

자율 군집 네트워크에서 군집 탐지 기법

김바울*, 김경덕**, 김상욱*

*경북대학교 전자전기컴퓨터학부

**위덕대학교 컴퓨터공학과

e-mail : paulkim@media.knu.ac.kr

A Technique of Cluster Detection to Self-Organized Network

Paul Kim*, Kyungdeok Kim**, Sangwook Kim*

*School of Electrical Engineering and Computer Science,
Kyoungpook National University

**Dept. of Computer Engineering, Uiduk University

요 약

다양한 네트워크에서 군집을 분석하고 그 구조를 발견하는 것은 그 네트워크의 복잡도를 낮추어 전체 시스템을 이해하고 관리하는데 중요하다. 특히 기본적인 컴퓨팅이 가능한 여러 기기들이 자율적으로 서로 통신하여 군집을 이루는 자율 군집 네트워크에서 군집을 정확하게 발견하는 것은 집단행동 서비스를 실현하는데 있어서 중요한 기술이다. 따라서 본 연구에서는 자율 군집 네트워크에서 군집 탐지 기법을 제안한다. 제안하는 기법은 군집을 발견하고 그 군집을 식별하기 위해 해당 네트워크에서 한 노드를 공유하는 두 개의 간선 쌍에 대해 계층 군집화를 수행하고 계층 간에 간선 유사도를 계산하여 비교한다. 계층 군집화를 통한 간선들은 트리 구조로 표현할 수 있으며 최적의 분할 밀도를 이용하여 노드들을 클러스터링한 후 최종 군집으로 분리 한다.

1. 서론

다양한 시스템이나 그 구조를 네트워크로 표현하는 것은 객체 간에 상호작용을 쉽게 이해할 수 있다. 이런 네트워크는 대부분 군집 또는 커뮤니티 구조를 가진 서브 네트워크의 결합으로 구성되어 있고, 이런 서브 네트워크를 군집이라고 한다 [1].

특히 기본적인 컴퓨팅이 가능한 다양한 기기들이 자율적으로 통신하여 군집을 이루는 자율 군집 네트워크에서 그 군집을 발견하는 것은 집단행동 서비스를 실현하는데 있어서 중요한 기술이다. 따라서 본 연구에서는 자율 군집 네트워크에서 커뮤니티 구조를 발견하고 그 군집을 식별하는 방법을 제안한다. 제안하는 방법은 군집을 식별하기 위해 해당 네트워크에서 한 노드를 공유하는 두 개의 간선 쌍들에 대해 계층 군집화를 수행한다. 이 때 두 개의 간선에 대한 유사도를 계산할 수 있으며, 계층 군집화를 통한 간선들은 트리 구조로 표현할 수 있다. 이렇게 트리 구조로 표현한 네트워크에서 최적의 분할 밀도를 이용하여 노드들을 클러스터링 함으로써 군집을 발견할 수 있다. 본 논문은 2장에서 관련 연구를 설명하고, 3장에서 자율 군집 네트워크의 특징에 대해 기술한 후, 4장, 5장에서 군집 발견 기법에 대해 설명한다.

2. 관련 연구

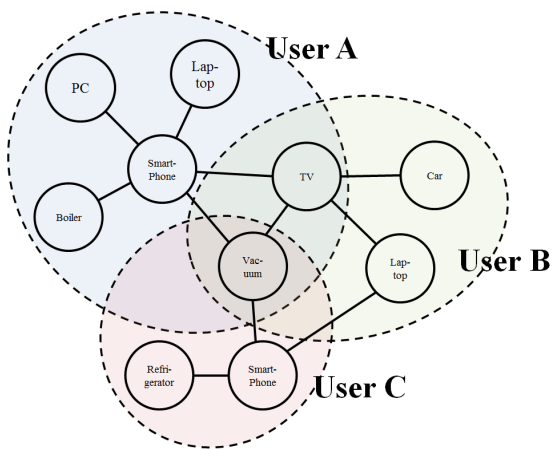
자율 군집 네트워크는 구조와 기능에 있어서 조직화된 시스템을 말하며, 다양한 분야에서 활용되고 있으며 연구되고 있다 [2][3]. 이는 구조화 측면에서 요소들이 특정한 방법으로 조직화되어야 하며, 기능적인 측면에서 서로 상호작용하여 전체 시스템이 특정한 목적을 달성하도록 해야 한다. 또한 중앙 집중적인 제어 없이 환경 변화에 민감한 요소들 사이에서 분산된 상호작용을 유도함으로써 군집 행동을 제어할 수 있다.

