

# A social network monitoring procedure based on community statistics

Joo Weon Lee<sup>a</sup>, Jaeheon Lee<sup>1,a</sup>

<sup>a</sup>Department of Applied Statistics, Chung-Ang University

---

## Abstract

Recently, monitoring and detecting anomalies in social networks have become an interesting research topic. In this study, we investigate the detection of abnormal changes in a network modeled by the DCSBM (degree corrected stochastic block model), which reflects the propensity of both individuals and communities. To this end, we propose three methods for anomaly detection in the DCSBM networks: One method for monitoring the entire network, and two methods for dividing and monitoring the network in consideration of communities. To compare these anomaly detection methods, we design and perform simulations. The simulation results show that the method for monitoring networks divided by communities has good performance.

Keywords: abnormal detection, network monitoring, social network, statistical process monitoring

---

## 1. 서론

최근 컴퓨터와 인터넷의 발전과 함께 시간의 흐름에 따라 다양한 연결망 데이터가 생성되고 있다. 이렇게 시간의 흐름을 반영하는 연결망을 동적 연결망(dynamic network)이라 한다. 대표적인 예로는 트위터(twitter), 인스타그램(instagram), 페이스북(facebook) 연결망이 있으며 개인 및 회사에서 주고받는 이메일(e-mail) 연결망 등이 있다.

사회 연결망 분석(social network analysis; SNA)은 이러한 사회 연결망 내에 속하는 객체들 간의 관계 및 의사소통 패턴을 모니터링한다. 이때 연결망 내 개인이나 집단에서의 의사소통 패턴이 갑작스럽게 변화한다면 이는 사회 연결망에 비정상적인 변화가 발생했다고 할 수 있다. 사회 연결망의 변화에 대한 가장 대표적인 예로는 미국의 에너지 회사인 엔론(Enron)의 이메일 연결망 (Priebe 등, 2005; Shetty와 Adibi, 2005)이 있으며, 이때 회사의 파업 직전에 회사의 간부들 사이에서 이메일의 전송 트래픽 횟수가 비정상적으로 급증한 것으로 나타났다. 이뿐만 아니라 테러 (Krebs, 2002; Vargas, 2012; Everton과 Cunningham, 2013), 사기 (Chau 등, 2006; Fire 등, 2012) 등에서도 사회 연결망 내에서 비정상적인 변화가 발생한 것을 알 수 있다. 따라서 사회의 주요한 사건 전후에 사회 연결망에 변화가 나타나기 때문에, 이를 가능한 한 빠르게 탐지하는 것이 사회 연결망 분석의 주된 목표 중 하나이다.

사회 연결망에서의 비정상적인 변화 탐지를 위한 방법 중 하나는 연결망을 통계적 공정 모니터링(statistical process monitoring; SPM) 절차에 적용하는 것이다. 통계적 공정 모니터링의 주요 도구인 관리도(control

---

This work was supported by the National Research Foundation of Korea (NRF) grant funded by the Korea government (MSIT) (No. 2020R1F1A1A01050674).

<sup>1</sup>Corresponding author: Department of Applied Statistics, Chung-Ang University, 84 Heukseok-Ro, Dongjak-Gu, Seoul 06974, Korea. E-mail: [jaeheon@cau.ac.kr](mailto:jaeheon@cau.ac.kr)

chart)는 연결망을 실시간으로 모니터링하여 연결망의 일시적 혹은 영구적인 변화와 같은 이상 변화를 탐지하고 감시할 수 있다. 예를 들어, Hosseini와 Noorossana (2018)는 EWMA (exponentially weighted moving average) 관리도와 CUSUM (cumulative sum) 관리도를 사용하여 연결망의 다양한 변화에 대해 연결정도(degree)의 평균과 표준편차를 모니터링하고 이에 대한 성능을 비교하였다.

Wilson 등 (2019)은 연결망 모형 중 DCSBM (degree corrected stochastic block model)을 Shewhart 관리도에 적용하여 연결망의 지역적 변화와 전역적 변화를 모니터링하였다. 이 논문에서 고려하는 DCSBM에 대해서는 2장에서 상세하게 설명할 것이다. 연결망 모형으로 DCSBM을 가정하고 비정상적인 변화를 탐지하는 최근 연구는 다음과 같다. Yu 등 (2018)은 Hotelling의  $T^2$  관리도를 사용하여 개인의 경향성을 모니터링하는 다변량 절차를 제안하였다. Yu 등 (2022)은 연결망에서 중심성의 요약통계량(summary statistics)에 기초한 방법, Preibe 등 (2005)이 제안한 스캔 방법(scan method), Yu 등 (2018)과 Wilson 등 (2019)이 제안한 모형에 기초한 방법(model-based method)에 대한 성능을 다각적으로 비교하고 평가하였다. 또한 Lee와 Lee (2021)는 DCSBM에서 사용할 수 있는 자기출발(self-starting) 모니터링 절차를 제안하고 그 성능을 평가하였다.

이 논문은 DCSBM 연결망의 비정상적인 변화 탐지를 위해 세 가지의 모니터링 방법을 제안한다. 이를 간략하게 설명하면, 첫 번째 방법은 연결망 전체를 분할 없이 모니터링하는 방법이다. 두 번째 방법은 연결망을 전체 커뮤니티의 개수로 분할하여 모니터링하는 방법이고, 세 번째 방법은 연결망을 좀 더 세부적으로 분할하여 이를 모니터링하는 방법이다. 그러므로 두 번째 방법과 세 번째 방법은 분할한 연결망 중 한 곳에서라도 변화를 탐지하였을 경우, 연결망에 비정상적인 변화가 발생하였다고 판단하는 것이다. 제안된 세 가지 방법의 성능을 비교하기 위해 연결망의 다양한 변화를 고려하여 모의실험을 설계하고 시행하였다. 그리고 변화의 유형에 따라 어떤 모니터링 방법이 가장 빠르게 변화를 탐지하여 성능이 우수한지 평가하였다. 이를 위해 성능 비교의 척도로써 평균런길이(average run length; ARL)를 사용하였다. 이때 런길이는 비정상적인 변화가 발생했다는 신호가 줄 때까지 관측한 연결망의 개수를 나타낸다.

이 논문의 구성은 다음과 같다. 2장에서는 이 논문에서 사용하는 연결망 모형인 DCSBM을 소개하고, 3장에서는 연결망을 분할하여 모니터링하는 세 가지 방법을 제안한다. 4장과 5장에서는 제안된 방법의 성능을 비교하기 위하여 모의실험 절차 및 설계를 제시하고 모의실험 결과에 대하여 설명한다. 6장에서는 사례 연구를 설명하고, 마지막 7장에서는 이 논문에 대한 결론을 제시한다.

## 2. DCSBM

시점  $t (= 1, 2, \dots)$ 에서의 동적 사회 연결망 모형을  $G_t = (V, E_t)$ 로 나타내자. 이때  $V = \{1, 2, \dots, n\}$ 은 연결망에 소속한 노드(node)의 집합이고, 이는 연결망에 속하는 각각의 객체를 나타낸다. 그리고  $E_t$ 는 시점  $t$ 에서의 엣지(edge)의 집합이고, 이는 노드들 간의 의사소통 관계를 나타낸다. 이 논문에서는 시점에 따라 노드의 개수  $n$ 은 변하지 않는다고 가정한다.

사회 연결망은 엣지의 특성에 따라 가중 연결망(weighted network)과 비가중 연결망(unweighted network)으로 구분된다. 엣지의 집합은  $E_t = \{e_{i,j}(t) : i, j \in V\}$ 로 표현할 수 있는데, 가중 연결망의 경우  $e_{i,j}(t)$ 는 시점  $t$ 에서 노드  $i$ 와  $j$  사이의 의사소통의 정도 또는 횟수를 나타낸다. 반면 비가중 연결망의 경우  $e_{i,j}(t)$ 는 노드  $i$ 와  $j$  사이의 의사소통의 유무를 나타내며, 의사소통을 한 경우에는  $e_{i,j}(t) = 1$ , 의사소통을 하지 않은 경우에는  $e_{i,j}(t) = 0$ 이 된다. 또한 연결망은 노드 간 의사소통 방향성의 여부에 따라 무방향 연결망(undirected network)과 방향 연결망(directed network)으로 구분된다. 무방향 연결망의 경우 노드  $i$ 와  $j$ 의 의사소통 방향을 구분하지 않기 때문에  $e_{i,j}(t) = e_{j,i}(t)$ 가 되고, 방향 연결망의 경우 의사소통 방향을 구분하기 때문에  $e_{i,j}(t) \neq e_{j,i}(t)$ 가 된다. 또한 일반적으로 동일한 노드 간의 연결, 즉 자가 루프(self-loop)는 고려하지 않기 때문에, 이 논문에서도  $e_{i,i}(t) = 0$ 을 가정한다. 연결망은  $e_{i,j}(t)$ 를 원소로 하는 인접 행렬(adjacency matrix)로 표현할 수 있으며, 이는 연결망 그래프를 수리적으로 표현하여 연결망 분석을 용이하게 한다.

