

Monitoring social networks based on transformation into categorical data

Joo Weon Lee^a, Jaeheon Lee^{1,a}

^aDepartment of Applied Statistics, Chung-Ang University, Korea

Abstract

Social network analysis (SNA) techniques have recently been developed to monitor and detect abnormal behaviors in social networks. As a useful tool for process monitoring, control charts are also useful for network monitoring. In this paper, the degree and closeness centrality measures, in which each has global and local perspectives, respectively, are applied to an exponentially weighted moving average (EWMA) chart and a multinomial cumulative sum (CUSUM) chart for monitoring undirected weighted networks. In general, EWMA charts monitor only one variable in a single chart, whereas multinomial CUSUM charts can monitor a categorical variable, in which several variables are transformed through classification rules, in a single chart. To monitor both degree centrality and closeness centrality simultaneously, we categorize them based on the average of each measure and then apply to the multinomial CUSUM chart. In this case, the global and local attributes of the network can be monitored simultaneously with a single chart. We also evaluate the performance of the proposed procedure through a simulation study.

Keywords: average run length, closeness centrality, degree centrality, exponentially weighted moving average (EWMA) chart, multinomial cumulative sum (CUSUM) chart, social network monitoring

1. Introduction

Social network analysis (SNA) provides information about social and individual relationships by graphically and mathematically monitoring global structures and local entities. Network surveillance applies statistical process monitoring (SPM) to a network and aims to quickly detect network anomalies. The network is monitored using control charts to detect anomalous behavior in each period of the network data. Therefore, network surveillance can help prevent conflicts and crises in social situations by identifying global and local perspectives. Many studies have addressed network-surveillance schemes. For example, Hosseini and Noorossana (2018) compared the performance of detecting outbreaks by applying the average and standard deviation of degree measures of a network to exponentially weighted moving average (EWMA) and cumulative sum (CUSUM) charts. Wilson *et al.* (2019) applied a degree corrected stochastic block model (DCSBM) to Shewhart charts to detect local and global changes. Yu *et al.* (2018) proposed a multivariate procedure to monitor an individual's propensity by using the Hotelling T^2 chart. Moreover, Yu *et al.* (2021) compared several monitoring

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

¹ Corresponding author: Department of Applied Statistics, Chung-Ang University, 84, Heukseok-ro, Dongjak-gu, Seoul 06974, Korea. E-mail: jaeheon@cau.ac.kr

Published 31 July 2022 / journal homepage: <http://csam.or.kr>

© 2022 The Korean Statistical Society, and Korean International Statistical Society. All rights reserved.

methods using the global network summary statistics, the scan-based method proposed by Preibe *et al.* (2005), and the model-based methods proposed by Yu *et al.* (2018) and Wilson *et al.* (2019).

Networks comprise nodes (vertices) and edges (links), where nodes represent actors or entities and edges represent communication between nodes. Networks can be divided into unweighted or weighted networks according to the types of edges. Edges in unweighted networks take a value of 0 or 1 depending on whether two nodes are communicating or not. Edges in weighted networks take non-negative values such as the number of communications between two nodes. Additionally, networks can be divided into two main categories: directed and undirected networks depending on whether the network is directional or not. An undirected network does not consider the direction of communication between two nodes. For example, given nodes i and j for $i \neq j$, communication between nodes i and j is the same as communication between nodes j and i . Otherwise, it is a directed network, or digraph. A network graph can be represented by an adjacency matrix that illustrates the edges between nodes. Network data can be analyzed more easily by mathematically representing the networks using adjacency matrices.

Centrality measures indicate which node is most important in the network. Freeman (1977, 1978) proposed degree, closeness, and betweenness centrality measures in unweighted networks. Degree centrality is the number of edges directly connected to a node and is representative of the node's popularity and communication activity. Scott (1991) argued that degree centrality can be regarded as local centrality. However, closeness and betweenness centralities are considered measures of global centrality in terms of the distance among various nodes. Closeness and betweenness centralities are associated with the number of nodes that lie within the shortest distance from all other nodes in the network, helping clarify the global network structure. Brandes (2001), Newman (2001), and Barrat *et al.* (2004) generalized these measures and applied them to weighted networks. Brandes (2001) and Newman (2001) considered the shortest path algorithm proposed by Dijkstra (1959) under the assumption that the most critical actor or entity is on the shortest path and can easily access all nodes. Opsahl *et al.* (2010) proposed a generalized form that combines the number of edges and their weights in order to consider both measures in a weighted network. Abassi *et al.* (2011) discussed the relationship between scholar's performance and SNA measures using normalized centrality measures. Abassi *et al.* (2013) also proposed hybrid centrality measures that combine existing centrality measures to improve the importance of nodes, and applied them to a co-author network to evaluate the proposed measures.

