Journal of Korea Multimedia Society Vol. 22, No. 6, June 2019(pp. 684-694)
https://doi.org/10.9717/kmms.2019.22.6.684

네트워크에서 퍼진 정보의 근원에 대한 Voronoi 추정방법

최 재 영†

Finding the Information Source by Voronoi Inference in Networks

Jaeyoung Choi†

ABSTRACT

Information spread in networks is universal in many real-world phenomena such as propagation of infectious diseases, diffusion of a new technology, computer virus/spam infection in the internet, and tweeting and retweeting of popular topics. The problem of finding the information source is to pick out the true source if information spread. It is of practical importance because harmful diffusion can be mitigated or even blocked e.g., by vaccinating human or installing security updates. This problem has been much studied, where it has been shown that the detection probability cannot be beyond 31% even for regular trees if the number of infected nodes is sufficiently large. In this paper, we study the impact of an anti-information spreading on the original information source detection. We consider an active defender in the network who spreads the anti-information against to the original information simultaneously and propose an inverse Voronoi partition based inference approach, called *Voronoi Inference* to find the source. We perform various simulations for the proposed method and obtain the detection probability that outperforms to the existing prior work.

Key words: Information Source Detection, Epidemic Diffusion Model, Voronoi Inference,

1. 서 론

최근 인터넷의 급속한 발전으로 인해 복잡한 네트워크에서 다양한 정보들이 전파되고 있다. 이런 정보 확산 현상은 삼성이나 애플과 같이 자신의 회사의 이득을 위해 새로운 제품을 되도록 많은 사람에게 빨리 알리고자 전파되는 경우도 있고 혹은, 정치인이나 유명인들이 자신의 신조를 사람들에게 알리는 목적으로 정치적인 견해가 전파되기도 한다. 하지만, 이와 동시에 Fig. 1과 같이 악성루머나 바이러스처럼 사람들에게 혹은 컴퓨터에 좋지 않는 정보도 이런 인터넷망을 타고 급속도로 퍼져나가기도 한다. 이는 인터넷망에서 익명성이 보장되는 경우에 대해서 일부 사람들이 이를 악용하여 발생하는 현상 중 하나이다. 루머나 악성코드처럼 좋지 않은 정보가 퍼지는 상황을 제어하는 방법은 크게 두 가지로 구분이 된다. 하나는 네트워크에 다양하게 연결된 여러 경로를 찾아서 긴급하게 이를 차단하는 방법이고 또 다른 하나는 정보를 처음 퍼뜨린 근원(Source)을 찾아서 더 퍼지지 않도록 막는 방법이다. 전자의 경우, 현재 거대하게 생성된 복잡한 네트워크에서 이미 흘러가고 있는 정보를 쉽게 차단하는 것은 어려움이 있다.

※ Corresponding Author : Jaeyoung Choi, Address: (62399) Eodeung-daero, Gwangsan-gu, Gwangju, Korea, TEL : +82-62-940-5436, FAX : +82-62-940-5436, E-mail : jychoi@honam.ac.kr

† Dept. of Automotive Engineering, Honam University
Receipt date : Dec. 28, 2018, Revision date : Apr. 18, 2019
Approval date : May 16, 2019

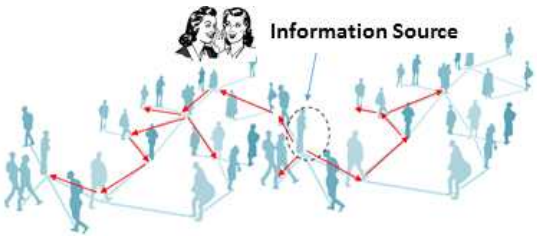

Fig. 1. Information Diffusion in Social Network.

따라서 이 문제와 관련하여 주로 처음 퍼뜨린 근원을 먼저 찾아내는 방법에 대한 연구가 전 세계적으로 많이 진행되었다.

정보의 근원을 찾는 문제를 처음 다뤘던 연구인 [1,2]에서는 현존하는 가장 좋은 방법으로 알려진 추정법 중 하나로써 최우추정량(Maximum Likelihood Estimator; MLE)을 사용하였다. 그 결과, 네트워크에 아주 많은 노드가 이 정보를 듣게 되면[1], 즉 감염되면 처음 정보를 퍼뜨린 근원을 찾을 수 있는 확률이 모든 노드가 같은 수의 이웃 노드를 가진 비교적 간단한 가장 간단한 정규트리(Regular tree)에서도 31%를, 그리고 일반적인 그래프에서는 10%를 넘지 못하는 것이 밝혀졌다. 이는 거대한 네트워크에서 매우 많은 노드가 근원이 될 수 있는 후보로 있는 경우에는 나쁘지 않은 결과이다. 이후 연구로서, 이를 더 높이고자 다른 여러 방법이 제안이 되었다. 그중 하나로서, 본 논문에서는 루머와 같은 근원이 퍼뜨리는 정보의 반대 정보를 네트워크에 있는 방어자(Defender)가 동시에 퍼뜨리는 방법을 고려하고 있다. 즉, 두 가지 다른 정보가 퍼진 현상을 관측한 후 적절한 추론 방법을 사용하면 얼마나 더 원래 정보의 근원에 대해서 잘 찾을 수 있는지에 대해 살펴보았다. 이는 다시 말하면 정보의 근원을 보다 더 잘 찾기 위하여 추가적인 다른 정보를 사용한 것이다. 예를 들면, 바이러스가 퍼지는 경우에도 백신과 같은 것이 동시에 퍼지게 되는 현상도 있고 또한 루머와 같은 한쪽으로 치우쳐진 정보만 퍼지는 것이 아니라 반대의 정보도 같이 퍼질 수 있는 상황이 있을 수 있다는 것에 착안하여 제안된 방법이다. 특히, 본 연구는 선행연구에 대한 확장된 연구로서 [3]에서 제안된 알고리즘이 가진 한계를 해결하기 위해 제안된 논문이다. 구체적인 내용은 3장에서 다루도록 한다.

1) 일반적으로 이 경우를 감염(Infection)되었다고 표현한다.