사회 연결망을 표현하기 위해 제안된 확률 모형으로 다양한 모형이 제시되었다. 예를 들면, 각 개인이나 개인들의 그룹인 커뮤니티(community)의 성향을 고려하지 않은 Erdos-Renyi 모형 (Erdos와 Renyi, 1959)을 필두로 Chung-Lu 모형 (Chung과 Lu, 2002), SBM (stochastic block model) (Holland 등, 1983; Snijders와 Nowicki, 1997; Nowicki와 Snijders, 2001), 그리고 DCSBM (degree corrected stochastic block model) (Karrer와 Newman, 2011) 등이 있다.

이중 SBM과 DCSBM은 비슷한 특성을 가진 노드들의 집합체인 커뮤니티(community)를 고려한 모형이다. 즉, 같은 커뮤니티에 속하는 노드들의 엣지의 크기는 서로 다른 집합에 속하는 노드들의 엣지의 크기보다 더 크며, 이것은 커뮤니티 내에 속하는 노드들 간의 의사소통 횟수가 더 많음을 나타낸다. SBM과 DCSBM에 대한 연구로 각 노드들을 커뮤니티로 구분하는 절차에 대해 많이 진행되었다 (Chin 등, 2015; Abbe 등, 2016; Zhang과 Zhou, 2016; Gao 등, 2018). 그러나 이 논문에서는 DCSBM에 의해 잘 설명이 되는 사회 연결망에 시간에 따라 관측할 때, 비정상적인 변화가 발생하는지 모니터링하는 절차에 대해 연구하고자 한다.

DCSBM은 Karrer와 Newman (2011)이 가중 연결망에 대해 제안하였으며, Xu와 Hero (2013)는 비가중 연결망에서의 DCSBM을 제안하였다. Wilson 등 (2019)은 동적 연결망에서의 DCSBM을 제안하였는데, 이는 무방향 가중 연결망(undirected weighted network)을 가정한 모형이다. 이 논문은 Wilson 등 (2019)이 제안한 무방향 가중 연결망에서의 DCSBM을 고려하는데, 이 연결망의 인접 행렬은 대각원소가 0인 대칭 행렬이 된다.

Karrer와 Newman (2011)과 Wilson 등 (2019)이 고려한 DCSBM은 실제 연결망과 유사한 두 가지 특성으로 (i) 커뮤니티 구조(community structure)와 (ii) 연결정도의 이질성(degree heterogeneity)을 반영한 모형이다. 이 연결망 모형  $G$ 에서  $V$ 는  $k(\geq 1)$ 개의 커뮤니티의 합집합( $V = V_1 \cup V_2 \cup \dots \cup V_k$ )으로 나타낼 수 있다. 또한 연결망에서 각 노드에 연결된 총 가중 엣지의 합을 연결정도(degree)라고 정의하는데, 노드  $i$ 의 연결정도  $d_i$ 는 연결망에서 노드  $i$ 와 의사소통을 하는 총 횟수를 의미하고 다음 식과 같이 정의된다 (시점  $t$ 는 생략했음). 이때 연결망 내에서 각 노드별 연결정도의 값이 다르기 때문에 연결정도의 이질성을 가진다고 할 수 있다.

$$d_i = \sum_{x \in V} e_{i,x}.$$

DCSBM은 모수  $(\theta, c, P)$ 에 의존하는데,  $\theta = \{\theta_i : i \in V\}$ 는 각 노드의 의사소통 경향성을 나타내는 모수이고,  $c = \{c_i : i \in V\}$ 는 각 노드가 속하는 커뮤니티의 라벨을 나타낸다. 또한  $P = \{P_{u,v} : u, v = 1, 2, \dots, k\}$ 는  $k \times k$  대칭 행렬이고,  $P_{u,v}$ 는 커뮤니티  $u$ 와  $v$ 의 의사소통 경향성을 나타내는 모수이다. 이 논문에서는  $(\theta, c, P)$ 가 주어질 경우 가중 엣지는 다음과 같이 서로 독립이고 모수가  $\lambda_{i,j}$ 인 포아송 분포를 따른다고 가정한다.

$$e_{i,j} \sim \text{Poisson}(\lambda_{i,j}), \quad \text{여기서 } \lambda_{i,j} = \theta_i \theta_j P_{c_i, c_j}, \quad (2.1)$$

이때 커뮤니티  $u$ 에 속하는 총 노드의 수는

$$n_u = \sum_{i \in V} \mathbb{I}(c_i = u)$$

와 같이 정의할 수 있으며, 여기서  $\mathbb{I}(\cdot)$ 는 주어진 조건을 만족하면 1, 아니면 0인 표시 함수(indicator function)를 나타낸다. DCSBM에서 연결망의 구별성(identifiable)을 위하여 일반적으로  $\theta_i$ 의 합은 노드  $i$ 가 속한 커뮤니티의 총 노드의 수와 동일하다는 제약을 준다. 즉,  $\theta_i$ 에 대하여 식 (2.2)와 같은 제약식이 주어진다. DCSBM에 대한 좀 더 상세한 내용은 Karrer와 Newman (2011)과 Wilson 등 (2019)을 참고할 수 있다.

$$\sum_{i: c_i = u} \theta_i = n_u. \quad (2.2)$$

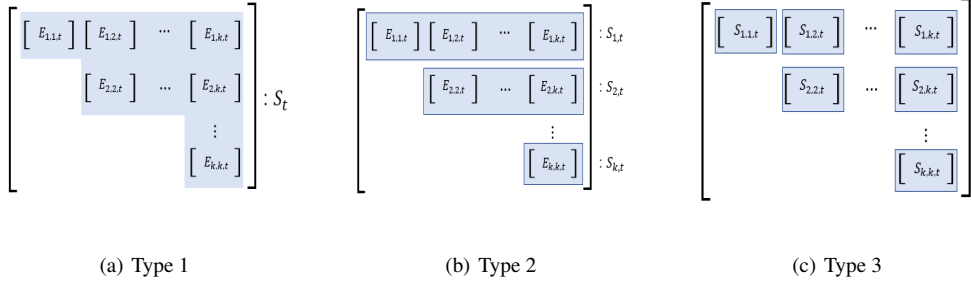


Figure 1: Three types of monitoring methods in social networks.

### 3. 모니터링 절차

이 장에서는 연결망의 비정상적인 변화 탐지를 위해 세 가지의 모니터링 방법을 제안하는데, 이 방법들은 통계적 공정 모니터링(statistical process monitoring; SPM)을 기반으로 한다.

연결망  $G_t$ 를 모니터링하기 위해 관리통계량을 설정하여 변화를 탐지하고자 한다. 관리통계량을 사용하여 모니터링하는 방법은 SPM의 가장 대표적인 도구인 관리도(control chart)를 사용하는 것이다. 즉, 관리도를 사용하여 시간에 따라 관측되는 관리통계량을 실시간으로 모니터링하여 연결망에 변화가 발생했는지 탐지하는 것이다. 관리도는 제1국면과 제2국면으로 나누어 적용된다. 제1국면에서는 미리 측정된  $m$ 개의 연결망  $G_t$  ( $t = 1, 2, \dots, m$ )에서 모형의 모수를 추정하는 단계이다. 이 추정량을 사용하여 관리도의 관리한계(control limit)를 설정할 수 있다. 이때 관리한계는 주어진 관리상태에서의 특성 (제1종 오류, 즉 오경보율  $\alpha$ , 또는 관리상태에서의 ARL인  $ARL_0$ 값)을 만족하도록 설정할 수 있다. 제2국면에서는 새로운 연결망  $G_t$  ( $t > m$ )에서 관리통계량을 계산하고 이 값이 관리한계를 벗어난 경우 연결망에서 비정상적인 변화가 발생하였다고 판단하게 된다.

