Advances in Industrial Engineering, Winter 2022, 56(1): 73-86
DOI: 10.22059/AIE.2022.339650.1828
RESEARCH PAPER

Enhancing the Performance of Monitoring the DCSBM Using Multivariate Control Charts with Estimated Parameters

Sara Abossedgh [a], Abbas Saghaei [*,a], Amirhossein Amiri [b]

a. *Department of Industrial Engineering, Science and Research Branch, Islamic Azad University, Tehran, Iran.*
b. *Department of Industrial Engineering, Faculty of Engineering, Shahed University, Tehran, Iran.*

Received: 27 February 2022, Revised: 20 March 2022, Accepted: 31 March 2022

Abstract

Many methods are applied to network surveillance for anomaly detection. Some quality control methods have been developed to monitor several quality characteristics simultaneously in different networks. In our study, we use three multivariate process monitoring techniques such as Hotelling's T^2, MEWMA, and MCUSUM to compare to the prior univariate control charts in the Degree-Corrected Stochastic Block Model (DCSBM), a random network model supporting the degree of each node based on Poisson distribution. By estimating parameters in Phase I from many charts, we apply ARL and SDRL metrics for the performance evaluation of multivariate control charts. The advantage of our method is detecting signals faster than previews ones by simulation and this is useful for defining the suitable method in different types of change. Furthermore, the quality of performance in different multivariate methods is displayed in detecting the shifts in the DCSBM. Finally, MCUSUM shows better performance for monitoring local and global changes than other methods.

Keywords:
Change Detection;
DCSBM;
Estimation Effect;
Multivariate Process Monitoring;
Random Graphs

Introduction

Social networks present interactions and communications between individuals or actors. Online social networks such as Facebook, Twitter, and Instagram are developed all around the world. Analysing the behaviour of the networks becomes significant in science.

Many studies have concentrated on controlling temporal networks and their methods had been quite different from static ones. Even the goals of network surveillance of these two approaches are different. In a deterministic network, we try to detect the random traffic flow which is the result of abnormal conditions as we say surveillance on networks [1]. However, many random variables are defined in dynamic networks and we consider network structure in abnormal conditions as we say surveillance of networks [1]. Although studying temporal networks is one of the most important topics in network surveillance, the objectives for studying this field cause many different names, such as temporal graphs, evolving graphs, time-varying graphs, time-aggregated graphs, time-stamped graphs, dynamic networks, dynamic graphs, dynamical graphs, and so on [2].

One of the methods of network surveillance is statistical process monitoring (SPM) [3]. There is a common tool to observe changes and signals in each social network during a

[*] Corresponding author: (A. Saghaei)
Email: a.saghaei@srbiau.ac.ir

particular period of time. With anomaly detection, we find significant changes that can lead to an abnormal situation but all different types of shifts don't necessarily reflect an unusual behaviour [4].

In statistical process monitoring, we consider two phases, Phase I and Phase II. These phases are different in their goals. Phase I is used to find the change over time, evaluate the process stability, and define the situation which is in-control with related parameters [5,6]. But in Phase II, we have to detect anomalies on the foundation of known parameter estimation in Phase I. The performance of this method is assessed through the run length which is defined as the number of chart statistics plotted until the chart signals [7]. Actually calculating the expected value of the run length, the average run length (ARL) and the standard deviation of the run length (SDRL) can be applied to assessing this method. However, these two metrics can show different performances in some conditions [6,8,9]. When the parameters are known, ARL and SDRL are both constant, but when the parameters are estimated and control limits are determined based on estimated parameters through in-control Phase I samples, both of them become random variables. In these cases, the effect of parameter estimation on the performance of the Phase II control chart should be noticed [6,7,10]. By using these kinds of metrics, we can investigate the proper control chart for each type of change. Moreover, the amount of efficiency for each method would be determined and compared with other methods.

