

# 외부 커뮤니티 연관도를 이용한 소셜 네트워크 분석

이현진\*, 지태창\*\*

## 요약

소셜 네트워크로부터 커뮤니티를 발견하기 위해서는 네트워크의 노드들에 대한 군집화 과정이 필요하다. 일반적인 군집화 알고리즘은 커뮤니티의 개수를 미리 설정해 주어야 한다는 약점을 가지고 있다. 커뮤니티의 개수에 의해서 군집화의 결과가 달라질 수 있기 때문에 커뮤니티의 개수 선정은 매우 중요하다. 본 논문에서는 커뮤니티 간의 구별을 위한 외부 커뮤니티 연관도를 정의하고, 이를 기존의 군집화 알고리즘과 결합하여 커뮤니티의 개수를 동적으로 발견하는 방법을 제안한다. 제안하는 방법은 커뮤니티 개수의 정확도와 커뮤니티 평균 순도를 기준으로 기존의 방법들과 비교하였다. 실험 결과 기존의 방법들에 비하여 제안하는 방법의 우수한 성능 확인할 수 있었다.

## Social Networks Analysis using External Community Relationship

Hyunjin Lee\*, Taechang Jee\*\*

## Abstract

A clustering process for nodes in a network is required to find communities from social networks. General clustering algorithm needs to be configured the number of communities in advance. The number of communities is a very important element because the result of clustering can be different, depending on it. In this paper, we define the external community relationship which is distinguished between communities. Using the external community relationship as an evaluation metric of clustering result, we propose a method to determine the number of communities dynamically. We compare the proposed method to existing methods based on the accuracy of the number of communities and the average purity of communities. Our results show favorable performance for these criteria compared to the existing methods that were evaluated.

Keywords : 소셜 네트워크, 소셜 네트워크 분석, 커뮤니티 발견, 군집화

## 1. 서론

소셜 네트워크 서비스(Social Networks Services: SNS)는 웹의 발전으로 인해 오프라인 소셜 네트워크가 온라인상에서 인적 관계를 형성해 줄 수 있도록 발전한 서비스이다. 소셜 네트워크 서비스는 동창, 동기, 지역 등 사람들을 일정 부류로 분류 해주는 서비스, 친구들과 연락

을 주고받을 수 있는 기능을 제공하는 서비스, 사용자들의 신뢰관계를 기반으로 추천기능을 제공하는 기능들을 제공하고 있다. 대표적인 소셜 네트워크 서비스로는 국외의 Facebook, Flickr, Twitter, Myspace 등과 국내의 Cyworld, me2day 등이다.

컴퓨터과학 분야에서는 소셜 네트워크에서 존재하는 웹 사용자들 간의 연결성 확장을 통한 검색 효율을 향상시키려는 연구와 사회현상과 소셜 네트워크상의 현상에 대한 분석, 네트워크의 효율적인 구성에 관한 연구들이 활발히 진행되고 있다[1]. 네트워크 형태를 잘 이해하는데 있어서 중요한 단계는 그 안에 있는 커뮤니티(community)를 찾는 것이다. 이러한 커뮤니티를 찾기 위하여 그래프 분할(graph partitioning)이

※ 제일저자(First Author) : 이현진  
접수일:2011년 01월 12일, 수정일:2011년 03월 07일,  
완료일:2011년 03월 30일  
\* 한국사이버대학교 컴퓨터정보통신학과  
hjlee@mail.kcu.ac  
\*\* 연세대학교 컴퓨터과학과

나 군집화(clustering)등 커뮤니티 발견을 위한 다양한 방법들이 제시되었다.

안용열등은 커뮤니티를 링크의 집합으로 재정의 하여 모든 노드들이 여러 커뮤니티에 속하는 상황을 해결하는 방법을 제시하였다[2].

Freeman은 메타데이터(metadata)와 콘텐츠(content)를 기반으로 유사그룹들을 동적으로 식별하는 위상 트리 방법 (topological tree method)을 사용하여, 자동적으로 소셜 네트워크 그룹을 조직하였다[3].