Because centrality measures in weighted networks are continuous variables, previous studies on SNA have generally used continuous statistics to evaluate performance through control charts. However, if several centrality measures are transformed into one categorical variable through classification, only a single chart can be used to monitor social networks more effectively. By using this method, the problem for monitoring social networks using centrality measures turns into that of monitoring categorical processes. Recently, Perry (2020) classified the observed networks over time into two categories based on the average values of transitivity and reciprocity measures, and applied them to an EWMA chart to monitor the hierarchical tendency. Furthermore, he compared the out-of-control average run length (ARL) performance of the proposed EWMA chart with that of the multinomial CUSUM chart proposed by Ryan *et al.* (2011).

A brief explanation of the multinomial CUSUM chart proposed by Ryan *et al.* (2011) is as follows: They proposed the multinomial CUSUM chart for monitoring in situations where items were classified into more than two categories, and they compared with multiple Bernoulli CUSUM charts. The multinomial CUSUM chart is useful when monitoring more than two categories with specified direction of out-of-control shifts, and unlike other charts which monitor categorical data, this chart

does not require aggregation of items into samples.

In this paper, we propose a procedure for creating a categorical variable having four categories by classifying degree and closeness centrality measures according to their average values in weighted networks, and then monitoring this variable using the multinomial CUSUM chart. By applying the four categories to the multinomial CUSUM chart, we can simultaneously monitor both local and global views of the network. We also compare the performance of the proposed procedure with that of the EWMA charts using degree centrality and closeness centrality individually.

The next sections are as follows: In Section 2, we define degree and closeness centrality measures, and in Section 3, we describe the control charts used in the simulations. In Section 4, we propose a procedure for monitoring one categorical variable transformed from two centrality measures, and in Section 5, we describe the simulation settings and discuss the simulation results. Finally, we provide our conclusions in Section 6.

2. Centrality measures

2.1. Degree centrality

Degree centrality indicates the node's communication activity and popularity. It has the advantage of being uncomplicated and easy to calculate. Degree is a local centrality measure because it is calculated based on the number of neighbors of a focal node. Freeman (1977, 1978) defined degree centrality as the number of directly connected nodes in a binary network. The degree centrality for node i in unweighted networks is defined as

$$D(i) = \sum_j^N x_{ij},$$

where x_{ij} is a binary value of 1 when nodes i and j interact with each other and 0 otherwise, and N is the total number of nodes.

However, since this definition does not take into account the weights representing the frequency of communications or the closeness between nodes, the use of this measure in weighted networks can lead to significant information loss. Barrat *et al.* (2004) proposed the degree centrality for node i in weighted networks as in Equation (2.1) considering weights.

$$D_w(i) = \sum_j^N w_{ij}, \quad (2.1)$$

where w_{ij} denotes the weight between nodes i and j .

2.2. Closeness centrality

Closeness centrality is the inverse of the shortest path between nodes, indicating that an efficient node is close to all the nodes. This centrality measure is related to global centrality. For example, a node may have a high degree centrality but a low value of closeness centrality in the network because the node is popular locally but not globally. There has been great interest in defining the shortest path between nodes. It assumes that the higher the number of intermediate nodes, the higher the communication cost. For example, a node with many intermediary nodes takes more time to acquire information, which distorts and delays the communication between nodes.

Freeman (1977, 1978) proposed the closeness centrality for node i in unweighted networks defined as

$$C(i) = \left[\sum_j^N d(i, j) \right]^{-1},$$

where $d(i, j)$ is the shortest distance between nodes i and j and is defined as

$$d(i, j) = \min(x_{ig} + \cdots + x_{gj})$$

for intermediary nodes g on paths between node i and j .

However, this measure has the disadvantage of losing a lot of the information contained in weighted networks. Newman (2001) considered the relationship between the weight and distance by using Dijkstra (1959)'s algorithm. The closeness centrality for node i in weighted networks proposed by Newman (2001) is given by

$$C_w(i) = \left[\sum_j^N d_w(i, j) \right]^{-1}, \quad (2.2)$$

where $d_w(i, j)$ is the shortest distance used Dijkstra's algorithm and is defined as

$$d_w(i, j) = \min\left(\frac{1}{w_{ig}} + \cdots + \frac{1}{w_{gj}}\right)$$

for intermediary nodes g on paths between node i and j .

3. Control charts

3.1. EWMA chart

The EWMA chart proposed by Roberts (1959) is useful for monitoring and detecting small shifts in the process parameter. Despite having similar performance to CUSUM charts, EWMA charts are widely used due to their ease and simplicity of application and implementation.