복잡한 네트워크에서 그 구조를 분석하여 커뮤니티 또는 군집을 발견하는 연구는 많이 연구되고 있다. 특히 시간에 따라 진화하는 동적 네트워크에 대한 연구가 최근 이슈가 되고 있으며, 네트워크가 진화하는 과정에서 커뮤니티 구조를 발견하는 연구가 활발하다 [4][5]. 또한 네트워크의 커뮤니티 사이에서 노드 간에 중복을 허용하는 커뮤니티 발견 방법들이 새롭게 연구 되고 있다 [6][7]. 하지만 이 방법들은 무작위 네트워크를 제외한 기본 네트워크를 전제로 하기 때문에 커뮤니티 발견의 정확도가 낮다. 따라서 본 연구에서는 자율 군집 네트워크를 분석하여 그 특징을 정의하고, 분석한 네트워크에서 커뮤니티를 발견하고 식별하는 방법을 제안한다.

3. 자율 군집 네트워크의 특성

자율 군집 네트워크를 지원하는 시스템은 그 시스템을 구성하는 여러 개체가 자발적 상호작용을 함으로써 객체 간에 행동, 규칙, 질서 등을 추론하여 새로운 환경에 유연하게 적응할 수 있다. 따라서 시스템 내, 외적인 변화가 발생하더라도 시스템을 구성하는 개별 요소들이 변화에 자동으로 반응하고 대처할 수 있다. 이런 자율 군집 네트워크는 그림 1과 같이 정점 집합 $V = \{v_1, \dots, v_n\}$ 과 간선 집합 $E = \{e_1, \dots, e_n\}$ 로 이루어진 그래프 $G = \{V, E\}$ 로 설명하고 표현할 수 있다. 여기에서 각 노드 v 는 자율 군집 시스템을 구성하는 각 요소를 말하며 그 요소 간의 인터랙션으로 간선 e 가 생성된다. 이렇게 그래프 이론을 자율 군집 시스템에 적용하는 것은 자율 군집 네트워크의 복잡도가 높더라도 그것을 그래프로 표현함으로써 분석이 용이하게 할 수 있다. 또한 자율 군집을 이루는 각 개체들을 노드로, 개체 간의 커넥션을 간선으로 맵핑한 그래프를 분석하여 커뮤니티를 발견할 수 있다. 그러므로 자율 군집 네트워크를 그래프로 표현하여 군집을 발견함으로써 전체 시스템을 유연하고 쉽게 제어할 수 있다.

자율 군집 시스템은 그 시스템을 구성하는 개별 요소들의 지역적인 개별 행동이 전체 시스템에 영향을 적게 주어야하며, 다른 요소들과 조화를 이루어야 한다[8]. 각 요소가 이런 제약들을 수용하면서 행동 규칙을 추론하기 위해서는 자율 군집 네트워크에서 커뮤니티를 정확하게 발견해야 한다. 이런 복잡한 네트워크에서 정확하게 커뮤니티를 발견하기 위해서는 해당 네트워크의 특징이 그래프에 나타나야 하며, 커뮤니티 발견 및 식별에 반영되어야 한다. 자율 군집 네트워크의 대표적인 특징은 간선이 가중치와 방향을 동시에 가지는 것이다. 또한 시간에 따라 진화하는 동적 네트워크이면서 노드의 유형이 여러 개인 Multi-partite 네트워크이다. 따라서 자율 군집 네트워크는



(그림 1) 자율 군집 네트워크의 예

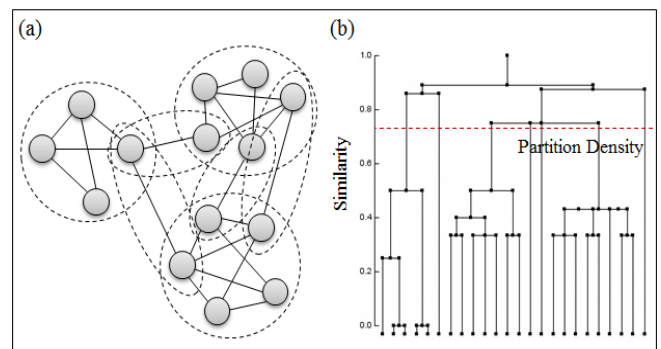
가중치 방향 그래프로도 표현할 수 있다.

자율 군집 네트워크의 간선은 두 개의 노드가 상호작용하여 서로 연결되는데, 이 연결된 간선이 일방적으로 인터랙션을 보내는 경우나 또는 받는 경우, 서로 인터랙션을 주고받는 경우 등의 유형이 존재한다. 이 간선의 방향은 두 개의 노드가 커넥션이 있을 때, 인터랙션을 가하는 주체에 따라 간선 방향이 결정된다. 이는 능동적 활동 유무를 가지는 노드들로 인해 방향성을 가지는 그래프 이론에 기반 한다. 예를 들어 i 노드가 j 노드에게 인터랙션을 줬다면, i 노드에서 j 노드까지의 방향을 가지는 간선을 부여할 수 있다. 이처럼 인터랙션으로 연결된 두 노드는 인터랙션을 일방적으로 받는 관계와 양방향으로 인터랙션을 갖는 두 가지의 커넥션 유형을 가질 수 있다. 또한 커뮤니티 구조의 특성상 커뮤니티 내부 노드들의 커넥션은 양방향을 가지는 경우가 많으며, 커뮤니티 간을 연결하는 커넥션들은 단방향을 가지는 경우가 대부분이다.