For network monitoring, the statistic or a vector of statistics has to be specified at first [3], because they present the perspective of a network. In some studies, central measures such as betweenness, closeness, and eigenvector are used to monitor a network [11]. Instead, modelling the network by a set of probability distributions of parameters can help in selecting a special statistic [3]. Many methods have been applied to monitor network measures. [12] recommended these methods in four categories: Control chart and hypothesis testing methods, Bayesian methods, Scan methods, and Time-series models.

We apply our surveillance strategy by the first approach. The first method is using control charts which include two Phases (Phase I and II). Different algorithms of control charts are designed. For univariate analysis, [13] highlighted the cumulative sum (CUSUM) as the primary method for analysing longitudinal networks. Also, Shewhart and exponentially weighted moving average (EWMA) control charts were used for detecting the change [13,14]. Generally, Shewhart is applied for detecting sudden large changes, and EWMA has the ability to find small and medium-sized changes [3]. Univariate control charts have been developed in some research [15,16,17].

When several related quality characteristics are considered to monitor, alternative methods in multivariate control charts can be more useful such as Hotelling's T^2, multivariate exponentially weighted moving average (MEWMA), and multivariate cumulative sum (MCUSUM) [18]. The performances of MCUSUM and MEWMA control charts are preferable to CUSUM and EWMA; furthermore, the efficiency of MEWMA is higher than MCUSUM in a wide range of changes [11]. Also, the most common multivariate control chart is Hotelling's T^2 control chart will be described further. Although the multivariate control charts use fewer points to detect the change compared to univariate control charts, the significant defect is the inability of representing the shifted variables [19].

Stochastic block models (SBM) are an increasingly popular category of models concerning community structure. The analysis of these networks has been considered by researchers today. Most studies about SBM are currently focused on different types of community detection or community recovery [20-23]. Vaca-Ramírez and Peixoto [24] presented a systematic analysis of this network and some research tried to monitor it in different ways.

Among different fields of studies, monitoring the Degree Corrected Stochastic Block Model (DCSBM) is an area of scientists' interest. Zwetsloot et al. [25] proposed a multivariate surveillance plan to monitor node propensity in the DCSBM. Some researchers [3,26] performed analysis and used surveillance methods for detecting local and global structural changes. Some articles have been trying to make improvements to the DCSBM [27-30]. Most of the studies caused some community detection research [31-40] that followed one of these approaches: model-based methods or spectral methods. Using model-based methods involves researchers to computational complexity and spectral methods lead to weakness of validation on synthetic and real data [31]. Anyway, improving knowledge in this area continues to defeat the limitations of different methods.

Much of the interest in social network analysis is based on monitoring the parameters of networks. Hosseini and Noorossana [41] applied EWMA and CUSUM control charts for degree measures to detect anomalies in a weighted undirected social network. Noorossana et al. [42] summarized social network analysis papers using EWMA, CUSUM, Shewhart, and other methods for a better understanding of dynamic changes. Mazrae-Farahani and Kazemzadeh [43] used Hotelling's T^2 and likelihood ratio test (LRT) statistics to monitor the network in Phase I. Moreover, applying Hotelling's T^2 and MEWMA is common in Phase II for network monitoring [44]. Also, MEWMA and MCUSUM control charts were employed for monitoring the average degree, average betweenness, and average closeness in another study [45].

Other approaches such as the likelihood-ratio test (LRT) and log-linear model are existed to detect any change in the network [46-49] and Zou and Li [50] proposed a Singular Value Decomposition (SVD)-based method to monitoring the dynamic network.

All current studies don't pay attention to the common methods of multivariate control charts for monitoring the DCSBM and comparing it with each other. Using effective methods can greatly improve the detection performance of monitoring.