We define $X_t, t = 1, 2, \dots$, as the independent observations at time t . The EWMA chart statistic, the upper control limit (UCL), and the lower control limit (LCL) are defined as follows:

$$\begin{aligned} Z_t &= \gamma X_t + (1 - \gamma) Z_{t-1}, \\ \text{UCL} &= \mu_0 + L \sigma_0 \sqrt{\frac{\gamma}{2 - \gamma} [1 - (1 - \gamma)^{2t}]}, \\ \text{LCL} &= \mu_0 - L \sigma_0 \sqrt{\frac{\gamma}{2 - \gamma} [1 - (1 - \gamma)^{2t}]}, \end{aligned} \quad (3.1)$$

where μ_0 is the in-control mean, σ_0 is the in-control standard deviation of X_t , γ ($0 < \gamma \leq 1$) is the smoothing coefficient, and L is the factor that determines the width of the control limits. The initial value of chart statistic, Z_0 , is usually set to be equal to μ_0 . For smoothing coefficients, it is well known that smaller values of γ are better at detecting small changes, and larger values are better at detecting large changes. If the value of γ is chosen, L can be determined by the practitioner according to the specified value of in-control ARL, ARL_0 .

As t increases, the above control limits converge to the asymptotic limits given by

$$\text{UCL} = \mu_0 + L\sigma_0 \sqrt{\frac{\gamma}{2-\gamma}}, \quad (3.2)$$

$$\text{LCL} = \mu_0 - L\sigma_0 \sqrt{\frac{\gamma}{2-\gamma}}. \quad (3.3)$$

In this paper, we use these asymptotic control limits to evaluate the performance of EWMA charts.

3.2. Multinomial CUSUM chart

Traditionally, SPM uses Shewhart p -charts and binomial CUSUM charts to monitor the proportion of non-conforming items in the process. To use these charts, we must wait until the required number of items ($n > 1$) are aggregated into samples to monitor the proportion of nonconforming items. As a result, these charts have the disadvantage of having an inherent delay in detecting shifts in the proportion, which causes slower detection of process changes. To overcome this disadvantage, Reynolds and Stoumbos (1999) proposed using a Bernoulli CUSUM chart when a continuous stream of inspected items is available. The Bernoulli CUSUM chart is a special case of the binomial CUSUM chart when $n = 1$.

Although these charts mentioned above can only be used in situations where items are classified into two categories, there are many situations where items are classified into more than two categories. For example, instead of classifying items in a process as good or bad, they can be categorized as good, fair, or bad. Similarly, the diagnosis of posttraumatic patients can be divided into four categories: survival, minor complications, major complications, and death. Ryan *et al.* (2011) proposed a multinomial CUSUM chart to monitor several categories by extending the Bernoulli CUSUM chart.

Let X_1, X_2, \dots be independent multinomial random variables where $X_t = u$ if the t th item classified into the u th category, $t = 1, 2, \dots, u = 1, 2, \dots, v$. Let $p_{0,u}$ and $p_{1,u}$ be the in-control and out-of-control probabilities of being classified into the u th category, respectively, where $p_{0,u} \neq p_{1,u}$, $\sum_{u=1}^v p_{0,u} = 1$ and $\sum_{u=1}^v p_{1,u} = 1$.

The multinomial CUSUM statistics are defined as

$$S_t = \max(0, S_{t-1} + R_t), \quad t = 1, 2, \dots, \quad (3.4)$$

where $S_0 = 0$ and R_t is the log-likelihood ratio score. The values of R_t are

$$R_t = \ln\left(\frac{p_{1,u}}{p_{0,u}}\right), \quad t = 1, 2, \dots, u = 1, 2, \dots, v, \quad (3.5)$$

when $X_t = u$. The multinomial CUSUM chart signals when $S_t > h$, where the control limit, h , can be determined satisfying the specified value of ARL_0 .

Note that the multinomial CUSUM procedure can be generalized to include samples of size $n > 1$ (Ryan *et al.*, 2011).

4. Proposed monitoring procedure

Let $w_{ij,t}$ be the number of communications between nodes i and j at time t . In this paper, we assume that the network is undirected and does not consider any node to communicate with itself, and also assume that $w_{ij,t}$ follows a Poisson distribution as

$$w_{ij,t} \sim \text{Poisson}(\lambda) \quad \text{for } i \neq j,$$

where λ represents the in-control mean of weights.

We can calculate degree centralities $D_{w,t}(i)$ and closeness centralities $C_{w,t}(i)$ for $i = 1, 2, \dots, N$ at time t defined in (2.1) and (2.2), respectively, and then can obtain the average degree centrality $\overline{\text{Deg}}_t$ and the average closeness centrality $\overline{\text{Close}}_t$ as

$$\overline{\text{Deg}}_t = \frac{\sum_{i=1}^N D_{w,t}(i)}{N}, \quad (4.1)$$

$$\overline{\text{Close}}_t = \frac{\sum_{i=1}^N C_{w,t}(i)}{N}. \quad (4.2)$$

These average measures can be classified into the following four categories to be applied to the multinomial CUSUM chart.

