

Delay-Tolerant Networks에서 영향력 추정의 분산 기법*

김찬명, 김용환, 한연희†

한국기술교육대학교 첨단기술연구소 지능네트워크연구소
e-mail:{cmdr, cherish, yhhan}@koreatech.ac.kr

An Influence Estimating Distributed Scheme in Delay-Tolerant Networks

Chan-Myung Kim, Yong-hwan Kim, Youn-Hee Han †

Laboratory of Intelligent Networks,
Advanced Technology Research Center

요 약

사회관계망에서 영향력 전파 문제는 네트워크에 가장 영향력을 끼칠 수 있는 노드들을 찾아 전체 네트워크에 영향력을 최대화 하는 것을 목적으로 한다. 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 매우 어렵다. 본 논문에서는 Delay-Tolerant Networks 환경에서 분산 방식으로 각자 노드가 k -Clique 구조로 커뮤니티를 구성하여 국지적 정보 (Local Information)만을 활용하여 자신의 영향력을 추정하는 방법을 제시하고 실험을 통해 제안 기법으로 산출한 노드들의 영향력이 전체 네트워크 관점에서 산출한 노드들의 영향력에 근접함을 실험을 통해 증명한다.

1. 서론

최근 사회 관계망 (Social Network)에서 영향의 확산 (Influence Propagation)에 대한 연구가 진행되고 있다 [1-4]. 영향은 새로운 아이디어, 정보, 소문의 전파로 인해 개인의 상태나 성질이 변화하는 것을 뜻하며 영향을 받은 개인은 자신과 교류하는 다른 상대에게도 영향을 주고 영향은 확산되어 네트워크 내에서 퍼져 나간다. 사회 관계망 네트워크에서 영향력 최대화 문제는 입소문 마케팅 전략에 기반을 둔 것으로 전체 네트워크에 영향력을 최대화시킬 수 있는 노드의 집합을 찾아내는 문제이다 [3].

한편, 컴퓨팅 환경의 변화에 따라 전통적인 유선망 네트워크뿐만 아니라 이동성을 지닌 노드들의 무선 모바일 네트워크에 대한 관심이 증대되고 있다. 특히, Delay-Tolerant Networks (DTNs) [5-8]는 유선망과 같이 노드와 노드가 항상 통신 가능한 상태에 있음이 보장되지 않는 네트워크로서 정보의 전달은 노드와 노드가 만날 때만 기회적으로 (Opportunistically) 이루어진다. 그러

므로, 메시지 전달 지연시간이 예측할 수 없이 길어질 수 있는 특성을 지니고 있으며, 메시지 전달을 위한 가장 적합한 노드를 선정하기 위하여 각 노드의 영향력을 측정하고 그러한 영향력이 높은 노드 위주로 메시지를 전달하여 전체적인 메시지 전달 지연시간을 단축하려는 연구가 시도되고 있다. 하지만, Delay-Tolerant Networks 환경에서는 전체 네트워크 정보를 정확히 알 수 없기 때문에 노드의 영향력을 정확히 측정하는 것은 매우 어렵다.

[5]는 Delay-Tolerant Networks에서 노드와 노드 사이의 접촉기록에 따라 각 노드가 자기 자신의 자신의 국지적 커뮤니티 (Local Community)를 구성하는 3가지 알고리즘 (SIMPLE, k -CLIQUE, MODULARITY)을 제시했다. 여기서 k -CLIQUE 방식은 k 개의 노드로 이루어진 무방향 그래프에서 노드 모두가 직접적으로 연결되어 있는 구조인 k -Clique 구조를 활용한다. k -Clique 구조를 지닌 커뮤니티에 속한 각 노드의 차수는 $k-1$ 을 유지한다.

노드는 자신과 전혀 교류가 없는 노드보다 자신이 속한 커뮤니티 내의 노드들에게 많은 정보와 영향력을 전파할 수 있다. 완전그래프 형태로 응집력이 높은 k -Clique 구조에서 커뮤니티 내의 노드들에게 영향력을 더 발휘할 수 있다.

본 논문에서는 Delay-Tolerant Networks 환경에서

* 본 논문은 2012년도 정부(교육과학기술부)의 재원으로 한국연구재단의 지원을 받아 수행된 연구 (2012-0002558)이며 또한 지식경제부와 한국산업기술진흥원의 전략기술 인력양성사업으로 수행된 결과입니다.