Boulet 등은 배치 커널 자기조직화지도(self-organizing map)와 관련된 라프라시안 방법을 사용하여 소셜 네트워크 분석을 하는 방법을 제안하였다[4].

강윤섭등은 커뮤니티 발견 문제를 공통 이웃 그래프 밀도(Common neighborhood sub-graph density)를 정의하여 이러한 유사도를 유사도 전파(affinity propagation)알고리즘과 결합하는 방법을 제안하였다[5].

일반적으로 커뮤니티를 구성하는데 있어서 자신이 속한 커뮤니티에 대한 유사도를 이용하여 동일 커뮤니티 여부를 결정하게 된다. 하지만, 외부 커뮤니티에 대한 유사도도 존재하고, 동일 커뮤니티에 속하는 노드들도 외부 커뮤니티에 대한 유사도는 서로 다른 경향을 보일 수 있기 때문에 커뮤니티의 충실도를 결정하는데 있어서 외부 커뮤니티 연관도를 고려해야 한다. 따라서 본 논문에서는 외부 커뮤니티 연관도를 정의하고 이를 군집화 알고리즘의 척도로 이용하여 커뮤니티를 찾는 방법을 제안하였다.

본 논문의 구성은 다음과 같다. 2장에서는 외부 커뮤니티의 연관도를 다루고, 3장에서는 외부 커뮤니티 연관도를 이용하여 소셜 네트워크에서 커뮤니티를 찾는 방법을 살펴본다. 4장에서는 실험과 결과를 분석하고 5장에서 결론을 맺는다.

## 2. 외부 커뮤니티 연관도

커뮤니티 발견을 위해서는 군집화 알고리즘이 필요하다. 군집화는 유사한 데이터 개체들의 집합인 군집으로 데이터를 분할함으로써 데이터 속에 숨겨져 있는 의미 있는 정보를 자동으로 발견하는 것이다. 커뮤니티를 발견하는 것은 정

확하게 해당 커뮤니티를 다른 커뮤니티와 구분하는 것이 아니라, 비슷한 특징을 가지는 커뮤니티들을 함께 모아둠으로써 사용자가 심도 있는 분석을 할 때 도움을 주고자 하는데 주목적을 지닌다.

좋은 커뮤니티 선택은 커뮤니티 내의 노드들은 서로 조밀하게(density) 연결되며 외부 커뮤니티의 노드들과는 성기게(sparse) 연결되는 노드의 그룹을 선택하는 것으로 정의된다. 즉, 같은 커뮤니티에 속한 노드 간에는 조밀한 연결이 이루어져야 하고, 외부 커뮤니티 사이에도 성긴 연결이 존재하게 된다.

일반적으로 커뮤니티를 구성하는데 있어서 자신이 속한 커뮤니티에 대한 유사도를 이용하여 동일 커뮤니티 여부를 결정하게 된다. 하지만, 외부 커뮤니티에 대한 유사도도 존재하고, 동일 커뮤니티에 속하는 노드들도 외부 커뮤니티에 대한 유사도는 서로 다른 경향을 보일 수 있기 때문에 커뮤니티의 충실도를 결정하는데 있어서 외부 커뮤니티 연관도를 고려해야 한다.

외부 커뮤니티 연관도  $R$ 은 각 커뮤니티의 외부 커뮤니티에 대한 연관도의 평균으로 정의된다.

$$R = \frac{\sum_{k \in K} R_k}{K} \quad (1)$$

여기서  $K$ 는 커뮤니티의 개수이다. 단일 외부 커뮤니티 연관도  $R_k$ 는 한 커뮤니티에 속한 노드들의 외부 커뮤니티와의 유사도에 대한 표준편차로 정의되며 이는 식 (2)와 같다.