본 논문의 구성은 다음과 같다. 먼저 2장에서는 본 논문과 관련된 여러 선행 연구들에 대한 소개를 비롯한 그 이후로 새롭게 제기된 다른 문제들에 대해서 간략히 소개한다. 3장에서는 본 논문에서 고려하고 있는 정보 확산에 대한 수학적인 모델을 자세히 기술하고 이런 모델에서 원래의 정보의 근원을 잘 찾을 수 있는 새로운 방법에 대해 제안한다. 그리고 4장에서는 본 연구가 가진 접근 방법에 대한 방어자 선택에 대한 두 가지 알고리즘을 제안하고 5장에서는 이런 알고리즘 기법으로 새롭게 제안된 방법이 정보의 근원을 얼마나 잘 찾을 수 있는지에 대한 다양한 네트워크 환경에서의 실험을 보여준 후, 마지막 장에서 논문에 대한 전체적인 요약을 하며 결론을 맺는다.

2. 관련 연구

2.1 단일 소스(Single Source) 추론

네트워크에서 정보가 퍼진 후 근원을 찾는 문제는 [1,2]에서 처음 제안이 되었다. 여기서는 소셜 네트워크뿐 아니라 일반적인 컴퓨터 네트워크에서 발생할 수 있는 바이러스나 루머와 같은 정보가 퍼져나가는 것을 고려하였다. 그리고 충분히 많은 시간 동안 정보가 퍼져나간 후에 네트워크 관리자가 이를 찾는 방법으로써 최우추정량을 제안하였다. 이는 네트워크 관리자는 정보가 언제 퍼지기 시작했는지에 대한 정보가 없이 다만, 임의의 시점에 네트워크를 관측했을 때, 확산된 형태(Snapshot)만 보고 그 근원에 대해 추론하는 방법이다. [1,2]에서는, 정보가 연결된 각 이웃들에게 확률적으로 퍼져나가는 모델을 고려하였기 때문에 위와 같은 확률에 기반 한 추정량을 사용하였다. 하지만, 해당연구에서는 그래프가 가장 다루기 쉬운 정규트리구조인 경우에만 최우추정량을 유한시간에 계산하는 것이 가능하다는 사실을 수학적으로 밝혔다. 이를 토대로 루머 중심(Rumor center)이라고 하는 새로운 그래프 중심에 대한 개념을 만들어 냈다. 이것은 다름이 아니라, 각 노드를 근원으로 가정한 경우, 현재의 정보가 퍼진 모양이 나타나는 경우의 수를 가장 많이 가진 노드를 말한다. 이를 바탕으로 최우추정량을 사용한 경우, 정규트리에서 원래 근원을 얼마나 잘 찾아낼 수 있는지에 대해서 이론적인 결과를 얻었다. 즉, 차수(Degree)가 2인

직선 그래프인 경우에는 시간이 충분히 지난 후에는 최우추정량으로 실제 근원을 찾을 확률이 0이 되지만, 차수가 3 이상인 경우에는 시간이 아무리 많이 지나고 네트워크에 감염된 노드가 많아져도 0보다는 큰 확률로 찾아낼 수 있다는 것을 증명하였다[2]. 하지만, 최우추정량만을 사용한 방법은 정보의 근원을 발견할 확률이 0.31을 넘지 못한다는 근본적인 한계도 밝혀냈다.

그 후에 이 연구를 기반으로 근원을 찾을 확률을 높일 수 있는 여러 가지 방법들이 제안되었다. [3]에서는 본 논문에서와 같이 원래 정보에 반대되는 정보를 퍼뜨림으로써 그 근원을 찾기 위해 두 가지 다른 정보가 같이 퍼지는 현상을 처음으로 제안하였다. 하지만, 이 논문에서는 제안하고 있는 네트워크에 보호자(Protector)라고 하는 반대의 정보를 퍼뜨리는 노드가 원래 정보를 그 이웃으로부터 전파가 된 후[3] 반대의 정보를 퍼뜨리는 한계적인 상황만을 고려하고 있다. 이 경우에서는 최우추정량이라는 가장 우수한 추정 방법을 사용해도 반대정보의 확산이 원래 정보의 근원을 찾는데 시간이 충분히 지나면 전혀 도움이 되지 않는다는 사실을 밝혀냈다. 본 논문에서는 이 연구를 기반으로 원래 정보를 듣고 나서 다른 정보를 퍼뜨리는 것이 아니라 임의의 시간에 동시에 퍼뜨리게 되면 얼마나 원래 정보의 근원을 찾는 데 도움이 되는지를 처음으로 분석한 논문이다. [4]에서는 네트워크에서 전체 감염된 노드를 근원에 대한 후보로 보는 것 대신에 이전에 미리 정보에 근원에 대한 사전 지식(Prior information)이 있는 경우를 고려하였다. 즉, 모든 감염된 노드를 고려하지 않고 일정한 후보 그룹에서 반드시 근원이 있다고 가정을 하였다. 저자들은 이 경우 사후추정량(Maximum a Posterior Estimation)을 사용했을 때, 근원을 찾을 확률이 1/2이 넘기도 하고 어떤 경우에는 확률 1로서 근원을 찾을 수 있다는 사실을 이론적으로 증명하였다. 하지만, 이것은 근원에 대한 기존의 정보를 알기 어려울 수 있다는 한계가 있었다. [5]는 네트워크에 정보가 퍼진 후에 관리자가 각 노드에게 추가적인 질문(누가 너에게 이 정보를 알려 주었는지 등)을 함으로써 근원을 파헤쳐 나가는 새로운 방법을 제안하였다. 실제 관리자가 이렇게 물어보는 경우 발생할 수 있는 비용을 고려하여 주어진 자산(Budget)에서 얼마나 이런 추가적인 질문들을 사용하면 더 잘 찾아내는지를 이론적으로 밝혔다.