† 한연희: 교신저자

k -CLIQUE 커뮤니티 구성 알고리즘을 통해 분산 방식으로 각자 노드가 스스로 자신의 영향력을 추정하는 방법을 제시한다. 노드 각각이 자신의 국지적 커뮤니티에 미치는 영향력으로 전체 네트워크에 파급될 영향력을 추정한다. 자신의 커뮤니티에 새로운 노드를 추가할 때마다 가상으로 활성화 시도를 하고 이러한 가상 활성화의 성공 여부로 자신의 영향력을 추정한다. 여러 노드와 교류고 높은 활동성을 지니면서 커뮤니티를 구성하는 노드들이 많은 노드에게 영향력을 끼칠 확률 또한 증가한다. 실험을 통해 제안 기법으로 산출한 노드들의 영향력이 전체 네트워크 관점에서 산출한 노드들의 영향력에 근접함을 실험을 통해 증명한다.

2. 네트워크 및 영향 전파 모델

1. Delay-Tolerant Networks (DTNs) 모델

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 일컫는다. 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 정해놓은 수 (예를 들어, 3번)만큼의 hello 메시지를 수신하지 못할 경우 노드 v_i 는 노드 v_j 가 떠났다(Leave)고 판단한다. 만약 α 가 임의의 노드와 접촉한 시각이고 β 가 접촉했던 노드가 떠난 시각이라면 $\beta - \alpha$ 를 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.

2. Independent Cascade 모델

Independent Cascade 모델 [1]은 여러 사회 관계망 연구에서 널리 쓰이는 동적 영향 전파 모델로 각 링크를 통해 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. 이 모델은 n 개의 초기 활성 상태 노드로부터 확산이 시작되며, 활성 상태의 노드는 다른 비활성 상태의 노드를 활성화시킬 수 있다. 비활성 상태에서 활성화상태로 전환할 수는 있으나 반대로

활성상태에서 비활성 상태로의 전환은 불가하다.

활성 상태의 노드 v_i 가 시간 t 에 비활성 상태의 노드 v_j 에 대해 $p(i, j)$ 의 확률로 활성화를 성공할 수 있고 활성화에 성공할 경우 시간 $t+1$ 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.

본 논문에서는 각각의 노드 v_i 가 노드 v_j 를 활성화할 확률 $p(i, j)$ 를 다음과 같이 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 접촉시간을 뜻한다. 즉, $p(i, j)$ 는 그러한 누적 접촉시간에 대하여 v_j 와의 누적 접촉시간 비율을 의미한다. 만약 v_j 와 많은 접촉 시간을 기록하고 있다면 $p(i, j)$ 값은 높아질 것이며, 그 만큼 v_i 입장에서 v_j 에 대한 활성화 성공률도 높아진다.

3. 제안 기법

제안 기법은 [5]에서 제시된 분산 방식 알고리즘을 토대로 k -CLIQUE 커뮤니티를 구성하며 각각의 노드가 독립적으로 수행하게 된다. k -CLIQUE 커뮤니티 구성 방법에서 활용하는 k -Clique 구조는 형성하는 기준이 엄격하고 구조 내의 노드끼리 응집력이 높은 특성을 지닌다. k -Clique 구조 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 방식으로 커뮤니티 구성을 수행하여 각각의 노드의 국지적인 영향력을 측정하고 전체 네트워크 시점에서의 영향력을 추정할 수 있다.

각각의 노드는 누적 접촉시간을 기준으로 자신과 가장 많이 접촉한 노드들을 Familiar Set으로 유지한다고 가정하며, 노드 v_i 가 노드 v_j 를 활성화할 확률 $p(i, j)$ 가 γ 보다 클 때 노드 v_i 는 노드 v_j 를 자신의 Familiar Set F_i 에 포함시킨다. k 값은 커뮤니티를 구성하는 기본 단위인 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 (그림 1 참조). 의사활성화는 Independent Cascade 모델을 따르며 새로 추가되는 v_j 의 인접노드들은

$p(l, j)$ 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 성공하면 v_i 는 자신의 N_i^C 를 업데이트한다 (즉, $N_i^C = N_i^C + 1$).

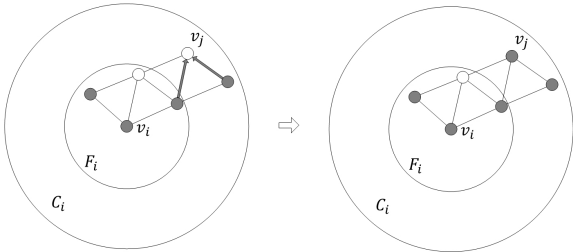


그림 1. 새로 추가된 v_j 에 대해 인접 노드들의 의사활성화 시도

제안 알고리즘에 사용된 기호는 다음 <표 1>과 같다.