$$R_k = \sqrt{\frac{1}{N_k} \sum_{i \in N_k} d^2(r_i, \bar{r})} \quad (2)$$

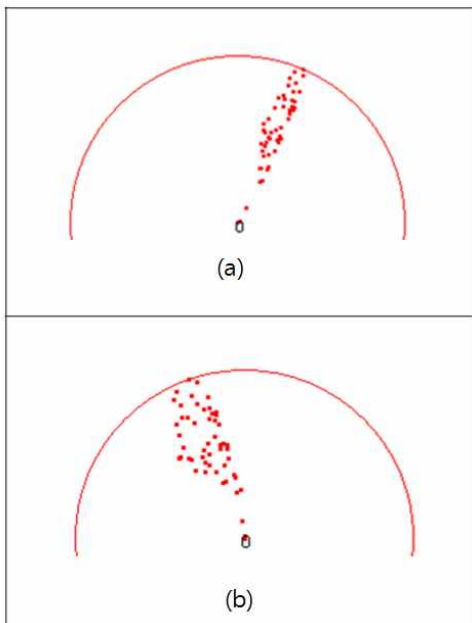
여기서,  $d(r_i, \bar{r})$ 는 두 벡터  $r_i$ 와  $\bar{r}$  사이의 거리이고,  $N_k$ 는  $k$  커뮤니티에 속한 노드들의 개수이며,  $\bar{r}$ 은  $r_i$ 들의 평균이다.  $r_i$ 는  $i$ 번째 노드와 외부 커뮤니티와의 유사도이다.

노드와 외부 커뮤니티와의 유사도는 일반적인 유사도가 아닌 노드가 영향 받는 방향으로 계산한다. 일반적인 유사도는 노드가 속한 커뮤니티를 결정하는데 사용한 척도이다. 이 척도와 커뮤니티 간의 방향을 이용하여 새로운 노드와 외부

커뮤니티와의 유사도를 계산한다. 새로운 유사도는 방향을 계산하기 쉬운 2차원 평면에서 구하게 된다. 커뮤니티 중심만을 2차원 평면으로 사상하고, 사상하는 방법은 다차원 척도법을 적용한다[6].

$$r_i = \sum_{k \in K, k \neq j} d_{ik} \cdot \overline{C_j C_k} \quad (3)$$

여기서,  $j$ 는  $i$  노드가 속한 커뮤니티의 번호이다.  $d_{ik}$ 는  $i$  노드와  $k$  커뮤니티간의 유사도이고,  $\overline{C_j C_k}$ 는  $j$  커뮤니티와  $k$  커뮤니티 사이의 방향을 의미한다. 다차원 척도법에 의해서 이 방향은 2차원 평면에서 구하게 된다.



(그림 1) 노드와 커뮤니티간의 유사도

그림 1은 노드와 커뮤니티간의 유사도를 도식화한 것으로 '0'은 커뮤니티의 중심을 의미하고, 점은 커뮤니티에 소속된 노드들이고, 원은 커뮤니티의 경계를 의미한다. (a)의 외부 커뮤니티 연관도의 값은 (b)의 외부 커뮤니티 연관도 값보다 더 작은 값을 보인다. (a)와 (b)를 비교 하였을 때 (a)가 (b)보다 노드들이 더 밀집되어 있다.

이는 (a)는 외부의 커뮤니티의 영향을 (b)에 비해 덜 받기 때문에 커뮤니티에 노드들이 유사한 연관도를 가지게 되어 밀집되어 있는 것이다. (b)의 경우는 노드들이 서로 다른 외부 커뮤니티 연관도를 보이기 때문에 밀집되어 있지 않고 퍼지는 현상이 발생된다.

### 3. 군집화 알고리즘과 외부 커뮤니티 연관도

#### 3.1 군집화 알고리즘

대상들을 군집화 하는 방법은 매우 다양하지만 모든 방법이 공통적으로 가지고 있는 기본전제는 군집 내의 객체들 간의 유사성을 극대화하고, 군집간의 유사성은 극소화하는 것이다. 군집화 알고리즘에는 K-Means 군집화 알고리즘과 같은 분할 기법(Partitional) 군집화 알고리즘과 계층적(Hierarchical) 군집화 알고리즘 등이 존재한다[7].