이 논문에서는 무방향 DCSBM을 고려하기 때문에, 연결망의 인접 행렬은 대각 원소가 0인 대칭 행렬이 된다. 제안하는 세 가지 방법은 다음과 같다. 첫 번째 방법 (Type 1)은 연결망 전체를 분할 없이 모니터링하는 방법이다. 따라서 전체 연결망의 가중 엣지를 모두 더한 값을 통계량으로 사용한다. 이 방법에서 사용하는 시점  $t$ 에서 통계량  $S_t$ 는 식 (3.1)과 같으며 서로 독립이고 모수가  $\lambda$ 인 포아송 분포를 따른다. 이때  $e_{i,j}(t)$ 는 시점  $t$ 에서의  $e_{i,j}$ 를 나타낸다.

$$\text{Type 1 : } S_t = \sum_i \sum_j e_{i,j}(t) \sim \text{Poisson}(\lambda), \quad \text{여기서 } \lambda = \sum_i \sum_j \lambda_{i,j}. \quad (3.1)$$

두 번째 방법 (Type 2)은 연결망에서 노드  $i$ 가 속하는 커뮤니티를  $u$  ( $= 1, 2, \dots, k$ )라 할 때, 커뮤니티  $u$ 에 속하는 모든 가중 엣지를 더한 값을 통계량으로 사용한다. 즉, 전체 연결망을 총  $k$ 개로 분할하여 모니터링하는 것이다. 이 방법에서 사용하는 통계량  $S_{u,t}$ 는 식 (3.2)와 같으며 서로 독립이고 모수가  $\lambda_u$ 인 포아송 분포를 따른다.

$$\text{Type 2 : } S_{u,t} = \sum_{i:c_i=u} \sum_j e_{i,j}(t) \sim \text{Poisson}(\lambda_u), \quad \text{여기서 } \lambda_u = \sum_{i:c_i=u} \sum_j \lambda_{i,j}. \quad (3.2)$$

세 번째 방법 (Type 3)은 전체 연결망을 좀 더 세부적으로 분할하여 모니터링하는 방법이다. 노드  $i$ 가 속하는 커뮤니티를  $u$ , 노드  $j$ 가 속하는 커뮤니티를  $v$ 라 할 때, 이에 속하는 모든 가중 엣지의 합을 통계량으로

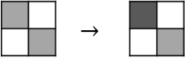
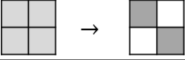
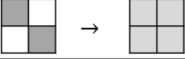
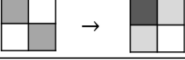
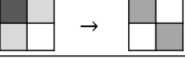
Changes	Network structure (P → P')
Intensified communication	
Split	
Merge	
Fragment	
Form	

Figure 2: Graphical representation of network changes.

사용한다. 따라서 동일한 커뮤니티 내에서 의사소통하는 가중 엣지의 합과 서로 다른 커뮤니티에서 의사소통하는 가중 엣지의 합을 각각 구분하여 모니터링하는 방법으로, 전체 연결망을 총  $k(k + 1)/2$ 개로 분할하여 모니터링하게 된다. 이 방법에서 사용하는 통계량  $S_{u,v,t}$ 는 식 (3.3)과 같으며 서로 독립이고 모수가  $\lambda_{u,v}$ 인 포아송 분포를 따른다.

$$\text{Type 3 : } S_{u,v,t} = \sum_{i:c_i=u} \sum_{j:c_j=v} e_{i,j}(t) \sim \text{Poisson}(\lambda_{u,v}), \quad \text{여기서 } \lambda_{u,v} = \sum_{i:c_i=u} \sum_{j:c_j=v} \lambda_{i,j}. \quad (3.3)$$

이해를 돕기 위해 세 가지 모니터링 방법을 Figure 1에 나타내었다 여기서  $E_{u,v,t}$ 는 분할된 연결망으로, 시점  $t$ 에서 커뮤니티  $u$ 와  $v$ 의 가중 엣지의 값을 원소로 가지는 소행렬을 나타낸다. Type 3에서는  $E_{u,v,t} = S_{u,v,t}$ 가 된다.

위 세 가지의 모니터링 방법에서 시점  $t$ 에서의 관리통계량을 정의하기 위해 먼저 다음과 같이 정규화를 수행한다. 가중 엣지의 합은 일반적으로 많은 원소들을 더한 값이기 때문에, 정규화된 통계량은 근사적으로 표준정규분포  $N(0, 1)$ 을 따른다고 할 수 있다.

$$\begin{aligned} \text{Type 1 : } Z_t &= \frac{S_t - \lambda}{\sqrt{\lambda}}, \\ \text{Type 2 : } Z_{u,t} &= \frac{S_{u,t} - \lambda_u}{\sqrt{\lambda_u}}, \\ \text{Type 3 : } Z_{u,v,t} &= \frac{S_{u,v,t} - \lambda_{u,v}}{\sqrt{\lambda_{u,v}}}. \end{aligned} \quad (3.4)$$

이때 식 (3.4)에 정의된 관리통계량에서  $Z_{u,t}$ 는 서로 다른  $u$ 에 대해 독립이고,  $Z_{u,v,t}$ 는 서로 다른  $u$ 와  $v$ 에 대해 독립이라는 것은 중요한 사실이다. 즉, Type 2와 Type 3에서 사용하는 각각의 관리도 절차는 서로 독립임을 알 수 있다.

연결망을 모니터링하기 위해 사용하는 관리통계량은 각각 다음과 같다.

$$\begin{aligned} \text{Type 1 : } & \max(|Z_t|), \\ \text{Type 2 : } & \max_u(|Z_{u,t}|), \\ \text{Type 3 : } & \max_{u,v}(|Z_{u,v,t}|). \end{aligned} \quad (3.5)$$

Table 1: Representation of parameter changes in social networks ( $k = 2$ )

$n$	Changes	$P$	$P^*$		
			Case 1	Case 2	Case 3
60	Intensified communication	$\begin{pmatrix} 0.50 & 0.10 \\ 0.10 & 0.50 \end{pmatrix}$	$\begin{pmatrix} 0.52 & 0.10 \\ 0.10 & 0.50 \end{pmatrix}$	$\begin{pmatrix} 0.54 & 0.10 \\ 0.10 & 0.50 \end{pmatrix}$	$\begin{pmatrix} 0.56 & 0.10 \\ 0.10 & 0.50 \end{pmatrix}$
	Split	$\begin{pmatrix} 0.30 & 0.30 \\ 0.30 & 0.30 \end{pmatrix}$	$\begin{pmatrix} 0.40 & 0.20 \\ 0.20 & 0.40 \end{pmatrix}$	$\begin{pmatrix} 0.42 & 0.20 \\ 0.20 & 0.42 \end{pmatrix}$	$\begin{pmatrix} 0.40 & 0.15 \\ 0.15 & 0.40 \end{pmatrix}$
	Fragment	$\begin{pmatrix} 0.40 & 0.20 \\ 0.20 & 0.40 \end{pmatrix}$	$\begin{pmatrix} 0.55 & 0.30 \\ 0.30 & 0.25 \end{pmatrix}$	$\begin{pmatrix} 0.55 & 0.35 \\ 0.35 & 0.30 \end{pmatrix}$	$\begin{pmatrix} 0.45 & 0.30 \\ 0.30 & 0.20 \end{pmatrix}$
150	Intensified communication	$\begin{pmatrix} 0.20 & 0.04 \\ 0.04 & 0.20 \end{pmatrix}$	$\begin{pmatrix} 0.21 & 0.04 \\ 0.04 & 0.20 \end{pmatrix}$	$\begin{pmatrix} 0.22 & 0.04 \\ 0.04 & 0.20 \end{pmatrix}$	$\begin{pmatrix} 0.23 & 0.04 \\ 0.04 & 0.20 \end{pmatrix}$
	Split	$\begin{pmatrix} 0.12 & 0.12 \\ 0.12 & 0.12 \end{pmatrix}$	$\begin{pmatrix} 0.14 & 0.10 \\ 0.10 & 0.14 \end{pmatrix}$	$\begin{pmatrix} 0.15 & 0.10 \\ 0.10 & 0.15 \end{pmatrix}$	$\begin{pmatrix} 0.14 & 0.08 \\ 0.08 & 0.14 \end{pmatrix}$
	Fragment	$\begin{pmatrix} 0.14 & 0.10 \\ 0.10 & 0.14 \end{pmatrix}$	$\begin{pmatrix} 0.16 & 0.13 \\ 0.13 & 0.12 \end{pmatrix}$	$\begin{pmatrix} 0.16 & 0.14 \\ 0.14 & 0.13 \end{pmatrix}$	$\begin{pmatrix} 0.15 & 0.13 \\ 0.13 & 0.12 \end{pmatrix}$