<표 1> 알고리즘 기호 설명

기호	설명
F_i	노드 v_i 의 Familiar Set
\widetilde{F}_{ij}	노드 v_i 의 이웃노드 v_j 에 대한 추정 Familiar Set
C_i	노드 v_i 의 국지적 커뮤니티
Π_i	노드 v_i 의 국지적 커뮤니티 C_i 안의 각 노드들의 Familiar Set을 원소로 지니는 집합 (즉, $\Pi_i = \{\widetilde{F}_{ij} \mid v_j \in C_i\}$)
γ	Familiar Set 임계값
N_i^C	노드 v_i 가 자신의 국지적 커뮤니티 내에서 의사 활성화(pseudo-activation)시킬 수 있는 노드 수
k	Clique의 크기

각각의 노드 v_i 는 독립적으로 제안 알고리즘을 수행하며, 노드가 DTN에 참여할 때에 다음과 같은 초기화 절차를 수행한다.

Initialization

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

다음 Procedure 1은 임의의 v_i 가 다른 노드 v_j 와 접촉이 되었을 때 마다 수행한다.

Procedure 1

1. v_j 의 C_j, F_j, Π_j 정보를 획득한다.
2. $v_j \notin C_i$ 일 때, 노드 v_j 의 F_j 와 C_i 가 공통으로 최소 $k-1$ 개의 노드를 공유 (즉, $|F_j \cap C_i| \geq k-1$) 하고 있다면 노드 v_j 를 C_i 에 추가한다.

2-1. C_i 에 속한 노드들 중 v_j 에 인접한 활성화 노드들은 v_j 에 대해 의사활성화를 시도한다.

2-2. Procedure 3을 수행한다.

다음 Procedure 2는 임의의 v_i 가 다른 노드 v_j 와 접촉 이후 v_j 가 떠날 때 마다 수행한다.

Procedure 2

1. $cd_{ij}, cd_{\max}(i)$ 를 업데이트 한다.
2. v_i 가 접촉한 각각의 노드 v_j 에 대해서 $p(i, j)$ 를 업데이트 한다.
3. v_i 가 접촉한 각각의 노드 v_j 에 대해서 만약 $p(i, j) \geq \gamma$ 이면 v_i 는 v_j 를 자신의 F_i 와 C_i 에 추가한다.
 - 3-1. C_i 에 속한 노드들 중 v_j 에 인접한 활성화 노드들은 v_j 에 대해 의사활성화를 시도한다.
 - 3-2. Procedure 3을 수행한다.
4. v_i 가 접촉한 각각의 노드 v_j 에 대해서 $p(i, j) < \gamma$ 이면 v_i 는 v_j 를 자신의 F_i 에서만 삭제한다.

Procedure 3

1. v_j 가 C_i 에 추가될 경우 Π_j 정보를 이용하여 C_j 안의 노드들을 C_i 에 추가할지 다음과 같이 판단한다. C_j 안의 각각의 노드 v_m 에 대해서 만약 \widetilde{F}_{jm} 과 C_i 가 공통으로 최소 $k-1$ 개의 노드들을 공유 (즉, $|\widetilde{F}_{jm} \cap C_i| \geq k-1$) 한다면 v_m 를 부가적으로 C_i 에 추가한다.
2. C_i 에 속한 노드들 중 v_m 에 인접한 활성화 노드들은 v_m 에 대해 의사활성화를 시도한다.

4. 실험 및 성능 분석

<표 2> 실험 데이터 집합

트레이스	Intel	Infocomm05
디바이스	iMote	iMote
프로토콜	Bluetooth	Bluetooth
기간(일)	5	3
노드 개수	20	41
총 접촉 횟수	560	22459
평균 접촉 횟수	0.084	4.6

제안 기법을 평가하기 위하여 본 장에서는 많은 DTN 알고리즘 제안 연구에서 활용하는 실제 노드의 접촉 기록이 있는 Hagle 프로젝트 데이터 집합[9] 중 'Intel'과 'Infocomm05' 데이터 집합으로 실험을 하였으며 각 데이터 집합의 정보는 <표 2>와 같다.

주어진 데이터 집합에 기반하여 각각의 노드가 제안 기

법으로 국지적 커뮤니티를 구성하도록 하면서 매 단위 시간마다 각각의 노드 v_i 에 대하여 영향력을 측정한다. 또한, 각 노드 v_i 가 국지적 커뮤니티 내에서 활성화할 수 있는 노드 수 N_i^C 와 전체 네트워크에서 활성화할 수 있는 노드 수 N_i 를 측정한다. N_i 가 높은 순으로 선택한 n 개의 노드들이 전체 네트워크에서 활성화할 수 있는 노드 수에 대해 N_i^C 가 높은 순으로 선택한 n 개의 노드들이 전체 네트워크에서 활성화할 수 있는 노드 수의 비율인 ρ 를 측정하고 시간 t 에 따른 변화를 분석한다. 즉, ρ 값이 높을수록 제안 기법으로 산출된 노드들의 영향력이 전체 네트워크 관점에서 산출한 노드들의 영향력에 근접하다는 것을 의미한다. Familiar Set 임계값인 γ 는 0.3을 사용하였다.