계층적 군집화 알고리즘은 개별 노드들 간의 거리를 비교하는 방법으로, 항상 계산 결과가 도출되고, 같은 차원의 다른 문제에 대해 항상 계산시간이 일정한 장점이 있지만, 노드의 개수나 노드를 구분하는 특징의 개수가 커지면 전체 계산시간이 기하급수적으로 증가하는 단점이 있다[8]. K-Means는 일정 시간 내에 계산이 종료할 수 있고, 노드의 개수나 노드를 구분하는 특징의 개수가 커져도 계산 시간이 점진적으로 증가하는 경향을 보인다.

본 논문에서는 커뮤니티를 구성할 때 K-means 군집화 알고리즘을 사용하였으며, 유사성 척도는 코사인 상관계수를 사용하였다. 초기 커뮤니티 형성을 위한 K-means 알고리즘은 표 1과 같다.

<표 1> K-means 알고리즘

1단계:	$K$ 개의 초기 중심값을 $C_i (i \in \{1 \dots K\})$ 를 초기화한다.
2단계:	$M$ 개의 입력데이터 $X_j (j \in \{1 \dots M\})$ 와 중심 $C_i$ 사이의 거리를 계산하고, 가장 짧은 거리의 중심에 할당한다.

$$u_{ij} = \begin{cases} 1 & \text{if } \frac{x_j c_i}{\sqrt{x_j^2 c_i^2}} \leq \frac{x_j c_k}{\sqrt{x_j^2 c_k^2}} \\ 0 & \text{otherwise} \end{cases} \quad (4)$$

where  $k \in \{1 \dots K, k \neq i\}, j \in \{1 \dots M\}$

3단계: 각 커뮤니티의 중심을 재계산한다.

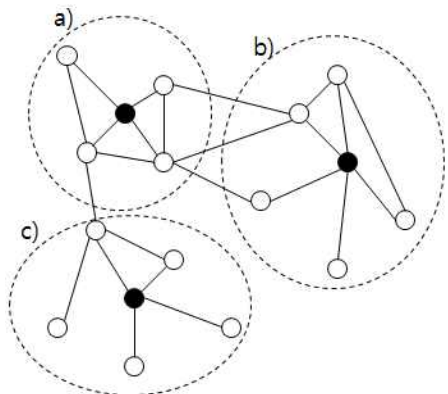
$$c_i = \frac{1}{N_i} \sum_k x_k \quad (5)$$

여기서  $N_i$ 는 커뮤니티  $i$ 에 할당된 입력 데이터의 수

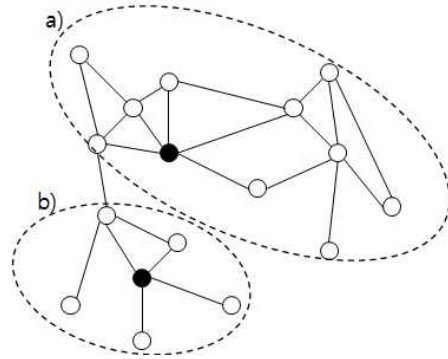
4단계: 입력데이터의 소속 커뮤니티에 변화가 없을 때 까지 2-3단계를 반복한다.

K-Means 알고리즘의 단점은 커뮤니티의 개수인  $K$ 를 미리 선정해야 한다는 것이다. 본 논문에서는 외부 커뮤니티 연관도를 사용하여 K-Means 군집화 알고리즘을 적용하여 커뮤니티의 개수를 자동으로 결정한다.