The Degree Corrected Stochastic Block Model (DCSBM)

Analysing the random graph model based on Erdős and Rényi (1960, 1961) research [51] has been developed. As we mentioned, one of the famous random models in network analysis is the SBM that comes from the integration of stochastic models presented by Holland (1981) and block models introduced by White (1976) [52]. The lack of the SBM is all nodes in the same community are stochastically equivalent which is far from the nature of the usual network. So, the Degree Corrected Stochastic Block Model (DCSBM) originates from the stochastic block model. It presents two characteristics of a network: degree heterogeneity and community structure. In this section, we briefly define the DCSBM based on previous research [3,53].

The probability distribution of the DCSBM is □ $(\cdot) = \mathbb{P}(\cdot\,|\theta, \pi, P)$ and $G_t = ([n], A_t)$ shows an undirected random network with $[n]$ actors (nodes), $[v]$ links (edges), and $A_t = \{a_{u,v}(t): u, v \in [n]\}$ is edge weights in time *t*. Also, A is the symmetric adjacency matrix of G. Each node of a network G is assigned to a community (block). The number of disjoint communities is k $(k \geq 1)$. In these networks, we have to define three parameters: (1) $\boldsymbol{\theta} = (\theta_1, \dots, \theta_n)$, shows the tendency of each node to connect with others, (2) $\boldsymbol{\pi} = (\pi_1, \dots, \pi_k)$ with constraints of $\pi_r > 0$, $\sum_{r\in[k]} \pi_r = 1$, defines the probability of a node belonging to a special community, and (3) $k \times k$ symmetric connectivity matrix (**P**) that each array $(P_{r,s})$ specifies the propensity of connection between nodes in communities r and s. To generate this network, we have to assign labels to all nodes ($\mathbf{c} = (c_1, \dots, c_n)$ are labels of n nodes), randomly by multinomial distribution. Then we calculate edges weights from the mean Poisson distribution as:

$$\mathbb{E}[A_{u,v}|c, \theta, P] = \theta_u \theta_v P_{c_u, c_v} , \tag{1}$$

Meanwhile, $m_{r,s}$ is the total weight of edges between community r and s if $r \neq s$ and twice the weight of edges if $r = s$. The mathematical form of the DCSBM is[30]:

$$P(A, G|\pi, \theta, P) = \prod_u \pi_{g_u} \prod_{u<v} \frac{(\theta_u \theta_v P_{g_u g_v})^{a_{uv}}}{a_{uv}!} e^{-\theta_u \theta_v P_{g_u g_v}} = \prod_r \pi_r^{n_r} \prod_u \theta_u^{d_u} \prod_{rs} P_{rs}{}^{m_{rs}/2} e^{-(1/2) n_r n_s P_{rs}} \prod_{u<v} \frac{1}{a_{uv}!} , \quad (2)$$

where n_r is the number of nodes in community r. This formation needs a special constraint that shows the sum of total θ_u is equal to the total number of nodes in each block:

$$\sum_{u:\, g_u = r} \theta_u = n_r \quad ; \quad r = 1, \ldots, k \ , \quad (3)$$

The block assignment is $G = \{G_u\}$. There is defined the average degree of nodes in block r:

$$d_r = \frac{1}{n_r} \sum_{u:\, g_u = r} d_u \ , \quad (4)$$

Using the maximum likelihood estimator (MLE) for each parameter has a closed-form solution [30]. We have:

$$\hat{\theta}_u = \frac{d_u}{d_{gu}} \ , \quad \hat{\pi}_r = \frac{n_r}{n} \ , \quad \hat{P}_{rs} = \frac{m_{rs}}{n_r n_s} \ . \quad (5)$$

According to a piece of research [3], for monitoring the DCSBM, the number of communities, k, is predefined and fixed, and the community labels are estimated at first, then $\hat{\mathbf{P}}$ and $\hat{\boldsymbol{\theta}}$ have been monitored. Because of having many $\hat{\theta}_u$ for monitoring, the sample standard deviation of $\hat{\theta}_u$ in each community for detecting the anomaly was chosen. Then in the monitoring plan, $\binom{k}{2}$ statistics (the propensity of connection between nodes in different communities) and k statistics (the propensity of connection between nodes in the same communities) both for $\hat{\mathbf{P}}$,and $sd(\hat{\theta}_u)$ in each community for $r = 1, \ldots, k$ are defined and δ is pooled estimate of $sd(\hat{\theta}_u)$ in all communities. For assigning θ_u to each node, a uniform distribution with δ_r (a multiplicative constant) parameters was selected:

$$\theta_u^0 \overset{iid}{\sim} U(1 - \delta_{c_u}^0, 1 + \delta_{c_u}^0), \quad (6)$$