이는 모든 통계량을 관리한계와 비교하지 않고 절대값이 가장 큰 통계량만 관리한계와 비교하는 것으로 충분하기 때문이다. 전체 오경보율이  $\alpha$ 로 주어진 경우 각 Type  $i$  ( $i = 1, 2, 3$ )에서 관리한계  $L_i$ 는 식 (3.6)과 같이 설정할 수 있다.

$$L_i = \Phi^{-1} \left[ (1 - \alpha/2)^{\frac{1}{r_i}} \right], \quad i = 1, 2, 3, \quad (3.6)$$

여기서  $\Phi^{-1}[\cdot]$ 는  $N(0, 1)$ 의 누적분포함수의 역함수를 나타내고,  $r_i$ 는 각 모니터링 방법에서 분할되는 연결망의 총 개수로서 Type 1인 경우  $r_1 = 1$ , Type 2인 경우  $r_2 = k$ , Type 3인 경우  $r_3 = k(k+1)/2$ 가 된다.

따라서 관리도 절차는 각 방법에서 정의된 식 (3.5)의 관리통계량이 관리한계  $L_i$ 를 벗어날 경우 연결망에서 비정상적인 변화가 발생했다는 신호를 주게 된다. 이와 같은 유형의 관리도를 Shewhart 관리도라고 부른다. Shewhart 관리도에서 관리상태의 ARL인 ARL<sub>0</sub>와 전체 오경보율  $\alpha$ 는 ARL<sub>0</sub> =  $1/\alpha$ 의 관계가 있음이 잘 알려져 있다.

식 (3.6)에서 분할되는 연결망의 총 개수를 나타내는  $r_i$ 는 각 방법에서 사용하는 관리도의 총 개수가 되는 데, 이때  $r_i$ 가 커질수록 관리한계  $L_i$ 값도 커진다는 사실에 유의해야 한다. 이는 여러 개의 관리도를 병행하는 절차의 장점과 단점이 동시에 존재한다는 것을 의미한다.

#### 4. 모의실험

이번 장에서는 동적 DCSBM을 기반으로 이 논문에서 제안한 세 가지의 모니터링 방법의 성능을 비교하기 위하여 모의실험을 실시하였다. 모의실험에서 생성하는 동적 연결망  $G_t$ 는 연결망에서 의사소통이 정상적인 흐름에 있는 경우 모수  $(\theta, c, P)$ 에서 생성되지만, 비정상적인 변화가 발생한 경우 모수  $(\theta, c, P^*)$ 에서 생성된다고 가정하였다. 즉, 이 논문에서는 각 노드의 의사소통 경향성을 나타내는 모수  $\theta$ 와 각 노드가 속하는 커뮤니티의 라벨  $c$ 는 고정되어 있고, 커뮤니티의 의사소통 경향성을 나타내는 모수  $P$ 만 비정상적인  $P^*$ 로 변화하는 상황을 고려한다.

또한 연결망에 두개의 커뮤니티 ( $k = 2$ )를 구성하고 연결망에 속하는 총 노드의 수는  $n = 60$ 과  $n = 150$ 을 고려하였다. 이때 각 커뮤니티에 속하는 노드의 수는  $2n_1 = n_2$ 와  $n_1 = n_2$ 로 설정하였다. 또한 연결망에서 노드의 의사소통 경향성을 나타내는 모수  $\theta_t$ 는 균등 분포  $U(0.5, 1.5)$ 에서 랜덤하게 생성하고 제약식 (2.2)를

만족하도록 다음과 같이 변환하였다.

$$\theta_i = \frac{n_{c_i} \theta_i}{\sum_{j:c_j=c_i} \theta_j}.$$

이 논문에서는 모니터링 방법의 성능을 평가하기 위해 다음과 같이 다섯 가지 연결망 변화의 유형  $P^*$  를 고려하였다. 즉, (i) 한 커뮤니티 내의 의사소통 증가(intensified communication), (ii) 커뮤니티 분할(split), (iii) 커뮤니티 병합(merge), (iv) 커뮤니티 분열(fragment), (v) 커뮤니티 생성(form)인 경우이다. 먼저 (i)에서 한 커뮤니티 내의 의사소통 증가는 커뮤니티 1 내에서 의사소통 경향성이 증가 ( $P_{1,1} \rightarrow P_{1,1}^*$ )하는 것이다. (ii)의 커뮤니티 분할은 하나의 커뮤니티가 두 개의 커뮤니티로 분할하는 것이고, (iii)의 커뮤니티 병합은 두 커뮤니티가 하나의 커뮤니티로 합쳐지는 것으로 (ii)의 커뮤니티 분할과 반대되는 개념이다. 또한 (iv)의 커뮤니티 분열은 연결이 약해지는 노드들이 발생해서 두 커뮤니티 중 하나의 커뮤니티가 해체되는 것이고, (v)의 커뮤니티 생성은 서로 강하게 연결되지 않은 노드들이 합류하여 새로운 커뮤니티가 형성되는 것으로 (iv)의 커뮤니티 분열과 반대되는 개념이다. 보다 쉬운 이해를 위하여 이러한 다섯 가지 연결망 변화의 유형을 Figure 2에 표현하였다. 이는 Yu 등 (2022)이 모의실험에서 고려한 변화의 유형과 유사하다.

모의실험에서 고려하는 다섯 가지의 연결망 변화의 구체적인 유형을 Table 1에 요약하였다. 각 변화 유형에 따라 세 가지 케이스(case)를 고려하였고, 이때 커뮤니티 분할과 커뮤니티 병합, 커뮤니티 분열과 커뮤니티 생성은 각각 서로 반대되는 개념을 가지므로 대표적으로 커뮤니티 분할과 커뮤니티 분열을 표에 나타내었다. 예를 들어  $n = 60$ 일 때 커뮤니티 병합(merge)의 case 1은 커뮤니티 분할(split)의 반대 경우이기 때문에, 다음과 같이 변화하는 것이다.

$$\begin{pmatrix} 0.40 & 0.20 \\ 0.20 & 0.40 \end{pmatrix} \rightarrow \begin{pmatrix} 0.30 & 0.30 \\ 0.30 & 0.30 \end{pmatrix}.$$

이 논문에서는 연결망의 가장 엣지에 대한 모수  $\lambda_{i,j}$ 가 알려져 있거나, 또는 제1국면에서 정확하게 추정되었다고 가정한다. 물론 제1국면에서의 추정 오차에 대한 영향 또한 중요한 연구 과제이며, 이는 향후 연구를 진행할 계획이다. 앞에서 언급한 바와 같이 모니터링의 성능 비교를 위해 ARL을 사용하였고, 관리상태에서의 ARL은  $ARL_0 = 200$  ( $\alpha = 0.005$ )으로 설정하였다. 이 논문에서는 처음 시점부터 변화가 발생하였다고 가정했으며, 제2국면에서의 모의실험 절차는 각 모니터링 방법 Type  $i$  ( $i = 1, 2, 3$ )에 대해 다음과 같다. 참고로, 처음 시점부터 변화가 발생하였다고 가정하고 계산한 ARL을 zero-state ARL이라고 부른다.