2.2 다중 소스(Multi Sources) 추론

앞 장에서는 주로 네트워크에 정보의 근원이 하나인 경우 그것을 추론하는 방법에 대한 연구였다면, [6,7,8]에서는 이보다 더 일반적인 경우로 여러 노드가 동시에 루머와 같은 정보를 퍼뜨리게 될 때, 이런 근원들의 집합을 잘 찾을 수 있는 방법들에 대해서도 살펴보았다. [6]에서는 네트워크에 k개의 근원이 있다고 가정하고 시간이 충분히 지난 후에 모든 가능한 근원들의 집합을 고려하여 가장 그럴듯한 후보를 찾고 이에 대한 실제로 찾을 확률에 대해 분석하였다. [7]에서는 같은 종류의 정보이지만 각각 다른 근원들이 다른 시간에 퍼뜨리는 일반적인 경우에 대해서도 모든 근원을 찾아낼 수 있는 추론 방법을 제안하였고 이것의 성능을 실험적으로 얻어냈다. [8]에서는 정보가 퍼져나가는 모델이 알려지지 않은 경우 이를 적절하게 학습하여 다중 근원을 찾는 방법을 제안하였다. 그리고 [9]에서는 네트워크에 같은 정보를 퍼뜨리는 근원들의 수에 대한 정보가 추가로 주어졌을 때 모든 근원을 동시에 찾아내는 방법을 제안하였다.

2.3 근원 숨기기 및 찾기 (Seeking and Hiding Problem)

마지막으로, 정보의 근원을 네트워크 관리자의 입장에서 얼마나 잘 찾아낼 수 있는가와 동시에 실제 근원의 입장에서 얼마나 자신을 잘 숨길 수 있는지에 대한 연구도 진행이 되었다. 특히, [10,11]에서는 자신의 정보를 보다 더 잘 숨기기 위한 방법으로 적응형 확산(Adaptive Diffusion)이라고 하는 새로운 확산 방법을 제안하였다. 저자들은 이 경우 네트워크 관리자가 사용할 수 있는 가장 좋은 최우추정량을 가지고 찾는다고 해도 근원을 잘 찾지 못한다는 것을 이론적으로 밝혔다. [12]에서는 근원과 네트워크 관리자를 각각 하나의 게임의 주체(Player)로 두고 근원은 정보가 퍼져나가는 비율을 조절함으로써, 그리고 관리자는 퍼진 노드 중 일정한 크기의 집단을 선택함으로써 각각의 주체들이 자신의 유익(Payoff)을

2) 본 논문에서는 이 결과에 대한 수학적인 Closed form을 제공하고 있다.

3) 여기서, 저자는 이를 소극적 보호자(Passive protector)라고 정의하고 있다.

Table 1. Taxonomy of Information Source Finding Problems

	Single Source	Multiple Sources
Snapshot	[1],[2]	[6],[7],[8]
Snapshot + Additional Information	[3],[4],[5] + this paper	[9]
Seeking and Hiding	[10],[11],[12]	Not yet

최대화 하는 게임이론(Game Theory)을 적용하여 분석하였다. 그리고 이런 게임모델에서 시간이 충분히 지난 후 궁극적으로 평형을 이루게 되는 내쉬 균형(Nash Equilibrium)은 어떻게 표현이 되는지를 분석하였다.

본 논문에서 고려하는 방법이 다른 최근 논문들과 어떤 관련성이 있는지를 Table 1에서 나타내었다. 즉, 본 연구는 기존의 정보가 어떻게 퍼져있는가에 대한 확산 형태에 더하여 다른 종류의 정보를 퍼뜨림으로써 얻게 되는 추가적인 정보(Additional Information)를 사용하게 되는 경우 근원에 대한 탐지 확률이 얼마나 더 증가하는지 살펴본 연구 중 하나이다. 특히, 최근 관련논문[3]에서 제안된 기존정보를 듣게 되면 반대정보를 퍼뜨리는 소극적인 보호자 방법과는 다르게 본 연구에서는 네트워크에서 방어자가 있는 경우 기존정보와 동시에 반대의 정보를 퍼뜨리는 경우 기존정보의 근원을 어떻게 찾을 수 있는지에 대한 방법을 제안하였다.

3. 시스템 모델 및 제안된 Voronoi 추정법

3.1 정보의 확산 모델(Diffusion models)

본 논문에서 고려하고 있는 네트워크는 $G=(V,E)$라는 그래프로 표기될 수 있는데, 여기서 V는 네트워크에 있는 모든 노드(Node)의 집합이고 E는 각 노드를 연결하는 변(Edge)들의 집합이다. 여기서 노드란, 소셜 네트워크에서 한 사람의 개체를 말하기도 하고 혹은 인터넷망에서 한 컴퓨터를 말하기도 한다. 그리고 이런 노드들을 서로 연결하는 것이 그래프에서 변으로 표현이 되는데 이는 각 노드의 소셜 연관성(Relationship) 혹은 각 컴퓨터의 연결성을 의미한다. 본 연구에서는 기존의 다른 논문들에서 가정한 것과 같이 정보가 퍼져나가는 상황에서 네트워크의 경계에서의 영향을 피하기 위해 노드가 충분히 많이 있고 네트워크에 있는 모든 노드가 하나로 연결된(Connected) 경우를 고려한다. 본 논문에서는 앞서 언급한 바와 같이 네트워크에 두 가지 정보에 대한 근원이 있다. 하나는 루머와 같은 원래 정보를 퍼뜨리는 근원이고 다른 하나는 이런 정보가 퍼지는 것을 대비해서 그 반대의 정보를 퍼뜨리는 근원이다. 전자를 저자는 정보의 근원이라고 표현하고 후자를 반대 정보의 근원(Defender: 방어자)이라고 표현하겠다. 그리고 각각을 s^* 와 p^*로 표현하도록 한다. 그런 후, Fig. 2와 같이 네트워크에서 두 가지 정보가 각각의 근원으로부터 그 이웃에게 전달되어서 퍼져나가게 되는데, 여기서 정보가 퍼지는 현상은 이미 잘 알려진 확산 방법 중 하나인 Susceptible-Infected (SI) 모델을 따른다고 가정한다. 즉, 이것은 한 노드가 이미 감염된 상태라고 하면 이와 연결된 이웃 노드는 비율이 $\lambda>0$인 지수분포(Exponential distribution)를 따라서 정보가 확률적으로 퍼져나가는 확산모델이다. 그리고 이와 반대의 정보가 퍼져 나가는 경우도 비슷한 방법으로 퍼진다고 가정한다. 즉, 그림과 같이 임의의 노드가 보호된(Defended) 상태[4]라고 하면 이 노드는 그 이웃에게 같은 비율 λ로 방어자가 준 정보를 지수분포로 퍼뜨린다고 가정한다. 실제로는, 퍼지는 비율은 네트워크의 환경이나 정보의 긴급성 등에 따라서 각 노드마다 다를 수 있으나, 좀 더 용이한 분석을 위해서 위와 같이 동일한 비율로 퍼져

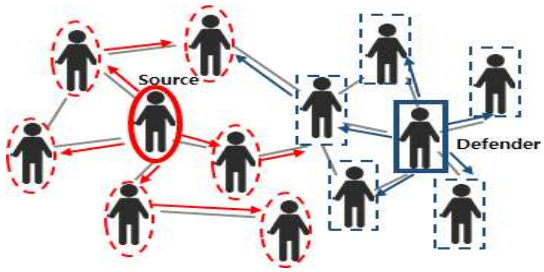

Fig. 2. Diffusion of the source's information (red) and defender's information (blue).