Some initial values set for $\mathbf{P}^0$ and δ_r^0 are:

$$\mathbf{P}^0 = \begin{pmatrix} 0.2 & 0.1 \\ 0.1 & 0.2 \end{pmatrix}, \delta_1^0 = \delta_2^0 = 0.5 \ . \quad (7)$$

So, generating a dynamic network with $n = 100$ nodes, $k = 2$ equally sized communities with a special t^* and m (number of the network in Phase I) is performed [3].

This kind of monitoring was organized to detect global and local changes in community structure for different cases of simulation. All assumptions of this paper are as mentioned by Wilson et al. [3], only the approach of controlling the charts is different and multivariate control charts are applied. A critical point of using this method is a high-performance of applying a single or a few multivariate charts instead of many univariate charts [54]. Also, detecting the change in different conditions and choosing the best method in each state cause to improve our usage of multivariate control charts and fill our research gap. We briefly present three multivariate control charts and apply them to monitor the DCSBM.

Monitoring methods: Multivariate control charts

Hotelling's T^2

Hotelling's T^2 chart is the analogue of the Shewhart $\bar{x}$ chart and using for monitoring the mean vector of the process. It is called a multivariate Shewhart control chart. Although this statistic is slow to detect small process shifts, it is the most applied in multivariate process control for detecting a general shift in the process mean vector for an individual multivariate observation [55].

When $\mathbf{x}_1, \mathbf{x}_2, \mathbf{x}_3, \ldots$ are sample mean vector of p quality characteristics and $\mathbf{x}_i$ is ith independent sample mean vector, in most conditions $\boldsymbol{\mu}$ and $\boldsymbol{\Sigma}$ are unknown. Then, as we replace $\boldsymbol{\mu}$ with $\bar{\boldsymbol{x}}$ and $\boldsymbol{\Sigma}$ with **S**, the Hotelling's T^2 statistic for $n = 1$ becomes:

$$T_i^2 = (\mathbf{x}_i - \bar{\mathbf{x}})' \boldsymbol{S}^{-1} (\mathbf{x}_i - \bar{\mathbf{x}}), \tag{8}$$

With $\bar{\mathbf{x}}$ and **S** being the vector of in-control value of means and covariance matrix, respectively and m is the number of samples in Phase I .

Multivariate exponentially weighted moving average (MEWMA)

MEWMA was developed by Lowry et al. (1992). It is the extension of EWMA. The statistic is:

$$T^2 = \mathbf{z}_i' \boldsymbol{\Sigma}_{\mathbf{z}_i}^{-1} \mathbf{z}_i \,, \tag{9}$$

where:

$$\mathbf{z}_i = \lambda \mathbf{x}_i + (1 - \lambda) \mathbf{z}_{i-1} \,. \tag{10}$$

$\mathbf{z}_i$ and $\mathbf{x}_i$ are the ith EWMA vector and the ith observation vector, respectively. If $Z_0 = 0$, $\boldsymbol{\lambda}$ is diagonal $p \times p$ matrix of the smoothing constant with $0 < \lambda_i \leq 1$. Also, two methods are defined for computing the $\boldsymbol{\Sigma}_{\mathrm{z}}$ [56].