- Step 1:** 시점  $t (= 1, 2, \dots)$ 에서  $\theta_i$  ( $i = 1, 2, \dots, n$ )를  $U(0.5, 1.5)$ 에서 랜덤하게 생성하고, 관리상태에서의 모수  $P$ 를 사용하여 식 (2.1)의  $\lambda_{i,j}$ 를 계산한다. 또한 이 값들을 이용하여 식 (3.1)–(3.3)의  $\lambda$ ,  $\lambda_u$ ,  $\lambda_{u,v}$ 를 계산한다.
- Step 2:** 모수  $(\theta, c, P^*)$ 에서 생성된  $G_t$ 에 대해  $S_t$ ,  $S_{u,t}$ ,  $S_{u,v,t}$ 를 계산하고, 식 (3.4)와 (3.5)에 의해 관리통계량을 계산한다. 이 값이 식 (3.6)에서 계산된 관리한계와 비교하여 관리한계를 벗어나지 않는다면  $t$ 를  $t + 1$ 로 증가시키면서 Step 2를 반복한다.
- Step 3:** 관리통계량이 관리한계를 벗어나서 변화의 신호가 발생하면 이때의 시점  $t$ 를 런길이로 기록한다.
- Step 4:** Steps 2–3의 과정을 10,000번 반복해서 10,000개의 런길이를 구한다.
- Step 5:** Steps 1–4의 과정을 10번 반복해서 총 100,000개의 런길이를 구하고, 런길이의 평균인 ARL을 계산한다.

## 5. 모의실험 결과

앞에서 설명한 모의실험 절차를 수행하여 각 변화 유형에 대한 ARL값을 Tables 2–6에 각각 제시하였다. 이때 각 케이스에 대해 제안된 세 가지 모니터링 방법 중 ARL값이 가장 작은 경우를 굵은 글씨체로 표기하였다.

Table 2: ARL values for the intensified communication scenario

	$(n_1, n_2) = (20, 40)$			$(n_1, n_2) = (30, 30)$		
	Type 1	Type 2	Type 3	Type 1	Type 2	Type 3
Case 1	166.46	<b>115.44</b>	133.43	105.78	<b>81.68</b>	96.44
Case 2	122.52	<b>48.97</b>	63.82	42.77	<b>23.95</b>	29.65
Case 3	85.00	<b>23.23</b>	29.01	19.01	<b>8.35</b>	10.24
	$(n_1, n_2) = (50, 100)$			$(n_1, n_2) = (75, 75)$		
	Type 1	Type 2	Type 3	Type 1	Type 2	Type 3
Case 1	126.30	<b>52.37</b>	63.79	45.58	<b>25.21</b>	31.52
Case 2	59.66	<b>11.20</b>	13.72	9.72	<b>3.80</b>	4.41
Case 3	28.10	<b>3.70</b>	4.28	3.34	<b>1.46</b>	1.57

Table 3: ARL values for the split scenario

	$(n_1, n_2) = (20, 40)$			$(n_1, n_2) = (30, 30)$		
	Type 1	Type 2	Type 3	Type 1	Type 2	Type 3
Case 1	45.68	3.07	<b>1.00</b>	196.30	1.24	<b>1.00</b>
Case 2	8.56	2.03	<b>1.00</b>	60.20	1.09	<b>1.00</b>
Case 3	32.01	2.45	<b>1.00</b>	4.40	1.01	<b>1.00</b>
	$(n_1, n_2) = (50, 100)$			$(n_1, n_2) = (75, 75)$		
	Type 1	Type 2	Type 3	Type 1	Type 2	Type 3
Case 1	59.76	5.70	<b>1.03</b>	198.68	1.80	<b>1.03</b>
Case 2	3.11	1.91	<b>1.00</b>	10.67	1.06	<b>1.00</b>
Case 3	4.32	1.80	<b>1.00</b>	1.59	<b>1.00</b>	<b>1.00</b>

### 5.1. 한 커뮤니티 내의 의사소통 증가(intensified communication)

연결망에서 커뮤니티 1 내에서의 의사소통 경향성이 증가( $P_{1,1} \rightarrow P_{1,1}^*$ )한 경우에 대한 결과가 Table 2에 나타나 있다.  $n = 60$ 인 경우 Case 1, Case 2, Case 3의  $P_{1,1}^*$ 값은 0.02씩 증가시켰고,  $n = 150$ 인 경우  $P_{1,1}^*$ 값은 0.01씩 증가시켰다.

Table 2의 결과를 살펴보면, Type 2 방법이 모든 Case에서 변화를 가장 빠르게 탐지하는 것으로 나타났다. 그 이유는 Type 1의 경우 전체 연결망에 대한 가중 엣지의 합을 통계량으로 사용하기 때문에, 커뮤니티 1에서의 변화가 크지 않은 경우 검정통계량의 변화가 크지 않은 것으로 판단된다. 또한 Type 3인 경우 Type 2에 비해 성능이 크게 떨어지는 않지만, 각각의 커뮤니티를 보다 세부적으로 모니터링하기 때문에 상대적으로 느리게 탐지된 것으로 판단된다. 따라서 하나의 커뮤니티 내에서만 변화가 발생한 경우 Type 2가 성능이 가장 좋음을 알 수 있다. 모든 모니터링 방법에서 각 커뮤니티의 노드 수가  $2n_1 = n_2$ 인 경우,  $n_1 = n_2$ 인 경우에 비해 커뮤니티 1에 속하는 노드의 수가 작기 때문에 비교적 느리게 변화가 탐지된 것으로 나타났다. 또한 당연한 결과이지만,  $P_{1,1}^*$ 값이  $P_{1,1}$ 값에 비해 더 커질수록 더 빨리 이를 탐지하는 것을 알 수 있다.

### 5.2. 커뮤니티 분할(split)과 커뮤니티 병합(merge)

커뮤니티 분할은 하나의 커뮤니티가 두개의 커뮤니티로 분할하는 것으로 Table 3에 이에 대한 성능의 결과를 나타내었다. 모의실험을 위해 Case 1은  $P_{1,1} \rightarrow P_{1,1}^*$ ,  $P_{2,2} \rightarrow P_{2,2}^*$ ,  $P_{1,2} \rightarrow P_{1,2}^*$ 의 증감의 크기를 동일하게 부여하였다. 예를 들어 총 노드의 수가 60일 때,  $P_{1,1}^*$ ,  $P_{2,2}^*$ ,  $P_{1,2}^*$ 는 모두 0.1씩 증가하거나 감소하였다. Case 2는  $P_{1,1}^*$ 과  $P_{2,2}^*$ 의 변화 크기를  $P_{1,2}^*$ 보다 크게 설정하였고, 반면 Case 3는  $P_{1,1}^*$ 과  $P_{2,2}^*$ 의 변화 크기를  $P_{1,2}^*$ 보다 작게 설정하여



Table 4: ARL values for the merge scenario

	$(n_1, n_2) = (20, 40)$			$(n_1, n_2) = (30, 30)$		
	Type 1	Type 2	Type 3	Type 1	Type 2	Type 3
Case 1	61.68	5.87	<b>1.00</b>	172.37	1.36	<b>1.00</b>
Case 2	11.12	3.39	<b>1.00</b>	82.47	1.15	<b>1.00</b>
Case 3	23.86	3.23	<b>1.00</b>	3.55	1.01	<b>1.00</b>
	$(n_1, n_2) = (50, 100)$			$(n_1, n_2) = (75, 75)$		
	Type 1	Type 2	Type 3	Type 1	Type 2	Type 3
Case 1	72.93	9.58	<b>1.03</b>	199.06	2.03	<b>1.03</b>
Case 2	3.50	2.53	<b>1.00</b>	12.12	1.13	<b>1.00</b>
Case 3	3.82	1.78	<b>1.00</b>	1.50	<b>1.00</b>	<b>1.00</b>

Table 5: ARL values for the fragment scenario

	$(n_1, n_2) = (20, 40)$			$(n_1, n_2) = (30, 30)$		
	Type 1	Type 2	Type 3	Type 1	Type 2	Type 3
Case 1	1.09	1.23	<b>1.00</b>	<b>1.00</b>	<b>1.00</b>	<b>1.00</b>
Case 2	1.74	1.57	<b>1.00</b>	<b>1.00</b>	<b>1.00</b>	<b>1.00</b>
Case 3	11.40	1.72	<b>1.00</b>	<b>1.00</b>	<b>1.00</b>	<b>1.00</b>
	$(n_1, n_2) = (50, 100)$			$(n_1, n_2) = (75, 75)$		
	Type 1	Type 2	Type 3	Type 1	Type 2	Type 3
Case 1	1.07	1.11	<b>1.00</b>	<b>1.00</b>	<b>1.00</b>	<b>1.00</b>
Case 2	1.13	1.31	<b>1.00</b>	<b>1.00</b>	1.01	<b>1.00</b>
Case 3	51.61	6.81	<b>1.00</b>	1.19	2.31	<b>1.00</b>

모의실험을 수행하였다.