4) 보호된 상태는, 감염된 상태와 반대로 방어자로부터 정보를 전달 받은 경우를 말한다.

나가는 것을 가정하였다. 후자의 경우를 본 논문에서는 Susceptible-Defended (SD) 라고 정의하겠다. 이런 SI 혹은 SD 모델에서는 만약 한 노드가 이웃 노드로부터 감염 혹은 보호가 된 후에는 시간이 지나도 그 상태가 변하지 않는다고 가정한다.

3.2 Voronoi 추정 방법(Voronoi Inference)

앞서 설명한 두 가지 정보의 확산 모델이 주어진 경우 최우추정량을 수학적으로 표현하면 다음과 같다. 먼저 I_N을 네트워크에서 N개의 노드에 정보가 퍼진 그래프라고 하고 D_M을 M개의 노드가 방어자로부터 그 반대 정보를 받은 노드들의 그래프라고 하자. 이 경우에 최우추정량은 아래와 같이 표현된다.

$$\hat{s}_{MLE} = \arg max_{v \in I_N} P(I_N, D_M | v, p^*). \quad (1)$$

즉, 네트워크에 방어자 p^*가 있는 경우 두 정보의 확산 그래프 I_N과 D_M을 만들어 냈을 확률이 가장 큰 노드를 찾는 것이 최우추정량이다. 하지만, 위의 식은 일반적인 경우 매우 계산하기 어려운(NP-hard) 문제라는 것이 잘 알려졌다. 그 이유는 앞의 모델에서 네트워크에 N개의 감염 노드와 M개의 보호 노드들이 퍼지게 되는 것은 확률적인 현상이기 때문이다. 즉, 네트워크 관리자가 임의의 시점에서 퍼져 있는 모습을 보고 발생될 확률을 계산할 때, 고려되어야 하는 확률적 경로(Sample path)에 대한 경우의 수가 지수적으로 많다. 따라서 본 논문에서는 최우추정량에 대한 하나의 근사적 방법으로 다음과 같은 Voronoi 분할법에 기반 한 추정법을 제안한다.

1) **Voronoi 추정법**: 일반적으로 Voronoi 분할(Partition)이라고 하는 것은 아래 Fig. 3과 같이 평면에서 노드들이 주어진 경우에 이들을 가장 균등하게 각각의 노드들에 대한 구획을 구분 짓는 방법을 말한다.

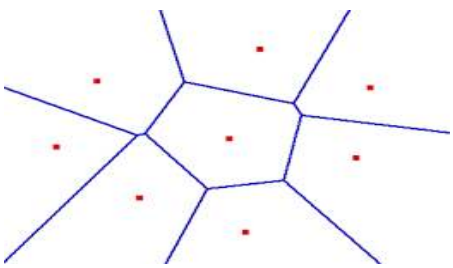

Fig. 3. Voronoi partition [The Wooden Boat Forum].

이와는 반대로, 본 연구에서는 각각의 정보가 퍼지게 된 경우에 대해서는, Fig. 4에서와 같이 각 정보에 대한 구획이 나누어졌을 때, 그 중심점이 어디 있는가를 찾는 방법이다. 이 경우, 네트워크에서 방어자의 위치가 네트워크 관리자에게 이미 알려진 상황에서 두 정보의 구획까지의 거리(Hop-distance)를 계산하여 원래 정보의 근원의 위치를 추정하는 기법을 제안한다. 이런 방법이 합리적인 이유는, 모델에서 두 정보는 같은 비율로 퍼져나간다고 가정했기 때문이다. 즉, 이 경우, 평균적으로 두 정보가 만들어내는 구획은 각각의 근원들로부터 비슷한 거리에 있을 확률이 가장 높을 것이다. 하지만, Fig. 4에서와 같이 고정된 방어자를 기준으로 경계(Boundary)까지 거리가 같은 감염된 노드는 여러 개가 있을 수 있다. 따라서 이 중에 감염된 그래프에서 루머 중심성[5](Rumor Centrality)가 가장 큰 노드를 선택하는 방법을 제안한다. 본 논문에서는 이 추정 방법을 Voronoi 추정법이라고 명명하고 그 추정치를 $\hat{s}_v$ 라는 기호로 표현하면 아래와 같은 식을 얻게 된다.

$$\hat{s}_v = \arg max_{v \in K(p^*)} R(v). \quad (2)$$

여기서, $K(p^*)$는 감염된 노드 중 경계선까지의 거리가 방어자 p^*와 같은 노드들의 집합을 의미하고 $R(v)$는 감염된 노드가 가진 루머 중심성을 뜻한다. 즉, Fig. 4에서 원래 정보가 퍼진 노드의 집합이 빨강색 원으로 표시가 되어 있고 방어자의 정보가 퍼진 노드들은 초록색 사각형으로 표시되어 있다. 이 경우, 방어자는 두 정보의 경계지점까지 2-hop의 거리에 있으므로 Voronoi 추정법은 먼저 감염된 노드 중

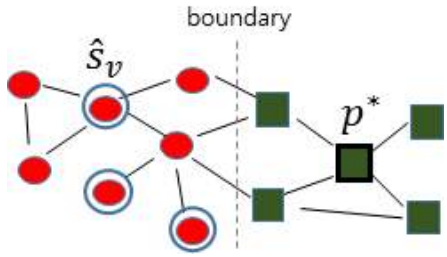

Fig. 4. Voronoi inference: Proposed method.

5) 감염 그래프에서 한 노드의 루머 중심성은 그 노드로부터 그래프가 발생될 수 있는 모든 샘플 경로들의 수를 더한 값이다. 자세한 계산 방법은 [1]을 참고하면 된다.