Multivariate cumulative sum (MCUSUM)

It is the extension of CUSUM. Some alternatives are proposed for MCUSUM chart. One of the statistics is [56]:

$$T_i^2 = [\mathbf{s}_i' \left(\frac{\Sigma}{n}\right)' \mathbf{s}_i]^{1/2} \,, \tag{11}$$

s_i is defined as:

$$\mathrm{s}_i = \begin{cases} 0 & if \quad C_i \leq k \\ (\mathbf{s}_{i-1} + \bar{\mathbf{x}}_i - \boldsymbol{\mu}_0)\left(1 - \frac{k}{C_i}\right) & if \quad C_i > k \end{cases} , \tag{12}$$

where $\mathbf{s}_i = 0$, $k > 0$, C_i is:

$$C_i = [(\mathbf{s}_{i-1} + \bar{\mathbf{x}}_i - \boldsymbol{\mu}_0)' \left(\frac{\Sigma}{n}\right)' (\mathbf{s}_{i-1} + \bar{\mathbf{x}}_i - \boldsymbol{\mu}_0)]^{1/2} \,, \tag{13}$$

and

$$k = 0.5\ \frac{(\mu_1-\mu_0)'(\frac{\Sigma_0}{n})^{-1}(\mu_1-\mu_0)}{[(\mu_1-\mu_0)'(\frac{\Sigma_0}{n})^{-1}(\mu_1-\mu_0)]^{1/2}}. \quad (14)$$

Also, Santos-Fernández [56] presented other forms of MCUSUM statistics.

Performance estimation of the methods

For applying multivariate control charts, choosing the number of variables and type I error are necessary. According to [3], $p = 4$ quality characteristics of our process include $sd(\hat{\theta})$ and the elements of **P** ($k * k$ symmetric matrix for k = 2) such as $\hat{p}_{1,1}$, $\hat{p}_{1.2}$, and $\hat{p}_{2,2}$ [3]. So, these four variables have to be controlled simultaneously in a unique control chart. On the other hand, type I error (α) for all Shewhart control charts was about 0.0027 ($\frac{1}{\mathrm{ARL}_0}$) for in-control conditions. It means around this point, all the charts show the first out of control point. Based on average ARL_0 for each parameter in the study by Wilson et al. [3], we use $\alpha_{overall} = 1 - (1-\alpha)^p$, probability of type I error for the joint control procedure, then we have $\alpha_{overall} = 0.013$ ($\mathrm{ARL}_0 = 76.92$). If we want to run control charts, we have to apply the same in-control point for getting ARL_0. By these assumptions, for discussing the effect of estimated parameters on the performance of T^2, MEWMA, and MCUSUM, we clarify and apply two different processes for monitoring the DCSBM: monitoring process with known and estimated parameters, and then describe our proposed method to monitor the DCSBM.

Monitoring process with known parameters

This monitoring process includes the following steps:

(1) The considered parameters ($p_{1,1}$, $p_{1,2}, p_{2,2}$) are known by $\mathbf{P}^0$, but $sd(\theta)$ and the covariance matrices are obtained by 10000 replications of Phase I networks. Then we get a unique mean vector and a covariance matrix for known parameters.

(2) Generate Phase II networks and include the mean vector and covariance matrix of known parameters for calculating the statistics and UCL to obtain run length (RL).

(3) Repeat Step 2 for 10000 times and record RL to achieve ARL.

So, running large enough simulations of the process causes results of monitoring with known parameters that express $m = \infty$ in Table 1. By choosing $\alpha_{overall} = 0.013$ for T^2, we can get the nearest ARL_0 (75.74). MEWMA and MCUSUM with UCLs of 13.86 and 9.3, respectively, are obtained. Defining ARLs for known parameters can be a prototype for choosing the appropriate m of estimated parameters in the next approach.