모의실험 결과 Type 3 방법이 모든 Case에서 가장 빠르게 변화를 탐지하는 것으로 나타났다. 이는 모든 커뮤니티에서 의사소통 경향성이 변화( $P_{1,1} \rightarrow P_{1,1}^*, P_{2,2} \rightarrow P_{2,2}^*, P_{1,2} \rightarrow P_{1,2}^*$ )하였기 때문에, Type 3가 가장 빠르게 변화를 탐지하는 것으로 판단된다. 특히 ARL값이 1에 가까운 값을 갖는다는 것은 변화가 발생하는 즉시 이를 탐지한다는 것을 나타낸다. 한편 커뮤니티를 고려하지 않고 연결망 전체를 모니터링하는 Type 1이 가장 변화를 늦게 탐지한다. 특히 각 커뮤니티의 노드의 수를 동일하게( $n_1 = n_2$ ) 설정한 경우, Case 1에서  $P_{1,1} \rightarrow P_{1,1}^*, P_{2,2} \rightarrow P_{2,2}^*, P_{1,2} \rightarrow P_{1,2}^*$ 의 증감이 서로 상쇄되는 효과가 있기 때문에 Type 1의 성능이 나쁘게 나타난 것으로 판단된다. 그러나 Case 2, 3과 같이 의사소통 경향성에 대한 변화의 크기를 다르게 설정한 경우, Type 1은 상대적으로 좀 더 나은 성능을 나타냈다.

또한 Type 1은 Case 1, 2에서는  $2n_1 = n_2$ 인 경우의 성능이  $n_1 = n_2$ 인 경우에 비해 더 좋게 나타났고, Case 3에서는  $n_1 = n_2$ 인 경우의 성능이 더 좋게 나타났다. 그러나, Type 2는 모든 Case에서  $n_1 = n_2$ 인 경우의 성능이 더 좋게 나타났다. 이는 변화의 크기가 큰 소행렬에 속하는 원소가 더 많을수록 변화가 각 방법의 통계량에 더 잘 반영되기 때문이라고 판단된다.

커뮤니티 병합은 두 커뮤니티가 하나의 커뮤니티로 합쳐지는 것으로 커뮤니티 분할과 반대되는 개념이다. Table 4에 이에 대한 성능의 결과를 나타내었다. 커뮤니티 분할과 마찬가지로 Type 3 방법이 모든 Case에서 가장 빠르게 변화를 탐지하는 것으로 나타났고, 전반적인 경향은 커뮤니티 분할의 경우와 유사하다.

Table 6: ARL values for the form scenario

	$(n_1, n_2) = (20, 40)$			$(n_1, n_2) = (30, 30)$		
	Type 1	Type 2	Type 3	Type 1	Type 2	Type 3
Case 1	1.12	1.36	<b>1.00</b>	<b>1.00</b>	<b>1.00</b>	<b>1.00</b>
Case 2	1.96	2.11	<b>1.00</b>	<b>1.00</b>	1.01	<b>1.00</b>
Case 3	14.73	2.44	<b>1.00</b>	<b>1.00</b>	1.01	<b>1.00</b>
	$(n_1, n_2) = (50, 100)$			$(n_1, n_2) = (75, 75)$		
	Type 1	Type 2	Type 3	Type 1	Type 2	Type 3
Case 1	1.08	1.14	<b>1.00</b>	<b>1.00</b>	<b>1.00</b>	<b>1.00</b>
Case 2	1.16	1.41	<b>1.00</b>	<b>1.00</b>	1.01	<b>1.00</b>
Case 3	60.41	1.51	<b>1.00</b>	1.23	3.40	<b>1.00</b>

Table 7: Representation of parameter changes in social networks ( $k = 3$ )

Changes	$P$	$P^*$					
		Case 1			Case 2		
Intensified communication	$\begin{pmatrix} 0.20 & 0.04 & 0.04 \\ 0.04 & 0.20 & 0.04 \\ 0.04 & 0.04 & 0.20 \end{pmatrix}$	$\begin{pmatrix} 0.22 & 0.04 & 0.04 \\ 0.04 & 0.20 & 0.04 \\ 0.04 & 0.04 & 0.20 \end{pmatrix}$			$\begin{pmatrix} 0.24 & 0.04 & 0.04 \\ 0.04 & 0.20 & 0.04 \\ 0.04 & 0.04 & 0.20 \end{pmatrix}$		
	$\begin{pmatrix} 0.12 & 0.12 & 0.12 \\ 0.12 & 0.12 & 0.12 \\ 0.12 & 0.12 & 0.12 \end{pmatrix}$	$\begin{pmatrix} 0.12 & 0.10 & 0.10 \\ 0.10 & 0.15 & 0.10 \\ 0.10 & 0.10 & 0.15 \end{pmatrix}$			$\begin{pmatrix} 0.15 & 0.10 & 0.10 \\ 0.10 & 0.15 & 0.10 \\ 0.10 & 0.10 & 0.15 \end{pmatrix}$		
	$\begin{pmatrix} 0.14 & 0.10 & 0.10 \\ 0.10 & 0.14 & 0.10 \\ 0.10 & 0.10 & 0.14 \end{pmatrix}$	$\begin{pmatrix} 0.16 & 0.13 & 0.13 \\ 0.13 & 0.12 & 0.13 \\ 0.13 & 0.13 & 0.12 \end{pmatrix}$			$\begin{pmatrix} 0.16 & 0.13 & 0.13 \\ 0.13 & 0.16 & 0.13 \\ 0.13 & 0.13 & 0.12 \end{pmatrix}$		

### 5.3. 커뮤니티 분열(fragment)과 커뮤니티 생성(form)

커뮤니티 분열은 연결이 약해지는 노드들이 발생해서 두 커뮤니티 중 하나의 커뮤니티가 해체되는 경우이며, Table 5에 이에 대한 성능의 결과를 나타내었다. Case 1은  $P_{1,1} \rightarrow P_{1,1}^*$ 은 비교적 크게 증가하고,  $P_{2,2} \rightarrow P_{2,2}^*$ 은 비교적 크게 감소한 경우를 나타낸다. 또한 Case 2는  $P_{1,1} \rightarrow P_{1,1}^*$ 과  $P_{1,2} \rightarrow P_{1,2}^*$ 를 비교적 크게 증가시켜, 커뮤니티 1 내에서의 의사소통 경향성과 커뮤니티 1과 2 사이의 의사소통 경향성이 비교적 크게 증가한 경우를 나타낸다. Case 3는  $P_{2,2} \rightarrow P_{2,2}^*$ 가 상대적으로 크게 감소한 경우를 나타낸다. 이는 커뮤니티 분할과 병합의 경우와 유사하게 모든 커뮤니티의 의사소통 경향성이 변화( $P_{1,1} \rightarrow P_{1,1}^*$ ,  $P_{2,2} \rightarrow P_{2,2}^*$ ,  $P_{1,2} \rightarrow P_{1,2}^*$ )하였기 때문에, 연결망을 세부적으로 분할하여 모니터링하는 Type 3 방법의 성능이 가장 좋은 것으로 나타났다. 특히 Type 3는 모든 Case에서 ARL값이 1이기 때문에 변화가 발생하는 즉시 이를 탐지하는 것을 알 수 있다.

전반적으로  $P_{1,1}^*$ 과  $P_{2,2}^*$ 가 비교적 크게 변화한 Case 1의 경우 빠르게 변화를 탐지하며, Case 3의 경우 상대적으로 느리게 변화를 탐지하였다. 특히 Type 1 방법의 성능이 제일 좋지 않음을 알 수 있다.