경계선까지 2-hop 거리에 있는 모든 감염된 노드를 선택한다. 위의 예시에서는 파랑 동그라미가 그려진 3개의 감염된 노드가 선택되는데, 이 중에서 루머 중심성이 가장 큰 노드는 맨 위에 있는 노드이고(3노드 중 가장 가운데 있다) 따라서 이 노드를 원래 정보의 근원으로 추정한다.

4. 방어자 선택 알고리즘

지금까지 네트워크에서 방어자가 주어진 경우 두 정보가 퍼진 가운데 원래 정보의 근원을 찾는 방법에 관한 설명을 하였다. 하지만, 이런 접근이 있기 전에 네트워크 관리자가 취할 수 있는 또 따른 방법 중 하나는 좀 더 영리하게 이런 방어자를 미리 심어 두는 것이다. 본 장에서는 네트워크에서 이런 방어자를 어떻게 선택하는지에 대한 두 가지 알고리즘을 제안한다.

4.1 거리 중심 알고리즘 (Distance Center based Defender Selection Algorithm: DICD)

이 방법은 주어진 네트워크에서 모든 노드의 거리 중심(Distance Center)을 찾아서 이 노드를 관리자는 방어자로 선정해 놓는 방법이다. 여기서 먼저, 거리 중심이라는 말은 다음과 같이 정의된다. 주어진 그래프 G에서 임의의 두 노드를 선택하고 이를 각각 v와 w라고 하자. 그러면 일반적으로는 루프(Loop)가 있는 그래프인 경우에는 이 두 노드를 연결하는 경로는 여러 개가 있을 수 있다. 그중에서 $d(v,w)$를 이런 경로 중 가장 짧은 경로(Shortest path)라고 하면 네트워크에서 노드 $v \in V$에 대해서 자신이 아닌 모든 다른 노드 w와의 거리를 합한 후 그것에 대한 역수를 $C(v)$라고 정의한다. 즉, $C(v) = 1/\sum_{w \in V} d(v,w)$ 일 때, 이런 $C(v)$를 최대화하는 노드가 네트워크에서 거리 중심인 노드이다. 이를 수학적으로 기술하면, 네트워크에서 방어자 p^*는 다음과 같이 표현 된다.

$$p^* = \arg max_{v \in V} C(v). \quad (3)$$

이런 거리 중심 알고리즘을 고려하는 이유는 일반적으로 루머나 바이러스와 같은 정보가 어디서부터 퍼지는 것에 대한 사전정보가 없는 경우 이것이 네트워크에서 랜덤하게 발생된다고 볼 수 있기 때문이다. 이 경우, 정보의 근원과 최대한 거리가 가까운 방어자를 선정하여서 네트워크에서 확률적으로 퍼지는 정보가 반대 정보와의 경계가 예측 불가능하지 않도록 하고자 하는 데 목적이 있다. 따라서 거리 중심 노드를 적절히 찾아서 방어자로 선택한다.

4.2 차수 중심 알고리즘 (Degree Center based Defender Selection Algorithm: DECD)

앞서 제안한 네트워크에서 거리 중심에 있는 노드를 선택하는 방법과는 다르게 차수 중심 알고리즘은 네트워크에서 각 노드를 기준으로 이웃 노드의 수가 가장 많은 노드를 방어자로 선택하는 알고리즘을 말한다. 이 방법 또한, 첫 번째 방법과 마찬가지로 루머처럼 정보가 발생되는 위치가 랜덤인 상황에서 차수가 가장 큰 노드를 선택하면 경계로 만들어지는 부분이 좀 더 예측 가능한 부분이 되도록 할 수 있다는 장점이 있다. 왜냐하면 이런 노드는 전체 네트워크에서 허브와 같은 역할을 하게 되는데 대부분의 노드가 이런 허브와는 아주 멀지 않게 연결이 되어 있기 때문이다. 본 방법에 대해 수학적으로 기술하면 아래와 같다. 먼저, 네트워크에 있는 각 노드 $v \in V$에 대해서 이것과 연결된 이웃 노드의 수를 $D(v)$라고 할 때, 방어자 p^*는 다음과 같이 표현될 수 있다.

$$p^* = \arg max_{v \in V} D(v). \quad (4)$$

차수 중심 방법은 본 논문에서 고려가 되는 네트워크가 페이스북(Facebook)이나 트위터(Twitter)와 같은 소셜 네트워크인 경우 이들의 분포가 멱급수(Power law)를 따른다고 알려져 있는데 이 경우에 잘 적용이 될 수 있는 방어자 선정 방법 중 하나다.

5. 분석 및 탐지 실험

본 장에서는 앞서 제안한 두 가지 알고리즘에 기반하여 네트워크에서 방어자를 선택하고 이를 토대로 시뮬레이션을 통해 두 가지 반대의 정보를 충분히 퍼뜨린 후 Voronoi 추정법을 사용하여 실제 그래프에서 정보의 근원을 얼마나 잘 찾을 수 있는가에 대한 실험을 진행하였다. 실험을 위해서 Matlab을 사용하였고 고려된 네트워크와 기본적인 실험 세팅에 대한 설명은 다음과 같다.