Table 1. ARLs of T^2, MEWMA, and MCUSUM based on estimated parameters comparing different m and known parameters ($m = \infty$) through simulation studies

		T^2					**MEWMA**					**MCUSUM**				
		m					m					m				
Sim.	**Change**	200	500	1000	1500	∞	200	500	1000	1500	∞	200	500	1000	1500	∞
0	None	72.89	72.44	75.5	75.6	75.74	73.61	72.55	75.54	75.49	75.49	72.62	73.23	75.39	75.73	73.35
1 $P^*_{1,1}=P^0_{1,1}+\varepsilon$	$\varepsilon = 0.01$	44.94	44.84	37.46	37.82	32.84	58.74	58.86	38.46	38.21	32.38	45.1	45.27	24.95	25.11	32.44
	$\varepsilon = 0.05$	10.6	10.45	8.47	8.56	7.06	24.03	24.05	8.45	8.43	6.47	10.48	10.46	5.98	6.08	9.013
	$\varepsilon = 0.1$	3.33	3.57	2.52	2.49	2.01	17.13	17.01	2.33	2.31	2.33	3.42	3.46	2.49	2.41	2.99
2 $P^*_{i,j}=P^0_{i,j}+\varepsilon$ $(i = 1, 2, j = 1, 2)$	$\varepsilon = 0.01$	25.43	24.63	23.75	23.18	20.54	38.37	38.54	23.96	23.99	21.07	24.84	24.96	17.46	17.21	21.06
	$\varepsilon = 0.05$	2.07	2.06	1.49	1.52	1.49	15.48	15.52	1.47	1.39	1.51	2.04	2.06	1.5	1.53	2.02
	$\varepsilon = 0.1$	3.12	3.1	1.48	1.53	1.48	16.55	16.61	1.51	1.37	1.47	3.02	3.16	1.5	1.46	2.46
3 $\delta^*_1 = \delta^0_1+\tau$	$\tau = 0.05$	52.57	52.53	44.52	44.62	39.92	66.22	65.65	40.57	40.44	29.96	52.4	52.47	29.39	29.69	49.07
	$\tau = 0.1$	42.53	42.84	40.96	40.91	32.56	56.4	56.44	42.5	43	35.12	42.8	42.08	33.93	34.47	37.3
	$\tau = 0.25$	13.27	12.72	11.51	11.44	8.96	26.67	26.11	7.53	7.7	5.99	13.12	13.36	4.46	4.21	10.46
4 $\delta^*_i = \delta^0_i+\tau$ $(i = 1, 2)$	$\tau = 0.05$	37.31	37.18	35	35.27	29.99	50.72	51.52	33.57	33.19	29.89	37.95	37.39	23.52	23.77	29.93
	$\tau = 0.1$	26.24	25.89	23.00	22.98	22.01	39.85	39.49	20.98	20.88	17.5	25.84	25.55	13.35	13.26	23.45
	$\tau = 0.25$	5.5	5.49	4.48	4.43	3.94	18.84	19.02	3.39	3.43	2.47	5.47	5.51	2.58	2.48	5
5 Merge comm.	$n = 50$	2.47	2.54	1.49	1.53	1.47	16.09	16.22	3.52	3.52	2.99	2.38	2.58	2.52	2.63	1.97
	$n = 100$	2.66	2.38	1.47	1.46	1.48	15.91	16.07	1.99	1.98	1.49	2.5	2.67	3.56	3.69	2.03
	$n = 500$	2.42	2.5	2	2.02	1.94	16.24	16	1.96	2.04	2.02	2.46	2.58	4.01	3.97	2.5
6 Split comm.	$n = 50$	19.33	20.01	17.55	17.31	15.04	33.84	32.44	17.47	17.28	14.1	19.39	19.54	9.58	9.17	16.28
	$n = 100$	21.58	21.66	20.4	20.07	7.98	35.83	36.08	21.01	21.07	17.54	22.06	22.33	18.52	18.47	18.53
	$n = 500$	21.62	21.55	19.51	19.51	5.98	35.37	35.1	23.57	23.65	19.5	21.44	21.86	14.04	13.89	16.5

Monitoring process with estimated parameters

Steps of this approach are stated below:

(1) Generate m networks in Phase I to estimate parameters (mean vector and covariance matrix).