- (i) Category 1 : $\overline{\text{Deg}}_t \geq \mu_d$ and $\overline{\text{Close}}_t \geq \mu_c$,
- (ii) Category 2 : $\overline{\text{Deg}}_t < \mu_d$ and $\overline{\text{Close}}_t \geq \mu_c$,
- (iii) Category 3 : $\overline{\text{Deg}}_t < \mu_d$ and $\overline{\text{Close}}_t < \mu_c$,
- (iv) Category 4 : $\overline{\text{Deg}}_t \geq \mu_d$ and $\overline{\text{Close}}_t < \mu_c$,

where μ_d is the in-control mean of degree and μ_c is the in-control mean of closeness which can be estimated from Phase I samples. In this paper, the values of μ_d , μ_c , and $p_{0,u}$ ($u = 1, 2, 3, 4$) in (3.5) are estimated from 100,000 simulated networks.

The procedure proposed in this paper is summarized as follows: First, $D_{w,t}(i)$ and $C_{w,t}(i)$ are calculated for N nodes at time t , and $\overline{\text{Deg}}_t$ and $\overline{\text{Close}}_t$ are calculated. After that, by determining which category these values belong to according to the above classification rule, the R_t in (3.5) and the multinomial CUSUM statistic, S_t , in (3.4) are calculated. Finally, by comparing the value of this statistic with the control limit h , it is determined whether or not to give a signal.

5. ARL performance

5.1. Simulation settings

Through simulation, the performance of the proposed multinomial CUSUM procedure is evaluated by comparing it with the EWMA procedure using degree centrality and closeness centrality separately. In the EWMA procedure, $\overline{\text{Deg}}_t$ in (4.1) and $\overline{\text{Close}}_t$ in (4.2) are used as X_t in (3.1), respectively. The multinomial CUSUM chart and the EWMA charts are compared based on their ARL values.

The settings of the simulation are determined by considering (i) the network to have $N = 30$ nodes, (ii) the in-control mean of weights as $\lambda = 1$ and 2, and (iii) in the presence of an anomalous event within the social network, the mean of weights between K (≤ 30) randomly chosen abnormal nodes to change from λ to $(1 + \delta)\lambda$. In our simulations, K , which refers to the number of abnormal nodes, takes on values of 2, 3, 4, and 6, while δ , which denotes the change in parameter, takes on values of 0.5, 1, 2, 3, 4, 5, 6, 7, and 8. For example, the case where $\lambda = 1$, $K = 2$, and $\delta = 2$ means that among 30 nodes, the weight of two nodes has changed to 3, and the weight of the remaining nodes is 1 without change.

Note that in order to apply the multinomial CUSUM chart, it is necessary to specify the out-of-control probabilities of each category, $p_{1,u}$ ($u = 1, 2, 3, 4$), in (3.5). Table 1 shows the in-control probabilities, $p_{0,u}$, estimated from Phase I samples and the out-of-control probabilities, $p_{1,u}$, desired

Table 1: The in-control and out-of-control probabilities for categories

		$p_{0,1}$	$p_{0,2}$	$p_{0,3}$	$p_{0,4}$
		0.4065	0.0899	0.4206	0.0830
$\lambda = 1$	K	$p_{1,1}$	$p_{1,2}$	$p_{1,3}$	$p_{1,4}$
	2	0.4445	0.0943	0.3822	0.0790
	3	0.5312	0.0974	0.3001	0.0714
	4	0.6503	0.0904	0.2036	0.0557
	6	0.8927	0.0393	0.0448	0.0232
		$p_{0,1}$	$p_{0,2}$	$p_{0,3}$	$p_{0,4}$
		0.4180	0.0777	0.4181	0.0863
$\lambda = 2$	K	$p_{1,1}$	$p_{1,2}$	$p_{1,3}$	$p_{1,4}$
	2	0.4497	0.0793	0.3879	0.0832
	3	0.5068	0.0823	0.3347	0.0763
	4	0.5957	0.0804	0.2578	0.0662
	6	0.8084	0.0549	0.0988	0.0379

Table 2: Values of L for the EWMA chart and values of h for the multinomial CUSUM chart

EWMA		$\gamma = 0.05, L = 1.878617$	
		$\gamma = 0.10, L = 2.147571$	
		$\gamma = 0.20, L = 2.359552$	
Multinomial CUSUM	$\lambda = 1$	$K = 2$	$h = 0.69$
		$K = 3$	$h = 1.68$
		$K = 4$	$h = 2.41$
		$K = 6$	$h = 3.146$
	$\lambda = 2$	$K = 2$	$h = 0.58$
		$K = 3$	$h = 1.30$
		$K = 4$	$h = 2.05$
		$K = 6$	$h = 3.15$

to be detected. These out-of-control probabilities are set to probabilities of $\delta = 2$ when $\lambda = 1$ and to probabilities of $\delta = 1$ when $\lambda = 2$ for each abnormal nodes size K .