커뮤니티 생성은 서로 강하게 연결되지 않은 노드들이 합류하여 새로운 커뮤니티가 형성되는 경우를 의미하며 이는 커뮤니티 분열과 반대되는 개념이다. Table 6에 이에 대한 성능 결과를 나타내었다. 커뮤니티 분열과 마찬가지로 Type 3 방법이 모든 Case에서 가장 빠르게 변화를 탐지하는 것으로 나타났고, 전반적인 경향은 커뮤니티 분열의 경우와 유사하다.

Table 8: ARL values when  $k = 3$ 

Scenario	Case	$(n_1, n_2, n_3) = (25, 25, 100)$			$(n_1, n_2, n_3) = (50, 50, 50)$		
		Type1	Type2	Type3	Type1	Type2	Type3
Intensified Communication	Case 1	153.21	<b>32.67</b>	46.08	18.79	<b>5.21</b>	6.96
	Case 2	88.29	<b>5.10</b>	6.88	3.13	<b>1.13</b>	1.19
Split	Case 1	1.05	1.08	<b>1.00</b>	196.10	34.95	<b>1.01</b>
	Case 2	1.02	1.06	<b>1.00</b>	7.55	1.18	<b>1.00</b>
Merge	Case 1	1.06	1.10	<b>1.00</b>	192.23	37.05	<b>1.02</b>
	Case 2	1.03	1.09	<b>1.00</b>	8.81	1.28	<b>1.00</b>
Fragment	Case 1	71.91	14.31	<b>1.00</b>	1.10	1.33	<b>1.00</b>
	Case 2	196.06	3.06	<b>1.00</b>	<b>1.00</b>	1.01	<b>1.00</b>
Form	Case 1	61.64	24.06	<b>1.00</b>	1.12	1.49	<b>1.00</b>
	Case 2	184.84	4.57	<b>1.00</b>	<b>1.00</b>	1.02	<b>1.00</b>

#### 5.4. 커뮤니티의 수가 $k = 3$ 인 경우

이 논문의 모의실험에서 고려한 커뮤니티의 수는  $k = 2$ 이다. 그러나 커뮤니티의 수가 증가할 때, 제안된 방법의 특성에 변화가 있는지 살펴보기 위해  $k = 3$ 인 경우 몇가지 변화에 대해 모의실험을 수행하였다.  $k = 3$ 인 경우 고려한 변화를 Table 7에 나타내었다. 이때 총 노드의 수는  $n = 150$ 이고, 각 커뮤니티에 속하는 노드의 수는  $4n_1 = 4n_2 = n_3$ 와  $n_1 = n_2 = n_3$ , 즉  $(n_1, n_2, n_3) = (25, 25, 100)$ 과  $(50, 50, 50)$ 으로 설정하였다. 나머지 설정은  $k = 2$ 인 경우와 동일하다.

모의실험 절차를 수행하여 각 변화 유형에 대한 ARL값을 Table 8에 제시하였다.  $k = 3$ 인 경우의 결과는  $k = 2$ 인 경우와 거의 유사한 경향을 나타내었다. 즉, 한 커뮤니티 내의 의사소통 증가(intensified communication)의 경우 Type 2의 성능이 가장 좋았지만, 나머지 변화에 대해서는 Type 3의 성능이 가장 좋게 나타남을 알 수 있다.  $k = 3$ 인 경우의 결과를 살펴볼 때, 다른  $k$ 에 대해서도 유사한 경향이 있을 것이라 예상된다.

## 6. 사례 연구

이번 장에서는 실제 사례에 대해 제안된 방법을 적용하여 그 성능을 알아보려고 한다. 사용한 연결망은 Farahani 등 (2017)이 사용한 사례이다. 이 사회 연결망은 2011년–2014년 동안 이란의 한 대학의 정보공학과 소속 대학원생 50명이 서로 주고 받은 이메일 횟수에 대한 것이다. 그들의 연구에서 이 사회 연결망을 포아송 분포에 적합시켰고 모수  $\lambda_{i,j}$ 에 대한 모형을 설정하였는데, 여기서는 그 모형에서 성별만 포함시킨 식을 사용하며 그 식은 다음과 같다.

$$\lambda_{i,j} = 10.2 + 1.3D_1 - 1.7D_2 + \epsilon_{i,j}, \quad (6.1)$$

여기서  $D_1$ 과  $D_2$ 는 성별을 나타내는 더미 변수인데, 노드  $i$ 와  $j$ 가 여성인 경우  $(D_1, D_2) = (0, 0)$ , 남성인 경우  $(1, 0)$ , 그리고 여성과 남성 또는 남성과 여성인 경우  $(0, 1)$ 으로 정의된다. 따라서 성별을 커뮤니티로 사용하고,  $k = 2$ 로 생각할 수 있다.

우리는 포아송 모형에서 식 (6.1)의 모수를 관리상태로 간주하여  $t \leq 30$ 의 연결망을 생성하였고,  $t \geq 31$ 의 연결망은 식 (6.2)의 이상상태의 모수에서 생성하였다. 이때  $n_1 = n_2 = 25$ 이고,  $\epsilon_{i,j}$ 는  $N(0, 1)$ 에서 생성하였다.

$$\lambda_{i,j} = 10.2 + (1.3 + \delta)D_1 - 1.7D_2 + \epsilon_{i,j}. \quad (6.2)$$

위의 식은 우리가 사용한 모형에서 남성 커뮤니티의 경향성이  $\delta$ 만큼 증가했다고 볼 수 있다. 즉, 한 커뮤니티 내의 의사소통 증가의 경우가 된다.

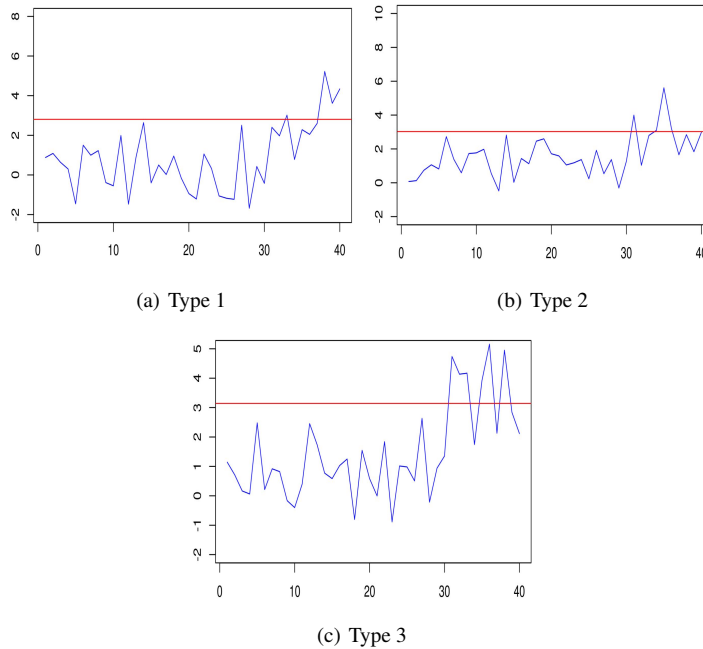


Figure 3: Control charts for three types of monitoring methods.

전체 오경보율은  $\alpha = 0.005$ 로 설정하였고, 식 (3.6)의 관리한계  $L_i$ 를 계산해서 각각  $L_1 = 2.807$ ,  $L_2 = 3.023$ ,  $L_3 = 3.144$ 를 얻었다. 그리고 식 (6.2)에서  $\delta$ 는 0.5를 고려하였다. 이와 같은 설정에서 각 방법을 사용하여 작성한 관리도를 Figure 3에 제시하였다. Figure 3을 살펴보면, Type 1, 2, 3는 각각  $t = 35, 31, 32$ 에서 관리한계를 벗어났기 때문에 Type 2, 3, 1의 순서로 변화를 빨리 탐지하는 것을 알 수 있다. 이 결과는 모의실험에서 나타난 특성과 동일한 결과이다.

## 7. 결론

사회 연결망은 연결망 내에 속하는 객체들 간의 관계뿐 아니라 의사소통 패턴과 같은 다양한 정보를 제공한다. 이러한 사회 연결망에서 개인이나 집단에서의 의사소통 패턴이 갑작스럽게 변화하였다면, 이는 연결망에서 비정상적인 변화가 발생하였다고 할 수 있다. 예를 들어 기업의 파업, 테러, 사기 등 사회에서의 주요한 사건 직전에 사회 연결망에서 비정상적인 의사소통 패턴이 발생하였다. 따라서 사회 연결망 분석에서 중요한 목표 중 하나는 가능한 한 빠르게 연결망에서의 변화를 탐지하는 것이다.