5.1 그래프(네트워크) 세팅

본 실험에서는 다음과 같은 세 가지 종류의 네트

워크에 대한 그래프를 고려한다. (1) 먼저, 가장 간단한 그래프인 트리구조에 대해서 실험을 진행하였고 특히, 트리구조 중에서도 네트워크의 모든 노드가 동일한 차수를 가진 정규 트리를 고려하였다. (2) 둘째로, 트리보다는 좀 일반적일 수 있는 합성 그래프(Synthetic graph)에 대해서 살펴보았는데, 가장 잘 알려진 합성 그래프 중에서 Erdos-Renyi(ER) 랜덤 그래프와 Small-World(SW) 그래프 그리고, Scale Free(SF) 그래프에 대해서 살펴보았다. ER 랜덤 그래프 같은 경우 네트워크에 노드가 주어지고 각 노드가 연결될 확률이 p>0인 베르누이 분포를 따라 동전 던지기를 하여 성공을 하면 연결하고 실패를 하면 연결을 하지 않는 방법으로 구성된 랜덤 그래프이다. 본 실험에서는 전체 그래프의 노드가 3000개인 경우 한 노드가 가진 평균 차수가 4가 되도록 네트워크를 형성시켰다. SW 그래프는 세상의 모든 사람이 많아도 6-hop으로는 다 연결이 되도록 구성이 되어 있는 특징을 가진 그래프이고 SF 그래프는 앞서 언급한 바 있는 멱급수 형태를 가진 이상적인 그래프이다. 각각에 대한 네트워크 형성에 대한 파라미터는 ER과 동일하게 적용했다. (3) 마지막으로, 실험에서 고려한 그래프는 실제 데이터를 가지고 만들어진 그래프(Real-world graph)로서 페이스북 그래프[13]와 US-power grid network [14]를 생성시켜서 실험하였다. 페이스북 네트워크인 경우는 총 4039노드가 88234개의 변으로 구성되어 있는 그래프를 형성하였다. 그리고 US-Power grid 네트워크의 경우 4941노드에서 6594개의 변을 가진 그래프로 형성하였다. 위와 같이 형성된 그래프 위에서 거리 중심 및 차수 중심으로 방어자를 선택한 후 랜덤하게 정보를 퍼뜨리는 근원을 선택하여 SI 및 SD 모델에 따라 네트워크에서 전체 확산된 노드의 수가 M+N=1000 이 되도록 한 후 제안한 추정법이 실제 근원을 찾는지 아닌지를 확인하는 실험을 진행하였다.

5.2 탐지확률(Detection Probability)

탐지확률의 척도는 아래와 같은 식을 사용하였는데, 이는 전체 시도된 정보의 확산 중에서 실제로 제안한 추정법이 원래 정보의 근원을 찾는 수로 나눠진 것이다.

$$\text{Detection Probability} = \frac{\#\text{of Detections}}{\#\text{of Trials}} \quad (5)$$

본 실험에서는 총 500번의 확산 현상을 생성시켰고 그중에서 Voronoi 추정법으로 정확히 정보의 근원을 찾는 횟수를 나누어 각 경우에 대한 탐지확률을 계산하였다. 실험에서 구한 정확성 지표인 탐지확률은 첫째로 원래 정보와 그 반대 정보의 근원들의 거리를 증가시키면서 확산을 시킨 경우에 대한 경향성을 위에서 설명한 3가지의 네트워크 그래프에 대해서 살펴보았다. 그리고 전체 확산이 된 노드의 수(감염 및 보호)가 증가할 때 탐지확률을 구하였다. 특히, 후자의 경우에는 기존연구[3]에서 제안한 소극적(Passive) 방어자인 경우와 비교하여 본 연구에서 고려하고 있는 적극적(Active) 방어자 즉, 정보가 이웃으로부터 들리고 반대의 정보를 전파하는 것이 아닌, 처음부터 동시에 다른 정보를 전파하는 경우에 대한 성능의 차이를 구하였다.

5.3 실험 결과

1) 정규 트리: 먼저 Fig. 5는 네트워크가 완벽하게 대칭인 정규 트리인 경우에 두 근원의 거리에 따른 탐지확률이 어떻게 되는지를 보여주고 있다. 그림의 결과로부터 알 수 있는 사실은, 두 근원의 거리가 가까울수록 탐지확률이 커진다는 사실이다. 그 이유는 가까운 거리에서 다른 정보가 퍼져나가는 경우 경계선이 두 근원의 가운데서 발생할 가능성이, 지수분포를 따르는 확산현상에 대한 모델에서는 더 크게 되고 본 연구에서 제안한 Vorinoi 추정법이 이 경우에는 정확하게 맞을 확률이 크기 때문이다. 또한 위에서 볼 수 있듯이, 정규트리에서 노드의 차수가 더 큰 경우(d=10)가 그렇지 않은 경우(d=3)에 비해서 탐지확

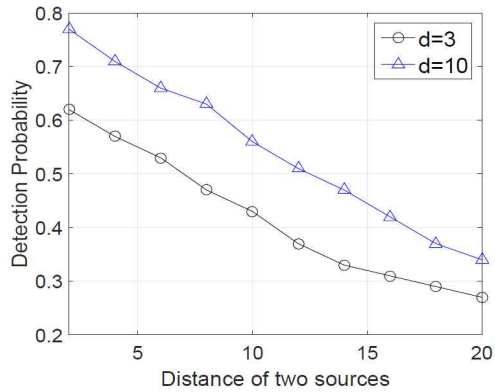

Fig. 5. Detection probability in regular tree w.r.t. distance of two sources (source and defender).

률이 더 높다. 그 이유는 제안된 추정법에서 적용되는 루머 중심성이 일반적으로 차수가 클 때 잘 동작한다고 알려져 있기 때문이다. 다음으로 Fig. 6에서는 차수가 3인 경우에 대해서 선행 연구에서 사용한 소극적 방어자보다 적극적 방어자를 사용하는 경우에 얼마나 성능이 좋아지는지를 보여주고 있다. 본 논문에서 제안된 Voronoi 추정법이 현재까지 잘 알려진 최적의 추정법인 최우추정법이 아님에도 불구하고 소극적 방어자에서 사용한 선행연구[3]의 최우추정량에 비해서도 그 성능이 월등히 좋아진다는 것을 확인할 수 있다. 이는 두 근원의 거리가 아주 멀리 않으면 제안한 Voronoi 추정법을 사용하여 원래 정보의 근원을 찾는 것이 소극적 방어자처럼 전파될 때까지 기다렸다가 찾는 것보다 훨씬 효율적이라는 사실을 나타낸다.