To determine the control limits for control charts, the in-control ARL value is set as 100. We design three different EWMA charts with $\gamma = 0.05, 0.1,$ and $0.2,$ and obtain the values of L in (3.2) and (3.3) using the software provided by Knoth (2021). The control limits, $h,$ for the multinomial CUSUM chart, are determined using 100,000 simulations. The values of L in (3.2) and (3.3) for the EWMA chart and the values of h for the multinomial CUSUM chart are displayed in Table 2. As shown in Table 1, in the multinomial CUSUM chart, since the values of $p_{0,u}$ and $p_{1,u}$ vary according to λ and $K,$ the values of h satisfying $ARL_0 = 100$ in each case are different. However, in the EWMA chart, the average centralities in (4.1) and (4.2) approximately follow a normal distribution, respectively, and $L,$ which represents the width of the control limits, depends only on $\gamma.$ Therefore, the values of L satisfying $ARL_0 = 100$ are determined to be the same value for a given $\gamma.$

The out-of-control ARL values for the EWMA chart are obtained from 10,000 simulations, while the values for the multinomial CUSUM chart are obtained from 100,000 simulations.

5.2. Simulation results

We look at how the occurrence of abnormal nodes changes the probability of each category. Figure 1 illustrates a scatterplot of the in-control and out-of-control probabilities for degree and closeness centralities of $\lambda = 1.$ In Figure 1(a) shows the scatterplot for the in-control state, and shows that the probabilities of category 1 and 3 components are similar. However, from Figure 1(b), it can be seen that when abnormal nodes of $K = 4$ and $\delta = 6$ occur, the probability of category 1 increases

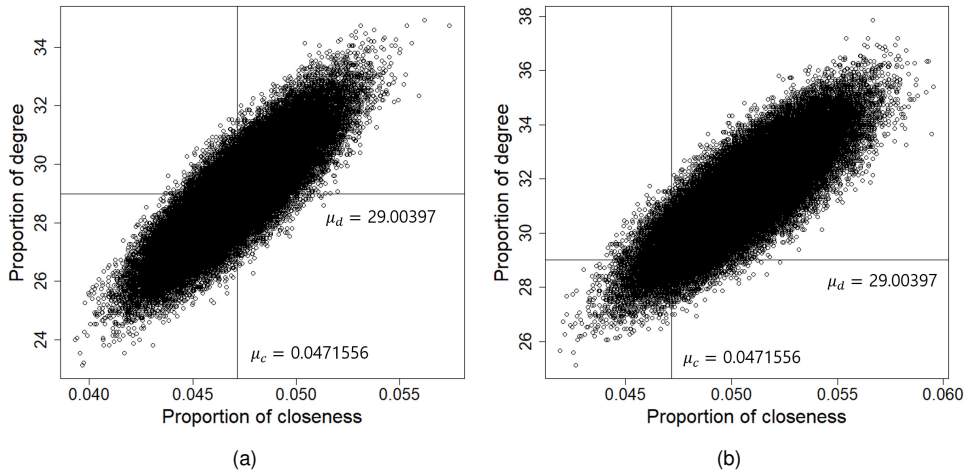


Figure 1: Scatterplots of the in-control and out-of-control probabilities based on degree and closeness centralities when $\lambda = 1$. (a) the in-control probabilities (b) the out-of-control probabilities when $K = 4$ and $\delta = 6$.

significantly, but that of category 3 decreases significantly. It is found that when $K = 2, 3$, and 4 , the probabilities of category 1 and 2 increase, while the probabilities of category 3 and 4 decrease. When $K = 6$, the probability of category 1 increases, but the probabilities of category 2, 3, and 4 decrease. Due to the change in the probabilities belonging to each category, the multinomial CUSUM chart can detect the occurrence of abnormal nodes.

The ARL performance for $\lambda = 1$ and $\lambda = 2$ are shown in Tables 3 and 4, respectively. In each row of the tables, the smallest ARL value is shown in boldface.

In Table 3 where $\lambda = 1$, the in-control mean of degree is $\mu_d = 29.00397$, whereas the in-control mean of closeness is $\mu_c = 0.0471556$. When $K = 2$ and δ is not large ($\delta \leq 3$), the multinomial CUSUM chart performs better than the EWMA charts. When δ is large, the EWMA chart using closeness centrality with $\gamma = 0.05$ performs better. However, even when δ is large, the performance of the EWMA chart using degree centrality is not good compared to the multinomial CUSUM chart. For example, when $\delta = 4$, the ARL for the multinomial CUSUM chart is 38.4, while the ARL for the EWMA chart using degree centrality with $\gamma = 0.05$ is 52.3. The multinomial CUSUM chart performs better for small δ at $K = 3$ and 4 , but the EWMA charts perform better as δ increases. When $K = 6$, the EWMA charts outperform the multinomial CUSUM chart for all the changes. When the abnormal nodes size is large ($K \geq 4$), the overall performance of the EWMA charts using $\gamma = 0.2$ is better.