<표 3> Intel 데이터 집합 실험 결과

n 의 비율	0.05	0.1
평균 ρ 값	0.914	0.952
최종적으로 측정된 ρ 값	0.888	0.95
최종 시간 활성화 노드 비율	0.8	0.95

<표 3>은 Intel 데이터 집합에서의 제안 기법의 실험 결과를 나타낸다. n 은 전체 네트워크 노드 수에 대해 각각 0.05, 0.1의 비율로 적용하였다. n 의 비율이 0.05일 때 평균 ρ 값은 0.914이고 n 의 비율이 0.1일 때 평균 ρ 값은 0.952를 나타냈다. 최종시간 $t=359312$ 에는 n 의 비율이 0.05일 때 ρ 값이 0.888이고 n 의 비율이 0.1일 때 ρ 값은 0.95를 나타냈다. 또한, n 의 비율이 0.05일 때 최종 시간에 전체 네트워크 노드 수의 80%에 해당하는 노드를 활성화시켰고 n 의 비율이 0.1일 때는 95%에 해당하는 노드를 활성화시켰다. 즉, 제안 기법으로 선택한 노드들이 시간이 지남에 따라 네트워크 내의 대부분의 노드에 영향력을 확산할 수 있었고 국지적 커뮤니티를 구성하여 추정된 영향력 높은 노드들이 전체 네트워크 관점에서 영향력이 높은 노드들과 영향력 전파 능력 관점에서 거의 유사한 영향력을 가짐을 알 수 있다.

<표 4> Infocomm05 데이터 집합 실험 결과

n 의 비율	0.05	0.1
평균 ρ 값	0.966	0.977
최종적으로 측정된 ρ 값	1.0	0.975
최종 시간 활성화 노드 비율	0.975	0.975

표 4는 Infocomm05 데이터 집합에서의 제안 기법의 실험 결과를 나타낸다. Infocomm05 데이터 집합에서의 실험 결과도 전체적으로 Intel 데이터 집합에서의 실험 결과와 유사함을 알 수 있다.

5. 결론

본 논문에서는 이동 노드들로 구성된 DTN에서 각 노드가 k -Clique 구조로 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 노드들이 전체 네트워크 관점에서도 영향력이 높음을 보였다. 본 연구의 결과는 DTN에서 메시지를 전달할 때 영향력이 높은 노드 위주로 메시지를 전달하여 전체적인 메시지 전달 지연시간을 단축하는 데에 기여할 수 있을 것으로 기대된다.

참고문헌

- [1] D. Lopez-Pintado, "Diffusion in complex social networks," Journal of Economic Literature, pages 573-590, 2004.
- [2] Y. Wang, G. Cong, G. Song, K. Xie, "Community-based Greedy Algorithm for Mining Top-K Influential Nodes in Mobile Social Networks," Proc. of the 16th ACM SIGKDD, pp.1039-1048, 2010.
- [3] D. Kempe, J. Kleinberg, É. Tardos. "Maximizing the spread of influence through a social network," Proc. of the 9th ACM SIGKDD, pp 137 - 146, 2003.
- [4] D. Gruhl, R. Guha, D. Liben-Nowell, A. Tomkins. "Information diffusion through blogspace," Proc. of the 13th international conference on World Wide Web, pp. 491 - 501, 2004.
- [5] P. Hui, E. Yoneki, S.-Y. Chan, J. Crowcroft, "Distributed Community Detection in Delay Tolerant Networks," Proc. of ACM Int. Workshop Mobility Evolving Internet Architecture, 2007.
- [6] A. Chaintreau, P. Hui, J. Crowcroft, C. Diot, R. Gass, J. Scott, "Impact of human mobility on the design of opportunistic forwarding algorithms," Proc. of INFOCOM, April 2006.
- [7] P. Hui, A. Chaintreau, J. Scott, R. Gass, J. Crowcroft, C. Diot, "Pocket switched networks and human mobility in conference environments," Proc. of WDTN, 2005.
- [8] P. Hui, J. Crowcroft, E. Yoneki, "BUBBLE Rap: Social-Based Forwarding in Delay Tolerant Networks," Proc. of MOHIHOC, 2008.
- [9] CRAWDAD data set, <http://crawdad.cs.dartmouth.edu>