2) 합성 그래프: 합성 그래프에서는 앞서 본 정규트리와는 다르게 그래프에 루프가 있을 수 있는데, 이 경우에 [1]에서 알려진 바와 같이 루머 중심성을 바로 계산하는 것이 매우 복잡하다. 따라서 이런 경우에는 Breath-First-Search (BFS) 트리 형태로 바꿔주는 근사적 방법을 사용하여 루머 중심성을 계산한다.[6] 방어자 선택에 대한 알고리즘은 ER에서는 거리 중심 알고리즘을 사용하였고, 네트워크의 거리가 상대적으로 작은 SW와 SF에서는 차수 중심 알고리즘으로 선택하였다. Fig. 7은 정규트리에서처럼 소스와 방어자의 거리에 따른 탐지 확률이 어떻게 변하는지를 보여주고 있는데, 합성 그래프에서 ER, SW 및

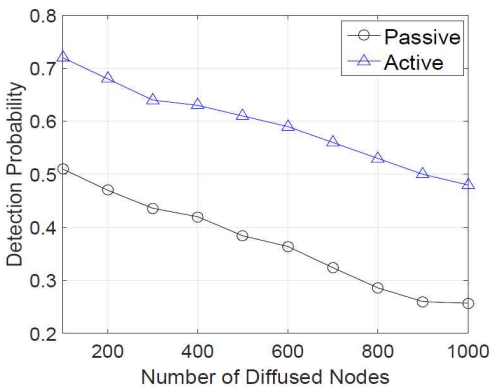

Fig. 6. Detection probability in regular tree (d=3): Active Defender vs. Passive Defender.

6) 이 부분에 대한 자세한 계산식은 [1]에 나타나 있다.

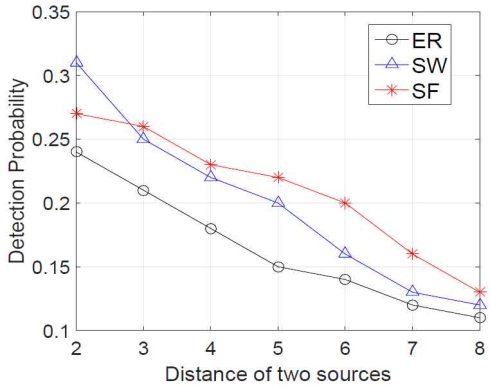

Fig. 7. Detection probability in synthetic graphs w.r.t. distance of two sources.

SF의 3가지 경우에 대해서 결과를 얻었다. 합성그래프는 루프가 있는 부분에 대한 한 번의 근사를 더 사용하기 때문에 정규트리에서의 결과보다는 성능이 좋지는 않지만, 여전히 두 근원의 거리가 가까운 경우가 그렇지 않은 경우보다 훨씬 원래 정보의 근원을 잘 찾는다는 사실을 확인할 수 있었다. 다음으로 Fig. 8에서는 ER과 SF의 두 가지 그래프에 대해서 적극적 방어자와 소극적 방어자인 경우 탐지확률이 얼마나 차이가 나는지를 보여주고 있다. 이 경우에도 비록 정규트리에서의 증가량 보다는 작지만, 소극적 방어자에 비해 적극적 방어자를 사용하는 경우가 대략 10% 이상 탐지확률이 증가한다는 것을 확인할 수 있었다.

3) 실제 네트워크: 이 경우는 실제 인터넷에 있는 데이터를 바탕으로 페이스북 그래프와 미국의 전기선 연결 그래프에 대한 실제 데이터를 바탕으로 네트

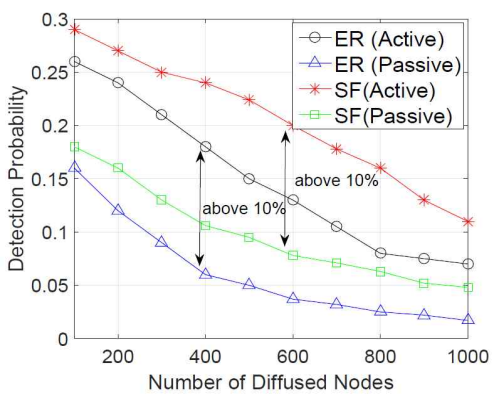

Fig. 8. Detection probability in synthetic graphs: Active Defender vs. Passive Defender.

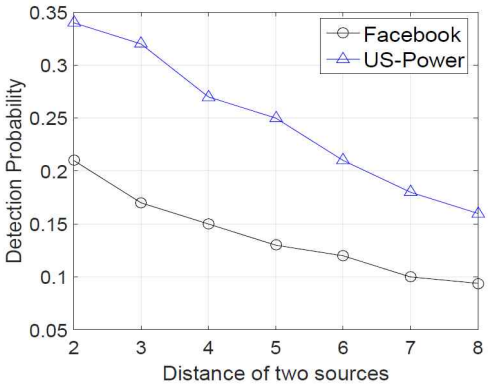

Fig. 9. Detection probability in real-world networks w.r.t. distance of two sources.

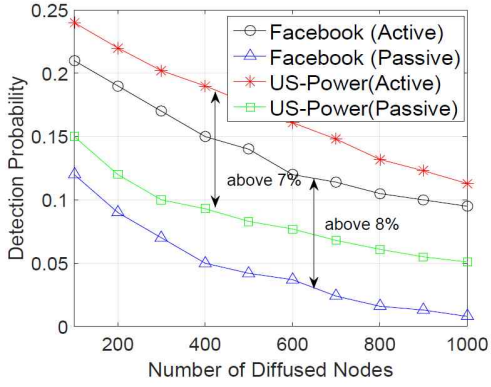

Fig. 10. Detection probability in real-world networks: Active Defender vs. Passive Defender.

워크를 형성한 후 제안한 추정 방법으로 결과를 얻었다. 먼저 페이스북 네트워크인 경우 전체 거리가 크지 않으므로 차수 기반 알고리즘 방법을 사용하여 방어자를 선택하였고, 반대로 US-power grid 네트워크에서는 거리 중심 알고리즘으로 방어자를 선택하였다. Fig. 9에서 이 두 가지 실제 네트워크에서의 적극적 방어자인 경우 두 근원에 대한 거리에 따른 탐지확률을 얻었고 합성 그래프와 마찬가지로 거리가 멀어질수록 탐지 확률이 낮아지는 현상을 확인할 수 있었다. 특히, 본 결과에서 페이스북 그래프에 비해서 US-Power 그래프가 더 탐지가 잘 되는 이유는 페이스북 그래프인 경우 작은 거리 안에도 아주 많은 노드가 존재하는 반면 US-Power는 상대적으로 더 적은 노드가 있기 때문에 나타나는 결과이다. Fig. 10에서는 앞에서 접근한 바와 같이 소극적인 방어자에 비해 적극적 방어자를 사용하는 경우 대략 7% 이상으로 성능 향상이 있음을 확인할 수 있었다. 이는 앞의 다른 그래프들에 비해서는 낮은 수치일 수 있으나, 실제 네트워크에서 노드 수가 아주 많은 경우에는 무시할 수 없는 증가량임을 확인할 수 있다.