To determine why the multinomial CUSUM chart performs poorly for large K and δ , we examine the probabilities of four categories for each change, as shown in Table 5 where $\lambda = 1$. As mentioned in Ryan *et al.* (2011), the multinomial CUSUM chart has a weakness in detecting shifts in misspecified directions and magnitudes. As K and δ increase, the direction and magnitude of the probabilities change differently from those specified in Table 1. For example, it can be seen that when $K = 3$ and $\delta = 8$, the probability of category 1 increases significantly compared to the magnitude specified in Table 1, and the probability of category 2 changes in the opposite direction to the direction specified in Table 1. In addition, when $K = 6$ and δ is large, the probability of category 1 is significantly larger than other probabilities. As such, when the probability of one category is representative of the entire network, the ARL performance of the multinomial CUSUM chart may deteriorate.

In Table 4 where $\lambda = 2$, the in-control mean of degree is $\mu_d = 57.99715$, whereas the in-control

Table 3: ARL values for EWMA charts, using degree and closeness centralities separately, and multinomial CUSUM charts when $\lambda = 1$

K	δ	EWMA						Multinomial CUSUM
		$\gamma = 0.05$		$\gamma = 0.1$		$\gamma = 0.2$		
		Degree	Closeness	Degree	Closeness	Degree	Closeness	
2	0.5	100.5	97.0	102.4	98.1	101.2	98.7	90.2
	1	95.7	91.7	98.6	93.4	100.7	92.2	81.6
	2	82.9	67.7	85.8	70.4	88.8	75.8	62.0
	3	66.8	49.9	71.3	54.1	76.5	58.7	47.5
	4	52.3	38.3	55.0	40.8	62.1	46.6	38.4
	5	42.7	31.4	46.0	32.5	50.7	36.8	33.4
	6	34.8	26.0	37.2	26.9	42.2	31.0	30.0
	7	28.8	23.1	30.4	23.8	35.1	26.9	27.4
8	24.5	20.9	25.8	21.4	29.2	23.3	24.4	
3	0.5	89.5	84.3	92.8	87.0	94.8	89.8	73.6
	1	67.2	56.5	69.8	60.2	76.4	64.3	50.2
	2	34.8	26.4	36.8	27.4	42.5	30.9	28.4
	3	21.4	16.2	21.8	16.1	24.9	17.4	19.5
	4	15.1	11.9	14.9	11.4	16.1	12.0	15.0
	5	11.6	9.7	11.1	9.0	11.4	9.1	12.7
	6	9.4	8.3	8.8	7.7	8.8	7.5	11.2
	7	7.9	7.5	7.3	6.9	6.9	6.6	10.3
8	6.9	6.9	6.2	6.2	5.9	5.9	9.6	
4	0.5	66.2	61.9	70.5	64.2	75.8	70.2	55.0
	1	36.9	29.6	37.1	31.2	42.5	35.6	31.6
	2	15.1	12.2	14.9	11.7	16.2	12.1	14.7
	3	9.4	7.8	8.8	7.2	8.8	6.9	10.0
	4	6.8	6.0	6.2	5.4	5.8	4.9	8.1
	5	5.4	5.0	4.8	4.5	4.4	4.0	7.2
	6	4.5	4.4	3.9	3.9	3.5	3.5	6.7
	7	3.8	4.1	3.4	3.6	3.0	3.2	6.4
8	3.4	3.8	2.9	3.3	2.5	2.9	6.3	
6	0.5	26.6	25.3	28.1	26.9	32.1	29.8	31.5
	1	11.6	10.6	11.1	10.0	11.5	10.2	13.6
	2	5.4	4.9	4.8	4.3	4.4	3.8	6.4
	3	3.6	3.4	3.1	3.0	2.7	2.6	5.3
	4	2.7	2.8	2.4	2.4	2.0	2.1	5.1
	5	2.3	2.4	2.0	2.1	1.7	1.8	5.0
	6	2.0	2.2	1.7	1.9	1.4	1.6	5.0
	7	1.8	2.1	1.5	1.8	1.2	1.5	5.0
8	1.6	2.0	1.2	1.7	1.1	1.4	5.0	

mean of closeness is $\mu_c = 0.08057656$. Similar to the results in Table 3, the multinomial CUSUM chart performs better for small values of K and δ , while the EWMA charts perform better for large values of K or δ . When K or δ are very large, the fact that the performance of the multinomial CUSUM chart deteriorates still occurs as shown in Table 3.