커뮤니티 개수 추정은 성능에 있어서 매우 중요하다 그림 2와 같이 커뮤니티 개수를 효과적으로 추정하지 못한 경우 a) 커뮤니티의 노드들이 b) 커뮤니티와 가까운 노드들과 c) 커뮤니티와 가까운 노드들이 혼재된다. 즉, a) 커뮤니티는 외부 커뮤니티와의 관계에서 서로 다른 속성을 가진 노드들로 구성되는 것이다. 하지만 그림 3과 같이 커뮤니티 개수를 효과적으로 설정한 경우 a) 커뮤니티의 노드들이 b) 커뮤니티와 가까운 노드들이 거의 존재하지 않게 된다. 즉, a) 커뮤니티는 외부 커뮤니티와의 관계에서 모두 유사한 속성을 가진 노드들로 구성되는 것이다.



(그림 2) 커뮤니티 개수를 효과적으로 설정하지 못한 경우



(그림 3) 커뮤니티 개수를 효과적으로 설정한 경우

### 3.2 외부 커뮤니티 연관도와 결합

외부 커뮤니티 연관도가 작으면, 해당 커뮤니티 내의 노드들이 외부 커뮤니티에 대해서 비슷한 성향을 보이고 있어서, 해당 커뮤니티를 더 분리할 필요가 없게 된다.

외부 커뮤니티 연관도가 크면, 해당 커뮤니티 내의 노드들이 외부 커뮤니티에 대해서 서로 다른 성향을 보이는 부분 집합들이 존재할 수 있게 된다. 그래서 해당 커뮤니티를 분리하는 것이 더 좋은 군집 결과를 보일 수 있다.

따라서 외부 커뮤니티 연관도가 충분히 작아질 때까지 커뮤니티의 개수를 증가시키면서 군집화를 실행시키게 된다.

외부 커뮤니티 연관도를 K-Means 군집화 알고리즘과 결합하는 방법은 표 2와 같다.

<표 2> K-means와 외부 커뮤니티 연관도의 결합

1단계:	$K = 3$ 으로 K-Means 군집화 알고리즘을 수행한다.
2단계:	수식 (1)의 외부 커뮤니티 연관도를 계산하여 $k_1 = 3, R_1 = R^3$ 를 저장한다.
3단계:	반복 횟수의 임계치인 $L$ 을 설정하고, 현재 반복 횟수를 $l = 1$ 로 설정한다.
4단계:	$K$ 를 1 증가시켜서 K-Means 군집화 알고리즘을 수행한다.
5단계:	수식 (1)의 외부 커뮤니티 연관도 $R^l$ 를 계산하고 커뮤니티의 개수 $k^l$ 를 저장한다.

다.  $t$ 는 현재 순번을 의미한다.

6단계: 직전 외부 커뮤니티 연관도와 차이  $r$ 를 계산한다.

$$r = R^{t-1} - R^t \quad (6)$$

7단계:  $r$ 이 0보다 크거나  $t-1=3$ 이고  $l < L$ 이면,  $l$ 을 1 증가시키고, 4단계에서 6단계를 반복한다.

8단계:  $r$ 이 0보다 작고  $t-1 > 3$ 고  $l < L$ 이면,  $k_2 = t-1, R^{k_2} = R^{t-1}$ 를 저장하고, 반복문을 빠져나온다.

9단계:  $l > L$ 면, 계산을 멈추고  $k_2 = t, R_2 = R^t$ 를 저장한다.

10 단계: 2단계에서 계산된 연관도  $R_1$ 과 반복문의 결과인  $R_2$ 을 비교하여 최소값  $R_{\min}$ 을 저장한다.

$$R_{\min} = \min(R_1, R_2) \quad (7)$$

$k_{\min}$  값을 수식 (7)에서 선택된  $R_{\min}$ 을 만족하는 커뮤니티 개수로 선택한다.