(2) Generate Phase II networks and include the mean and covariance matrix of Phase I for calculating the statistics and UCL to obtain RL.
(3) Repeat Steps 1 and 2 for 10000 times and record RL to achieve ARL.
(4) If the values of ARLs are equal to or smaller than ARLs of known parameters, report *m* as the sufficient number of Phase I networks. Otherwise, go to Step 1 and change *m*.

In order to find the minimum value for *m*, Table 1 displays ARLs for simulations by 3 multivariate methods (T^2, MEWMA, and MCUSUM) based on estimated parameters using *m* = 200, 500, 1000, and 1500 samples. By considering the trend of ARLs in Table 1, we conclude that ARLs in all 3 methods are coming closer to known parameters ARLs ($m = \infty$) with increasing *m* until 1000. After *m* = 1000, so many small changes in ARLs occur. So, we expect that *m* = 1000 leads to the best performance of multivariate control charts.

Performance of proposed monitoring methods with estimated parameters for *m* = 1000

Table 2 illustrates simulations separately for each parameter by Shewhart, mentioned in [3].

Table 2. ARLs and SDRLs of T^2, MEWMA, and MCUSUM based on estimated parameters by comparing to Shewhart when *m* = 1000 through simulation studies

Sim.		Change	Shewhart ARL	T^2 ARL	T^2 SDRL	MEWMA ARL	MEWMA SDRL	MCUSUM ARL	MCUSUM SDRL
0		None	76.92	75.5	72.65	75.54	72.43	75.39	72.42
1	$P^*_{1,1} = P^0_{1,1} + \varepsilon$	$\varepsilon = 0.01$	59.84	37.46	32.40	38.46	34.30	24.95	21.75
		$\varepsilon = 0.05$	10.12	8.47	6.50	8.45	5.46	5.98	3.98
		$\varepsilon = 0.1$	2.04	2.52	1.51	2.33	0.19	2.49	0.45
2	$P^*_{i,j} = P^0_{i,j} + \varepsilon$ $(i = 1, 2, j = 1, 2)$	$\varepsilon = 0.01$	28.3	23.75	21.02	23.96	23.02	17.46	15.47
		$\varepsilon = 0.05$	1.68	1.49	0.55	1.47	1.40	1.5	0.50
		$\varepsilon = 0.1$	1.01	1.48	0.14	1.51	0.28	1.5	1.29
3	$\delta^*_1 = \delta^0_1 + \tau$	$\tau = 0.05$	57.86	44.52	41.47	40.57	37.42	29.39	27.45
		$\tau = 0.1$	48.97	40.96	40	42.5	39.52	33.93	31.96
		$\tau = 0.25$	16.2	11.51	8.52	7.53	6.48	4.46	3.40
4	$\delta^*_i = \delta^0_i + \tau$ $(i = 1, 2)$	$\tau = 0.05$	40.4	35	32.10	33.57	30.36	23.52	20.59
		$\tau = 0.1$	25.99	23.00	18	20.98	19.01	13.35	11.52
		$\tau = 0.25$	4.23	4.48	1.46	3.39	2.41	2.58	1.96
5	Merge comm.	$n = 50$	1.74	1.49	0.5	3.52	1.50	2.52	0.95
		$n = 100$	1.67	1.47	0.49	1.99	0.69	3.56	1.47
		$n = 500$	1.59	2	0.69	1.96	0.72	4.01	1.76
6	Split comm.	$n = 50$	29.77	17.55	15.55	17.47	16.63	9.58	7.56
		$n = 100$	24	20.4	15.93	21.01	19.03	18.52	16.49
		$n = 500$	24.9	19.51	15.51	23.57	20.54	14.04	11.02

Our research finds the first out-of-control signal, reports ARLs for this process, and then compares this univariate approach to the multivariate mentioned methods. Figs. 1-2 depict Table 2 for each case of change separately.