From the results in Tables 3 and 4, it is recommended to use the multinomial CUSUM chart when the direction and magnitude of change in category probabilities can be predicted to some extent, or when the magnitude of change is expected to be small. The advantage of using the multinomial CUSUM chart is that we can monitor several centrality measures simultaneously instead of monitoring one centrality measure with a single chart.

6. Conclusions

In recent years, there have been many studies focused on monitoring and detecting changes in SNA and providing information to help prevent conflicts and crises. Accordingly, the networks were observed over time and the observed networks were monitored by applying control charts. Most studies

Table 4: ARL values for EWMA charts, using degree and closeness centralities separately, and multinomial CUSUM charts when $\lambda = 2$

K	δ	EWMA						Multinomial CUSUM
		$\gamma = 0.05$		$\gamma = 0.1$		$\gamma = 0.2$		
		Degree	Closeness	Degree	Closeness	Degree	Closeness	
2	0.5	98.0	99.0	99.5	101.2	100.8	100.5	90.7
	1	91.2	91.1	92.1	93.1	95.2	95.4	71.5
	2	67.9	58.6	71.2	60.9	76.5	67.3	49.6
	3	48.1	37.7	52.7	40.4	58.2	46.7	37.4
	4	36.6	27.9	38.6	29.0	44.0	33.5	30.1
	5	27.6	22.5	29.6	23.1	34.0	25.9	26.1
	6	22.6	19.3	23.7	19.7	27.1	21.8	23.2
	7	18.8	17.5	19.1	17.2	21.1	19.1	20.9
3	0.5	79.1	83.5	83.3	84.6	87.0	90.2	60.8
	1	48.6	45.1	51.8	47.7	58.9	53.4	38.2
	2	22.3	17.8	23.3	17.8	26.7	19.7	19.7
	3	14.0	10.9	13.6	10.4	14.6	10.5	14.1
	4	9.9	8.3	9.4	7.7	9.5	7.5	11.5
	5	7.8	7.0	7.1	6.2	6.9	6.0	10.1
	6	6.4	6.2	5.8	5.6	5.3	5.1	9.3
	7	5.4	5.6	4.8	5.0	4.4	4.6	8.7
4	0.5	48.8	51.0	52.7	53.7	58.9	59.2	41.0
	1	22.3	20.1	23.4	20.1	26.8	22.6	20.6
	2	10.0	8.1	9.4	7.6	9.6	7.4	10.3
	3	6.3	5.4	5.7	4.9	5.4	4.4	7.8
	4	4.7	4.3	4.2	3.8	3.7	3.4	6.8
	5	3.8	3.8	3.3	3.3	2.9	2.9	6.4
	6	3.2	3.4	2.8	2.9	2.4	2.5	6.2
	7	2.8	3.1	2.4	2.7	2.1	2.3	6.1
6	0.5	17.1	17.4	17.5	17.4	19.2	19.0	19.2
	1	7.8	7.0	7.2	6.4	6.8	6.0	8.4
	2	3.8	3.4	3.3	3.0	2.9	2.6	5.3
	3	2.6	2.5	2.3	2.2	1.9	1.9	5.0
	4	2.1	2.1	1.8	1.9	1.5	1.5	5.0
	5	1.8	1.9	1.5	1.7	1.1	1.3	5.0
	6	1.5	1.8	1.1	1.5	1.0	1.2	5.0
	7	1.2	1.7	1.0	1.4	1.0	1.1	5.0
8	1.0	1.6	1.0	1.3	1.0	1.1	5.0	

that applied control charts individually used continuous measures as statistics. However, when several continuous measures are classified into one categorical variable and applied to a control chart, there is an advantage that several measures can be monitored using a single control chart.

In this paper, we proposed a procedure for classifying degree and closeness measures, that represent local and global perspectives, respectively, based on their average values and detecting the occurrence of abnormal nodes using the probabilities change of the classified categories. The performance of the proposed procedure was compared with that of the EWMA charting procedure using degree and centrality measures individually through simulation.

From the simulation results, it was found that the multinomial CUSUM chart showed better performance for small changes but poor performance for large changes. When K and δ were large, the performance of the CUSUM chart deteriorated because the probability of each category changed differently from the direction and magnitude specified in advance. Therefore, the proposed procedure is recommended when the direction and magnitude of change in category probabilities can be predicted or when it is desired to detect small changes in the social network.