## 4. 실험환경 및 결과

### 4.1 커뮤니티 평균 순도

커뮤니티 발견 알고리즘의 정확도 기준으로 발견한 커뮤니티의 평균 순도 (average purity)를 사용했다[9]. 그 정의는 다음과 같다.

$$\frac{1}{K} \sum_k \sum_{i \neq j \text{ and } i, j \in C_k} \frac{\delta(i, j)}{|C_k|^2} \quad (8)$$

$K$ 는 발견된 커뮤니티의 개수이며  $\delta(i, j)$ 는 노드  $i$ 와  $j$ 가 실제로도 같은 커뮤니티에 속하면 1, 아니면 0의 값을 갖는다.  $|C_k|$ 는 발견된 커뮤니티 중  $k$ 번째 것의 크기를 의미한다.

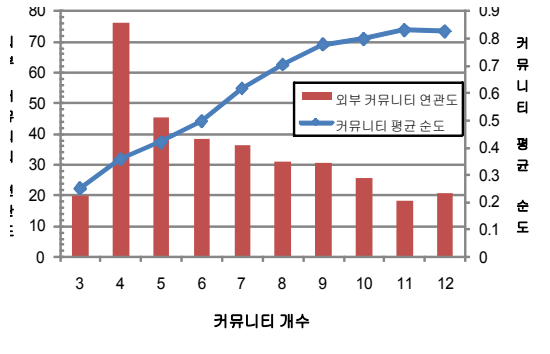
### 4.2 커뮤니티 개수에 따른 외부 커뮤니티 연관도

외부 커뮤니티 연관도가 커뮤니티를 발견하는데 영향을 끼치는 것을 확인하기 위하여 커뮤니티 개수를 변화시키면서 외부 커뮤니티 연관도를 계산하였다.

실험에는 College football[10], Political blogs[11] 데이터 세트를 사용했다. 이 데이터 세트들은 각 데이터 점이 속한 커뮤니티 정보까지 포함하고 있다. College football은 12개, Political blogs는 3개의 커뮤니티로 구성되어 있

다. College football은 115개의 football team이 존재하고, Political blogs는 1,494개의 blogs로 구성되어 있다.

외부 커뮤니티 연관도를 군집화 알고리즘과 결합할 때 커뮤니티의 개수는 3부터 의미가 있게 된다. 커뮤니티의 개수가 2개이면, 무조건 외부 커뮤니티 연관도는 0이 되어서 최적으로 인식된다. 따라서 커뮤니티의 개수를 3부터 1씩 증가시키면서, 외부 커뮤니티 연관도와 커뮤니티 평균 순도를 측정하였다.

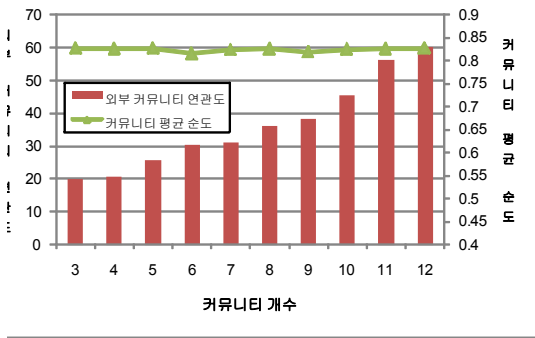


(그림 4) College football에 커뮤니티 개수를 증가시키면서, 외부 커뮤니티 연관도와 커뮤니티 평균 순도를 계산한 결과

College football에 대한 결과는 그림 4와 같다. 커뮤니티 개수가 3일 때 외부 커뮤니티 연관도는 낮았고, 커뮤니티 평균 순도는 가장 낮았다. 외부 커뮤니티 연관도는 커뮤니티 개수가 4부터 급격히 증가했다가, 커뮤니티 개수가 증가하면서 점점 감소하고 있고, 이에 반비례하여 커뮤니티 평균 순도는 증가하는 것을 확인할 수 있었다. 커뮤니티 개수가 3일 때의 외부 커뮤니티 연관도는 19.9이고, 커뮤니티 개수가 11일 때 외부 커뮤니티 연관도 18.5로 가장 낮았다. 이때 커뮤니티 평균 순도는 0.83이었다.