자율 군집 네트워크를 구성하는 노드들은 다양한 커넥션을 가지고 있으며, 각 커넥션은 인터랙션 정도에 따라 서로 다른 가중치를 가진다. 이는 네트워크를 구성하는 노드들의 인터랙션 양과 질을 정량적으로 측정함으로써 가중치 값을 부여할 수 있다. 여기에서 인터랙션 양은 커넥션이 일어나는 횟수, 유지 시간 등을 기반으로 하며, 인터랙션 질은 위치, 신뢰도, 친밀도 등을 기반으로 한다. 따라서 자율 군집 네트워크를 분석하여 각 간선의 가중치 값과 방향성을 정확하게 측정함으로써 좀 더 정확한 군집을 발견하고 식별할 수 있다.

4. 군집 발견

일반적인 네트워크에서 커뮤니티를 발견하고 식별하는 방법은 다양하다. 하지만 위와 같은 네트워크 특성을 수반하고 있는 자율 군집 네트워크에서 커뮤니티 구조를 발견하기 위해서는 그림 2와 같이 중복을 허용하는 계층 군집화(Hierarchical Clustering) 방법이 필요하다[7]. 계층 군



(그림 2) 커뮤니티 네트워크와 덴드로그램

집화는 노드나 간선을 계층적으로 클러스터링하기 때문에 아주 작은 단위의 커뮤니티가 모여 큰 커뮤니티를 구성하고, 이것이 최적의 커뮤니티 집합이 될 때까지 클러스터링하는 특징이 있다. 또한 시스템의 각 요소가 하나 이상의 군집에 소속되어야 하므로 노드의 중복이 허용되어야 한다. 따라서 중복을 허용하는 계층 군집화 방법은 자율 군집 네트워크처럼 구성 요소들 간에 포함 관계를 가지는 시스템에 적합하다. 중복을 허용하는 계층 군집화 방법은 해당 네트워크의 간선들을 계층적으로 클러스터링 하는 방법으로 간선 간에 유사도(Similarity)와 분할 밀도(Partition Density)를 계산해야 한다. 간선 간에 유사도는 수식 (1)과 같이 계산할 수 있다.

$$S(e_{ik}, e_{jk}) = \frac{|n_+(i) \cap n_+(j)|}{|n_+(i) \cup n_+(j)|} \quad (1)$$

여기에서 e_{ik} 와 e_{jk} 는 노드 k 를 공통으로 가지는 간선들이며, $n_+(i)$ 는 간선 e_{ik} 의 인접 노드 수이다. 이렇게 네트워크의 모든 간선들에 대한 유사도를 구한 후, 이 값을 바탕으로 계층 군집화를 수행 한다. 계층 군집화는 간선 쌍들의 모든 유사도 값을 오름차순으로 정렬하여 유사한 간선끼리 클러스터링 하는 과정이며, 동시에 분할 밀도를 계산하는 중간 과정이기도 하다. 이 과정에서 그룹핑되는 커뮤니티들이 발생하며 전체 간선이 하나의 커뮤니티로 포함될 때까지 반복한다. 이 때 각 커뮤니티에 대한 분할 밀도 D 는 수식 (2)와 같이 계산한다.

$$D = \frac{2}{M} \sum_c m_c \frac{m_c - (n_c - 1)}{(n_c - 2)(n_c - 1)} \quad (2)$$

여기에서 m_c 와 n_c 는 각각 커뮤니티 C 의 간선과 노드의 개수이다. 이 때 분할 밀도는 $0 \leq D \leq 1$ 의 값을 갖게 된다. 따라서 이 분할 밀도는 계층 군집화를 시작할 때 커

뮤니티가 하나도 없기 때문에 0이고, 분할 밀도가 1이면 전체 네트워크가 하나의 커뮤니티인 경우이다. 이렇게 클러스터링 진행 과정 속에 발견된 군집의 모듈성(Modularity)을 계산하게 된다. 만약 이 분할 밀도 값이 최대가 되는 군집 집합이 발견되면 이 군집 집합이 주어진 네트워크에서 최적의 군집이 된다.