마지막으로, Table 2에서는 제안된 두 종류의 방어자 선정 알고리즘을 적용한 탐지확률을 다양한 그래프에 대해서 얻었다. 위의 결과에서 확인할 수 있는 사항은, ER이나 US-Power grid 네트워크와 같이 일반적으로 그래프 반경이 큰 경우에는 거리 중심으로 방어자를 선택하는 것이 좋고 SF 나 SW 그리고 페이스북과 같이 전체 그래프 반경이 크지 않고 인기 있는 특정 노드를 중심으로 많은 다른 노드가 연결된 경우에는 차수 중심으로 방어자를 선정하는 것이 효과적이라는 것을 알 수 있다.

6. 결 론

본 논문에서는 네트워크에서 루머와 같은 정보가 퍼진 경우 그 근원이 어디인지를 찾아내는 방법에 대한 연구로 반대정보를 같이 퍼뜨리는 방어자가 있는 경우에 대해 살펴보았다. 특히, 방어자가 이웃으로부터 정보를 듣는 순간 다른 정보를 퍼뜨리는 소극적인 방어자가 아니라 동시에 다른 정보를 같이 퍼뜨리는 적극적인 방어자인 경우 얼마나 더 원래 정보의

Table 2. Detection probability for two defender selection algorithms (M+N=500)

	거리	ER	SW	SF	Facebook	US-Power
DICD	3	0.27	0.23	0.22	0.23	0.28
	6	0.22	0.17	0.19	0.15	0.24
	9	0.17	0.13	0.16	0.10	0.22
DECD	3	0.21	0.26	0.28	0.24	0.23
	6	0.17	0.21	0.23	0.21	0.16
	9	0.14	0.17	0.21	0.19	0.13

근원을 잘 찾을 수 있는지에 대해 다양한 네트워크 토폴로지에 대해서 살펴보았다. 이를 위해 Voronoi 추정법이라고 하는 새로운 방법을 제안하였고 그 효율성도 실험을 통해 입증하였다. 이는 [15]에서와 같이 최근 소셜네트워크에서 클라우드나 포그 컴퓨팅 기반 보안 프레임워크를 개발하는데도 유용하게 사용될 수 있을 것으로 기대한다.

REFERENCE

[1] D. Shah and T. Zaman, "Detecting Sources of Computer Viruses in Networks: Theory and Experiment," *Proceedings of ACM International Conference on Measurement and Modeling of Computer Systems*, pp. 203-214, 2010.

[2] D. Shah and T. Zaman, "Rumor Centrality: A Universal Source Detector," *Proceedings of ACM International Conference on Measurement and Modeling of Computer Systems*, pp. 199-210, 2012.

[3] J. Choi, S. Moon, J. Shin, and Y. Yi, "Estimating the Rumor Source with Anti-Rumor in Social Networks," *Proceedings of IEEE International Conference on Network Protocols Workshop on Machine Learning*, pp. 1-6, 2016.

[4] W. Dong, W. Zhang, and C. Tan, "Rooting Out the Rumor Culprit from Suspects," *Proceedings of IEEE International Symposium on Information Theory*, pp. 2671-2675, 2013.

[5] J. Choi, S. Moon, J. Woo, K. Son, J. Shin, and Y. Yi, "Rumor Source Detection under Querying with Untruthful Answers," *Proceedings of IEEE International Conference on Computer Communications*, pp. 2259-2267, 2017.

[6] J. Jaing, S. Wen, S. Yu, Y. Xiang, and W. Zhou, "K-Center: An Approach on the Multi-Source Identification of Information Diffusion," *IEEE Transactions on Information Forensics and Security*, Vol. 10, No. 12, pp. 2616-2626, 2015.

[7] F. Ji and W. Tay, "An Algorithmic Framework for Estimating Rumor Sources With Different Start Times," *IEEE Transactions on Signal Processing*. Vol. 65, No. 10, pp. 2517-2530, 2017.

[8] Z. Wang, C. Wang, J. Pei, and X. Ye, "Multiple Source Detection without Knowing the Underlying Propagation Model," *Proceedings of Association for the Advancement of Artificial Intelligence*, pp. 217-223, 2017.

[9] Z. Chen, K. Zhu, and L. Ying "Detecting Multiple Information Sources in Networks under the SIR Model," *IEEE Transactions on Network Science and Engineering,* Vol. 3, Issue 1, pp. 17-31, 2016.

[10] G. Fanti, P. Kairouz, S. Oh, and P. Viswanath, "Spy vs. Spy: Rumor Source Obfuscation," *Proceedings of ACM International Conference on Measurement and Modeling of Computer Systems*, pp. 271-284, 2015.

[11] G. Fanti, P. Kairouz, S. Oh, K. Ramchandran, and P. Viswanath, "Rumor Ource Obfuscation on Irregular Trees," *Proceedings of ACM International Conference on Measurement and Modeling of Computer Systems,* pp. 153-164, 2016.

[12] W. Luo, W.P. Tay, and M. Leng, "Infection Spreading and Source Identification: A Hide and Seek Game," *IEEE Transaction on Signal Processing*, Vol. 64, No. 16, pp. 4228-4243, 2016.

[13] J. Leskovec and J. McAuley, "Learning to Discover Social Circles in Ego Networks," *Proceedings of Annual Conference on Neural Information Processing Systems*, pp. 539-547, 2012.

[14] D.J. Watts and S.H. Strogatz, "Collective Dynamics of Small-World Networks," *Nature*, Vol. 393, No. 6684, pp. 440-442, 1988.

[15] M. Shin and S. Kim, "A Study on the Security Framework for IoT Services Based on Cloud and Fog Computing," *Journal of Korea Multimedia Society*, Vol. 20, No. 12, pp 1928-2939, 2017.

최 재 영

2008년 고려대학교 수학과 (학사)
2013년 고려대학교 수학과(석사)
2018년 한국과학기술원 전기 및 전자공학부(Ph.D.)
2018년～현재 호남대학교 미래자동차공학부 조교수

관심분야 : 소셜네트워크, 데이터 마이닝, 통계적 추론, 자동차통신, 네트워크 보안