Political blogs의 실험 결과는 그림 5와 같다. Political blogs의 경우 기본 커뮤니티의 개수가 3이고, 외부 커뮤니티 연관도는 지속적으로 증가하였다. 제안하는 알고리즘으로는 커뮤니티 개수가 5일 때 종료하지만, 그 후에 최소값이 존재하는지 확인하기 위하여 12개까지 실험을 계속하였다. 실험결과 외부 커뮤니티 연관도는 지속적으로 증가하는 것을 확인할 수 있었다. 커뮤니티 개수가 3일 때의 외부 커뮤니티 연관도는

19.9이고, 커뮤니티 개수가 5일 때의 외부 커뮤니티 연관도는 25.7이다. 또한, 커뮤니티 개수가 12일 때의 연관도는 60.3이었다. 그러므로 최종 커뮤니티 개수는 3이 된다. 커뮤니티 개수가 3인 경우 외부 커뮤니티 연관도는 19.9였고, 커뮤니티 평균 순도는 0.827이었다.



(그림 5) Political blogs에 커뮤니티 개수를 증가시키면서, 외부 커뮤니티 연관도와 커뮤니티 평균 순도를 계산한 결과

#### 4.3 기존 알고리즘과 성능 비교

본 논문에서 제안한 외부 커뮤니티 연관도 방법의 성능을 기존 알고리즘과 비교하였다. 비교를 한 기존의 방법들은 Newman의 leading eigenvector 방법[12]과 Ghosh 등이 제안한 확장된 modularity-maximization 방법[13]이다. 실험에 사용한 데이터는 앞에서 사용한 데이터와 같은 College football과 Political blogs의 2개의 데이터 세트를 사용하였다. 실험 결과는 <표 3>과 같다.

제안하는 외부 커뮤니티 연관도를 이용한 방법은 College football 데이터에서는 가장 높은 커뮤니티 평균 순도를 기록했으나, Political blogs에 대해서는 두 번째 커뮤니티 평균 순도를 나타냈다. 커뮤니티 개수를 살펴보면, 외부 커뮤니티 연관도를 이용한 방법과 Newman의 방법이 거의 유사한 정확도로 커뮤니티 개수를 찾아내는 것을 확인할 수 있다. 각각을 살펴보면, 제안하는 방법이 기존 방법들을 압도하는 성능을 보이지는 않는다. 하지만, 커뮤니티 개수와 커뮤니티 평균 순도를 같이 생각해 보면, 외부 커뮤니티 연관도를 이용한 방법은 정확한 커뮤니티 개수를 발견하면서, 높은 커뮤니티 평균 순도를 보이는 것을 확인할 수 있다.

<표 3> 제안하는 외부 커뮤니티 연관도를 사용한 방법과 기존 알고리즘과의 비교 실험 결과

	College football		Political blogs	
	커뮤니티 개수	커뮤니티 순도	커뮤니티 개수	커뮤니티 순도
제안하는 방법	11	0.83	3	0.827
Newman	12	0.5865	2	0.892
Ghosh	6	0.5808	4	0.6457

### 5. 결 론

본 논문에서는 소셜 네트워크에서 커뮤니티를 찾기 위하여 군집화 방법을 적용하였다. 이때 커뮤니티의 개수에 따라 결과가 달라질 수 있기 때문에 커뮤니티의 개수 선정은 매우 중요하다. 따라서 이러한 개수 선정에 있어서 커뮤니티 간의 구별을 위한 외부 커뮤니티 연관도를 정의하고, 이를 기존의 군집화 알고리즘과 결합하여 커뮤니티의 개수를 동적으로 발견하였다. 최적의 커뮤니티를 찾기 위하여 군집화를 수행하는데 있어서 커뮤니티의 개수를 증가시키면서 최적의 커뮤니티를 찾도록 하였다. 따라서 최적의 커뮤니티의 개수를 찾는 시간을 단축하기 위하여 외부 커뮤니티 연관도가 높은 커뮤니티에 대해서만 분할/결합 등의 작업을 수행하는 유연한 군집화에 대해서 연구가 필요하다.