5. 군집 식별

자율 군집 네트워크는 시간에 따라 진화하기 때문에 각 시간에 발견한 군집 집합 간을 서로 비교하여 변화하는 군집을 정확하게 식별해야 한다. 네트워크가 진화하는 과정은 $t-1$ 시간의 네트워크에서 노드와 간선이 추가되거나 삭제됨으로써 표현할 수 있고, 이를 확장하면 노드와 간선으로 이루어진 커뮤니티의 생성과 소멸, 결합으로 이해할 수 있다. 그러므로 시간에 따라 변화하는 네트워크를 그 네트워크를 구성하는 커뮤니티의 진화라고 할 수 있다. 이렇게 시간에 따라 노드의 추가와 삭제, 간선의 생성과 소멸 등이 자유로운 자율 군집 네트워크에서 군집을 정확하게 식별하는 것은 전체 시스템의 유연성을 높여 빠르게 변화된 환경에 적응할 수 있다. 하지만 동적 네트워크에서 커뮤니티를 식별하는 것은 NP-hard 문제로 계산 양이 많다. 이를 해결하기 위해서는 그림 3과 같이 군집을 발견하기 위해 클러스터링을 수행하는 과정 속에 군집을 식별하는 알고리즘을 포함해야 한다. 이렇게 함으로써 계산 시간을 줄일 수 있고 빠르게 변화하는 환경 변화에 적응할 수 있다. 또한 단 시간에 많은 변화가 일어나더라도 매 클러스터링 과정 속에 포함되어 있기 때문에 정확하게 커뮤니티들을 식별할 수 있다.

6. 요약

네트워킹이 가능한 다양한 기기들이 자율적으로 서로 통신하여 군집을 이루는 자율 군집 네트워크에서 군집을 발견하는 것은 시스템의 복잡도를 낮추고 변화하는 환경에 유연하게 적응할 수 있으며, 집단행동 서비스를 실현하는데 있어서 중요한 기술이다. 따라서 본 연구에서는 자율 군집 네트워크에서 군집 발견 및 식별 기법을 제안하였다. 제안하는 방법은 자율 군집 네트워크의 특성을 분석하고 이를 토대로 커뮤니티를 발견하고 방법과 식별하는 방법이다. 이 방법은 계층 군집화를 수행하여 계층 간에 간선 유사도를 비교하고 분할 밀도를 계산하여 클러스터링을 수행 한다. 또한 동적 네트워크에 유연하게 적응하기 위해 시간에 따라 변화하는 네트워크에서 각 시간에 발견한 군집을 서로 비교함으로써 정확한 군집을 식별하는 방법을 제안하였다.

Link Community Detection Algorithm	
Input:	$e^t = \{e_0, e_1, \dots, e_n\}$
Output:	$C^t = \{C_0, C_1, \dots, C_q\}$
1	<i>DetectLinkCommuni</i> ese(e^t)
2	for $i \leftarrow 0$ to $ e^t - 1$
3	<i>update</i> (e_i^t)
4	$C^t \leftarrow$ <i>Clustering</i> (e_i^t)
5	<i>IdentifyCommunities</i> (C^{t-1}, C^t)
6	$C^{t-1} \leftarrow C^t$
7	end for
8	end <i>DetectLinkCommunities</i>

(그림 3) 커뮤니티 발견 알고리즘

Acknowledgement

본 연구는 지식경제부 및 한국산업기술평가관리원의 산업융합원천기술개발사업(정보통신)의 일환으로 수행하였음. [10041145, 자율군집을 지원하는 웰빙형 정보기기 내장 소프트웨어 플랫폼 개발]

참고문헌

- [1] S. Fortunato, "Community Detection in Graphs," *Journal of Physics Reports*, vol. 486, pp. 75-174, 2010.
- [2] D. Garlaschelli, A. Capocci and G. Caldarelli, "Self-Organized Network Evolution Coupled to Extremal Dynamics," *Nature Physics*, vol. 3, pp. 813-817, 2007
- [3] C. Prehofer and C. Bettstetter, "Self-Organization in Communication Networks: Principles and Design Paradigms," *IEEE Communication Magazine*, vol. 43, no. 7, pp. 78-85, 2005.
- [4] D. Chakrabarti, R. Kumar and A. Tomkins, "Evolutionary Clustering," *Proceedings of 12th ACM SIGKDD International Conference on Knowledge Discovery and Data Mining*, pp. 554-560, 2006.
- [5] F. Folino and C. Pizzuti "Multiobjective Evolutionary Community Detection for Dynamic Networks," *Proceedings of the 12th Annual Conference on Genetic and Evolutionary Computation*, 2010.
- [6] G. Palla et al., "Uncovering the Overlapping Community Structure of Complex Networks in Nature and Society," *Nature*, vol. 435, no. 814, 2005.
- [7] Y. Ahn, J. P. Bagrow and S. Lehmann, "Link Communities Reveal Multiscale Complexity in Networks," *Nature*, vol. 466, no. 7307, 2010.
- [8] W. R. Ashby, "Principles of the Self-Organizing Dynamic System," *Journal of General Psychology*, vol. 37, pp. 125-128, 1947.