이 논문은 사회 연결망 모형 중 DCSBM을 기반으로 하였다. DCSBM은 실제 연결망과 유사한 특징을 많이 가지고 있기 때문에, 이 모형을 기반으로 연결망을 모니터링하는 세 가지의 방법을 제안하고 각 방법의 성능을 비교하였다. 첫 번째 방법(Type 1)은 연결망 전체를 분할 없이 모니터링하는 방법이다. 두 번째 방법(Type 2)은 연결망을 총 커뮤니티의 개수로 분할하여 모니터링하는 방법이고, 세 번째 방법(Type 3)은 연결망을 세부적으로 분할하여 커뮤니티 안팎에서 발생하는 의사소통을 각각 모니터링하는 방법이다.

연결망의 변화에 따른 제안된 방법의 성능을 비교하기 위해 모의실험을 수행하였다. 모의실험을 위해 다섯 가지의 연결망의 변화 유형을 제안하였고, 각 유형별로 세 가지 케이스로 변화를 부여하였다. 모의실험 결과 하나의 커뮤니티 내에서만 변화가 발생한 경우 Type 2 방법의 성능이 가장 좋았지만, 커뮤니티 분할,

커뮤니티 병합, 커뮤니티 분열, 커뮤니티 생성에서는 Type 3 방법의 성능이 가장 좋은 것으로 나타났다. 물론 커뮤니티의 의사소통 경향성의 변화만을 고려했기 때문에 Type 2와 3의 성능이 좋을 것이라고 기대되었지만, 사회 연결망에서 커뮤니티 경향성의 변화가 예상되는 상황에서는 Type 2와 3, 특히 Type 3 방법의 사용을 권장한다.

향후 좀 더 다양한 변화의 유형에 대하여 이를 효율적으로 탐지할 수 있는 절차와, 앞에서 잠시 언급한 바와 같이 제1국면에서 모수를 추정할 경우 추정 오차로 인한 영향에 대해 연구를 진행할 계획이다.

## References

- Abbe E, Bandeira AS, and Hall G (2016). Exact recovery in the stochastic block model, *Information Theory IEEE Transactions on*, **62**, 471–487.
- Chau DH, Pandit S, and Faloutsos C (2006). Detecting fraudulent personalities in networks of online auctioneers, In *Proceedings of European Conference on Principles of Data Mining and Knowledge Discovery*, Berlin, Germany, 103–114.
- Chin P, Rao A, and Vu V (2015). Stochastic block model and community detection in sparse graphs: A spectral algorithm with optimal rate of recovery, In *Proceedings of The 28th Conference on Learning Theory*, Paris, France, 391–423.
- Chung F and Lu L (2002). The average distances in random graphs with given expected degrees, *Proceedings of the National Academy of Sciences*, **99**, 15879–15882.
- Erdos P and Renyi A (1959). On random graphs I, *Publicationes Mathematicae*, **6**, 290–297.
- Everton SF and Cunningham D (2013). Detecting significant changes in dark networks, *Behavioral Sciences of Terrorism and Political Aggression*, **5**, 94–114.
- Farahani EM, Kazemzadeh RB, Noorossana R, and Rahimian G (2017). A statistical approach to social network monitoring, *Communications in Statistics-Theory and Methods*, **46**, 11272–11288.
- Fire M, Katz G, and Elovici Y (2012). Strangers intrusion detection-detecting spammers and fake profiles in social networks based on topology anomalies, *Human Journal*, **1**, 26–39.
- Gao C, Ma Z, Zhang AY, and Zhou HH (2018). Community detection in degree-corrected block models, *The Annals of Statistics*, **46**, 2153–2185.
- Holland PW, Laskey KB, and Leinhardt S (1983). Stochastic blockmodels: First steps, *Social Networks*, **5**, 109–137.
- Hosseini SS and Noorossana R (2018). Performance evaluation of EWMA and CUSUM control charts to detect anomalies in social networks using average and standard deviation of degree measures, *Quality and Reliability Engineering International*, **34**, 477–500.
- Karrer B and Newman ME (2011). Stochastic block models and community structure in networks, *Physical Review E*, **83**, 016107-1 – 016107-10.
- Krebs VE (2002). Mapping networks of terrorist cells, *Connections*, **24**, 43–52.
- Lee JW and Lee J (2021). Self-starting monitoring procedure for the dynamic degree corrected stochastic block model, *The Korean Journal of Applied Statistics*, **34**, 25–38.
- Nowicki K and Snijders TA (2001). Estimation and prediction for stochastic blockstructures, *Journal of the American Statistical Association*, **96**, 1077–1087.
- Priebe CE, Conroy JM, Marchette DJ, and Park Y (2005). Scan statistics on Enron graphs, *Computational and Mathematical Organization Theory*, **11**, 229–247.

- Shetty J and Adibi J (2005). Discovering important nodes through graph entropy the case of Enron email database, In *Proceeding of the 3rd International Workshop on Link Discovery*, Chicago, USA, 74–81.
- Snijders TA and Nowicki K (1997). Estimation and prediction for stochastic blockmodels for graphs with latent block structure, *Journal of Classification*, **14**, 75–100.
- Vargas JA (2012). Spring awakening: How an Egyptian revolution began on facebook, *The New York Times*, Sunday Book Review, Available from: <https://www.nytimes.com/2012/02/19/books/review/how-an-egyptian-revolution-began-on-facebook.html>
- Wilson JD, Stevens NT, and Woodall WH (2019). Methods for monitoring multiple proportions when inspecting continuously, *Journal of Quality Technology*, **43**, 237–248.
- Xu KS and Hero AO (2013). Dynamic stochastic block models: Statistical models for time evolving networks, In *Social Computing, Behavioral-Cultural Modeling and Prediction*, Washington DC, USA, 201–210.
- Yu L, Woodall WH, and Tsui KL (2018). Detecting node propensity changes in the dynamic degree corrected stochastic block model, *Social Networks*, **54**, 209–227.
- Yu L, Zwetsloot IM, Stevens NT, Wilson JD, and Tsui KL (2022). Monitoring dynamic networks: A simulation-based strategy for comparing monitoring methods and a comparative study, *Quality and Reliability Engineering International*, **38**, 1226–1250.
- Zhang AY and Zhou HH (2016). Minimax rates of community detection in stochastic block models, *The Annals of Statistics*, **44**, 2252–2280.

Received March 23, 2023; Revised May 5, 2023; Accepted May 8, 2023

# 커뮤니티 통계량에 기반한 사회 연결망 모니터링 절차

이주원<sup>a</sup>, 이재현<sup>1,a</sup>

<sup>a</sup>중앙대학교 응용통계학과

---

## 요약

최근 사회 연결망에서 비정상적인 변화를 모니터링하는 절차는 흥미로운 연구 주제이다. 이 논문은 사회 연결망 모형 중 커뮤니티와 개인들의 경향성을 모두 고려한 동적 연결망 모형인 DCSBM (degree corrected stochastic block model)을 가정하고 이 연결망 내의 변화를 모니터링하는 절차를 고려하였다. 이때 커뮤니티의 비정상적인 변화 탐지를 위해 세 가지의 모니터링 방법을 제안하였다. 또한 제안된 방법의 성능을 평가하기 위해 모의실험을 설계하고 수행하였다. 커뮤니티의 경향성 변화에 대한 모의실험 결과 연결망을 커뮤니티에 따라 분할하여 모니터링하는 방법이 전반적으로 빠르게 변화를 탐지하여 성능이 더 좋음을 알 수 있었다.

주요용어: 비정상적인 변화 탐지, 사회 연결망, 연결망 모니터링, 통계적 공정 모니터링

---

이 논문은 정부 (과학기술정보통신부)의 재원으로 한국연구재단의 지원을 받아 수행된 연구임 (No. 2020R1F1A1A01 050674).

<sup>1</sup>교신저자: (06974) 서울특별시 동작구 흑석로 84, 중앙대학교 응용통계학과. E-mail: jaeheon@cau.ac.kr