최근 기술 발달로 인하여 스마트폰의 활성화로 인하여 사진, 동영상 등 다양한 미디어에 대한 소셜 네트워크 서비스가 주목받고 있다. 따라서 다양한 미디어를 이용한 소셜 네트워크 서비스를 제공하기 위한 연구가 추가적으로 필요하다.

### 참 고 문 헌

[1] 최창현, 박건우, 이상훈, “지식검색 서비스에서의 소셜 네트워크 기반 영향력 지수 알고리즘”, 한국컴퓨터정보학회 논문집, Vol .14, No. 10, 2009.

[2] Yong-Yeol Ahn, James P. Bagrow, Sune Lehmann, “ Link communities reveal multiscale complexity in networks”, nature, vol. 466, pp.761-764. 2010.

[3] Richard Freeman, “Topological Tree Clustering of Social Network Search Results”, in Proceedings of the Eight International Conference on Lecture Notes in Computer Science (LNCS 4481), Springer, pp. 760-769, 2007.

[4] Romain Boulet, Bertrand Jouvea, Fabrice Rossi, Nathalie Villa, “Batch kernel SOM and related Laplacian methods for social network analysis”, Neurocomputing, Vol. 71, PP. 1257-1273, 2008.

[5] 강윤섭, 최승진, “공통 이웃 그래프 밀도를 사용한 소셜 네트워크 분석”, 정보과학회논문지:컴퓨팅 실재 및 레터, Vol. 16, No. 4, pp. 432-436, 2010.

[6] 장익진, “다차원 척도 분석법”, 연암사, 1998.

[7] Earl Gose, Richard Johnsonbugh and Steve Jost, “Pat tern Recognition and Image Analysis”, Prentice Hall, 1996.

[8] Douglass R. Cutting, David R. Karger, Jan O. Pedersen and John W. Tukey, “Scatter/Gather: a cluster-based approach to browsing large document collections”, In Proc. of the 15th annual international ACM SIGIR, June, pp. 318-329, 1992.

[9] Y. Wang, H. Song, W. Wang and M. An, “A microscopic view on community detection in complex networks”, Proceeding of the 2nd PhD Workshop on Information and Knowledge Management, New York, USA, pp. 57-64, 2008.

[10] M. E. J. Newman, “Modularity and community structure in networks”, Proceedings of the National Academy of Science, USA, 103(23):8577-8582, 2006.

[11] L. A. Adamic and N. Glance, “The Political Blogosphere and the 2004 U. S. Election: Divided They Blog”, Proceedings of the 3rd International Workshop on Link Discovery, Chicago, Illinois, pp. 36-43, 2005.

[12] M. E. J. Newman, “Finding community structure in networks using the eigenvectors of matrices”, Physical Review E, 74(3), 19 pages, 2006.

[13] R. Ghosh and K. Lerman, “Structure of Heterogeneous Networks”, International Conference on Computational Science and Engineering, vol. 4, pp.98-105, 2009.

**이 현 진**



1996년 : 순천향대학교 전산학과 학사  
 1998년 : 연세대학교 대학원 컴퓨터과학 석사  
 2002년 : 연세대학교 대학원 컴퓨터과학 박사

2003년~현재 : 한국사이버대학교  
 컴퓨터정보통신학과 부교수  
 관심분야 : 이러닝, 기계학습, 데이터마이닝

**지 대 창**



1997년 : 연세대학교 컴퓨터과학과 학사  
 1999년 : 연세대학교 컴퓨터과학과 석사  
 2004년 : ~ 현재 연세대학교 컴퓨터과학과 박사과정

1999년 ~현재: LG CNS 부책임 연구원  
 관심분야 : 패턴인식, 데이터마이닝, 스마트폰 기술