Finally, it would be interesting to study procedures that use measures other than degree and close-

Table 5: Probabilities of categories when $\lambda = 1$

K	δ	$P_{1,1}$	$P_{1,2}$	$P_{1,3}$	$P_{1,4}$
2	0.5	0.4141	0.0927	0.4100	0.0833
	1	0.4226	0.0920	0.4028	0.0826
	2	0.4445	0.0943	0.3822	0.0790
	3	0.4688	0.0977	0.3572	0.0764
	4	0.4929	0.0999	0.3350	0.0722
	5	0.5108	0.1008	0.3166	0.0718
	6	0.5280	0.0987	0.3030	0.0703
	7	0.5442	0.0954	0.2884	0.0720
3	8	0.5653	0.0924	0.2699	0.0724
	0.5	0.4315	0.0907	0.3944	0.0834
	1	0.4665	0.0922	0.3617	0.0796
	2	0.5312	0.0974	0.3001	0.0714
	3	0.5920	0.0999	0.2465	0.0616
	4	0.6491	0.0944	0.2012	0.0553
	5	0.6979	0.0861	0.1636	0.0524
	6	0.7431	0.0731	0.1320	0.0518
4	7	0.7773	0.0612	0.1086	0.0528
	8	0.8080	0.0478	0.0888	0.0554
	0.5	0.4622	0.0888	0.3667	0.0823
	1	0.5235	0.0917	0.3089	0.0759
	2	0.6503	0.0904	0.2036	0.0557
	3	0.7493	0.0791	0.1287	0.0429
	4	0.8292	0.0593	0.0772	0.0344
	5	0.8882	0.0375	0.0462	0.0282
6	6	0.9231	0.0214	0.0285	0.0270
	7	0.9476	0.0121	0.0158	0.0245
	8	0.9632	0.0056	0.0088	0.0225
	0.5	0.5488	0.0859	0.2855	0.0798
	1	0.6902	0.0759	0.1745	0.0595
	2	0.8927	0.0393	0.0448	0.0232
	3	0.9724	0.0113	0.0083	0.0080
	4	0.9938	0.0021	0.0013	0.0028
6	5	0.9984	0.0002	0.0002	0.0012
	6	0.9994	0.0000	0.0000	0.0006
	7	0.9998	0.0000	0.0000	0.0002
	8	0.9999	0.0000	0.0000	0.0001

ness centrality measures as criteria for categorization. In the future, we will study the classification procedures using two or more different measures and their performance for various types of change.

References

- Abbasi A, Altmann J, and Hossain L (2011). Identifying the effects of co-authorship networks on the performance of scholars, *Journal of Informetrics*, **5**, 594–607.
- Abbasi A and Hossain L (2013). Hybrid centrality measures for binary and weighted networks, *Complex Networks*, Springer, Berlin, Heidelberg.
- Barrat A, Barthelemy M, Pastor-Satorras R, and Vespignani A (2004). The architecture of complex weighted networks, *Proceedings of the National Academy of Sciences*, **101**, 3747–3752.
- Brandes U (2001). A faster algorithm for betweenness centrality, *The Journal of Mathematical Sociology*, **25**, 163–177.
- Dijkstra EW (1959). A note on two problems in connexion with graphs, *Numerische Mathematik*,

- 269–271.
- Freeman LC (1977). A set of measures of centrality based on betweenness, *Sociometry*, **40**, 35–41.
- Freeman LC (1978). Centrality in social networks conceptual clarification, *Social Networks*, **1**, 215–239.
- 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.
- Knoth S (2021). R software package ‘spc’: Statistical process control– Calculation of ARL and other control chart performance measures.
- Newman MEJ (2001). Scientific collaboration networks. II. Shortest paths, weighted networks, and centrality, *Physical Review E*, **64**, 016132.
- Opsahl T, Agneessens F, and Skvoretz J (2010). Node centrality in weighted networks: generalizing degree and shortest paths, *Social Networks*, **32**, 245–251.
- Perry MB (2020). An EWMA control chart for categorical processes with applications to social network monitoring, *Journal of Quality Technology*, **52**, 182–197.
- Priebe CE, Conroy JM, Marchette DJ, and Park Y (2005). Scan statistics on Enron graphs, *Computational and Mathematical Organization Theory*, **11**, 229–247.
- Reynolds MR Jr and Stoumbos ZG (1999). A CUSUM chart for monitoring a proportion when inspecting continuously, *Journal of Quality Technology*, **31**, 87–108.
- Roberts SW (1959). Control chart tests based on geometric moving averages, *Technometrics*, **1**, 239–250.
- Ryan AG, Wells LJ, and Woodall WH (2011). Methods for monitoring multiple proportions when inspecting continuously, *Journal of Quality Technology*, **43**, 237–248.
- Scott J (1991). *Social Network Analysis: A Handbook*, Sage Publications, Inc.
- Wilson JD, Stevens NT, and Woodall WH (2019). Modeling and detecting change in temporal networks via a dynamic degree corrected stochastic block model, *Quality and Reliability Engineering International*, **35**, 1363–1378.
- 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 (2021). Monitoring dynamic networks: A simulation-based strategy for comparing monitoring methods and a comparative study, *Quality and Reliability Engineering International*, **39**, 1226–1250.

Received March 17, 2022; Revised April 7, 2022; Accepted April 7, 2022