho_2$ 의 변화를 보여준다. Fig. 8 (a)에서 볼 수 있듯이  $n=2$ 일 때 Local과 Global 모두 최종 시간에 전체 네트워크 노드의 97%를 활성화시켰다.

Fig. 8 (b)에서 볼 수 있듯이  $n=4$ 일 때는 Local의 경우 최종 시간에 전체 네트워크 노드의 97%를 활성화시켰고 Global의 경우 네트워크의 모든 노드를 활성화시켰다. 즉, Infocomm05 데이터 집합에서의 실험 결과 또한 제안 기법으로 선택한 노드가 네트워크 내에서 시간이 지남에 따라 충분히 영향력을 확산할 수 있음을 보여준다.

## V. 결 론

입소문 마케팅에 기반을 둔 영향력 전파 문제는 네트워크에 가장 많은 영향력을 행사할 수 있는 노드들을 찾아 전체 네트워크에 영향력을 최대화 하는 것이 목적이다. 본 논문에서는 DTN에서 효율적인 메시지 전파를 목적으로 각 노드와 노드 사이의 누적 접촉 기록에 기반으로 한 Independent Cascade 영향력 전파 모델을 적용하였다. 임의의 DTN에서 각 노드가  $k$ -CLIQUE 알고리즘으로 국지적 커뮤니티를 구성하고 커뮤니티에 추가되는 노드에 대해 의사 활성화를 시도하며 자신의 영향력을 추정하는 방법을 제안하였다. 실험을 통해 제안 기법으로 국지적 커뮤니티를 구성하여 추정한 영향력이 높은 노드가 전체 네트워크 관점에서도 영향력이 높음을 보였다. 본 연구의 결과는 DTN에서 임의의 소스 노드가 메시지를 전체 네트워크로 전파시키고자 할 때 영향력이 높은 노드 위주로 메시

지를 전달하여 전체적인 메시지 전파 시간을 단축하는 데에 기여를 할 수 있을 것으로 기대된다.

## 참 고 문 헌

- [1] D. Lopez-Pintado. "Diffusion in complex social networks," in *proc. Games and Economic Behavior*, vol. 62, no. 2, pp. 573-590, Mar, 2008.
- [2] D. Kempe, J. Kleinberg, and É. Tardos. "Maximizing the spread of influence through a social network," in *proc. ACM SIGKDD Int. Conf. Knowledge discovery and data mining*, pp. 137 - 146, Aug, 2003.
- [3] D. Gruhl, R. Guha, D. Liben-Nowell, and A. Tomkins. "Information diffusion through blogspace," In *proc. Int. Conf. World Wide Web*, pp. 491 - 501, May, 2004.
- [4] Y. Wang, G. Cong, G. Song, and K. Xie, "Community-based greedy algorithm for mining top-k influential nodes in mobile social networks," in *proc. ACM SIGKDD Int. Conf. Knowledge discovery and data mining*, pp. 1039-1048, Jul, 2010.
- [5] P. Hui, E. Yoneki, S. Y. Chan, and J. Crowcroft, "Distributed community detection in delay tolerant networks," in *proc. ACM Int. Workshop Mobility Evolving Internet Architecture*, pp. 7-14, Aug, 2007.
- [6] A. Chaintreau, P. Hui, J. Crowcroft, C. Diot, R. Gass, and J. Scott, "Impact of human mobility on opportunistic forwarding algorithms," *IEEE Trans. Mobile Compu.*, vol. 6, no. 6, pp. 606 - 620, Jun, 2007.
- [7] P. Hui, A. Chaintreau, J. Scott, R. Gass, J. Crowcroft, and C. Diot, "Pocket switched networks and human mobility in conference environments," in *proc. the ACM SIGCOMM workshop on Delay-tolerant networking*, pp. 244-251, Aug, 2005.
- [8] J. Leguay, A. Lindgren, J. Scott, T. Friedman, and J. Crowcroft, "Opportunistic content distribution in an urban setting," in *proc. 2006 SIGCOMM workshop on Challenged networks*, pp. 205 - 212, Sep, 2006.

- [9] P. Hui, J. Crowcroft, and E. Yoneki, "BUBBLE rap: social-based forwarding in delay tolerant networks," in *proc. MOHIHOC*, pp. 241 - 250, May, 2008.
- [10] Y.Kim, "Information Delivery Scheme by Using DTN in Battlefield Environment," *J. KICS*, vol. 36, no. 3, pp. 260-267, 3, 2011.
- [11] *CRAWDAD data set*, from <http://crowdad.cs.dartmouth.edu/>.

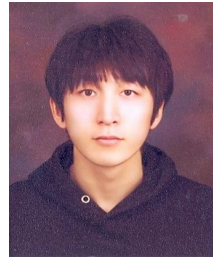
**김 찬 명 (Chan-Myung Kim)**



2009년 8월 한국기술교육대학교 멀티미디어공학과 졸업  
2011년 8월 한국기술교육대학교 컴퓨터공학과 석사  
2011년 8월~현재 한국기술교육대학교 컴퓨터공학과 박사 과정

<관심분야> DTNs, Social Networks

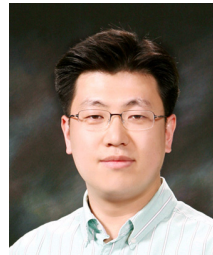
**김 용 환 (Yong-hwan Kim)**



2005년 3월 한국기술교육대학교 공학 학사  
2008 8월 한국기술교육대학교 정보미디어공학과 석사  
2010 8월 한국기술교육대학교 컴퓨터공학과 박사 수료  
<관심분야> Social Networks,

Future Internet

**한 연 희 (Youn-Hee Han)**



1996년 2월 고려대학교 수학과 (이학사)  
1998년 5월 고려대학교 컴퓨터학과(이학석사)  
2002년 2월 고려대학교 컴퓨터학과(이학박사)  
2002년 2월~2006년 2월 삼성

종합기술원 전문연구원

2006년 3월~현재 한국기술교육대학교 부교수

<관심분야> Mobility Management, Social Networks, Future Internet