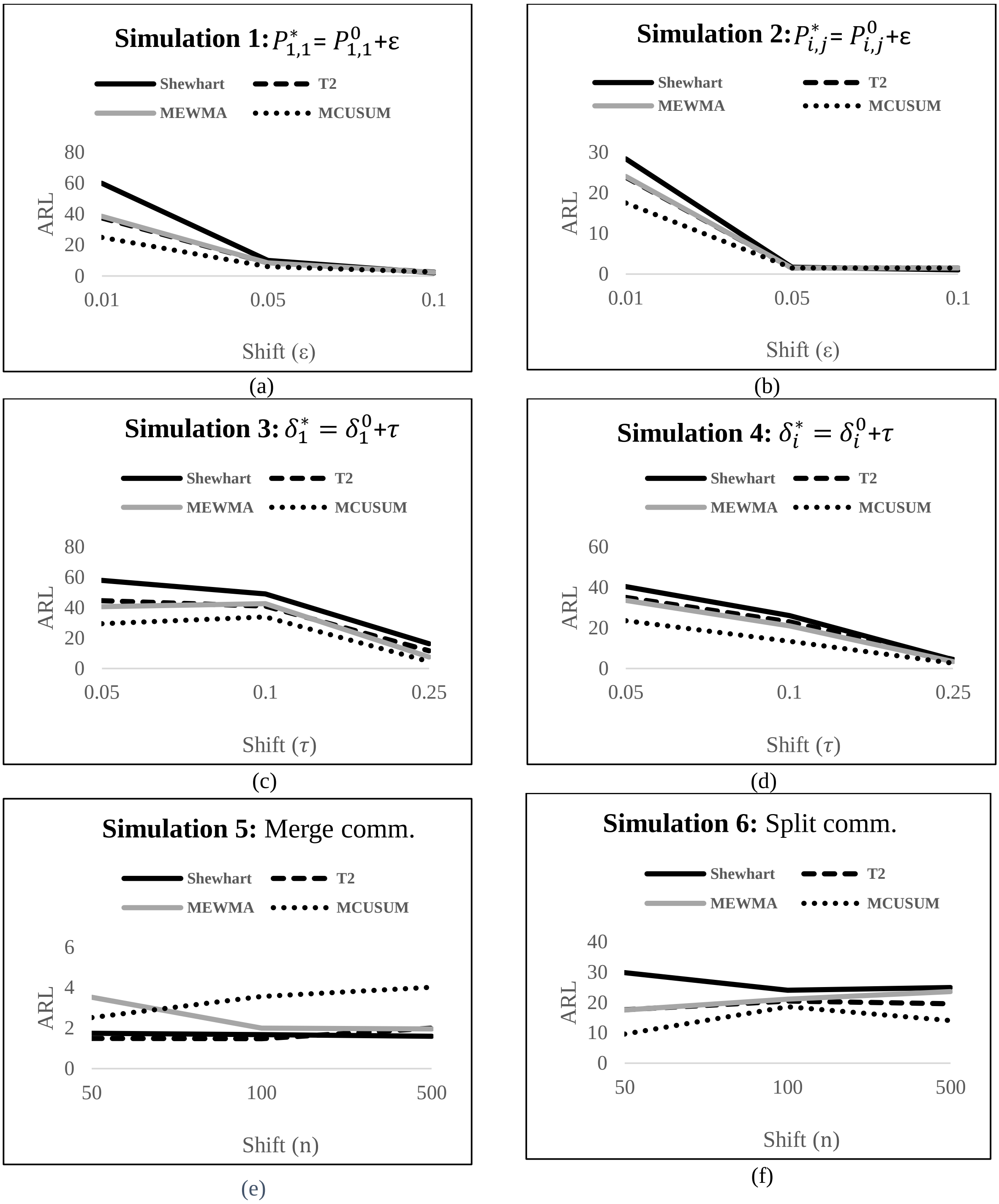

Fig. 1. Comparison of ARLs among Shewhart, T^2 , MEWMA, and MCUSUM

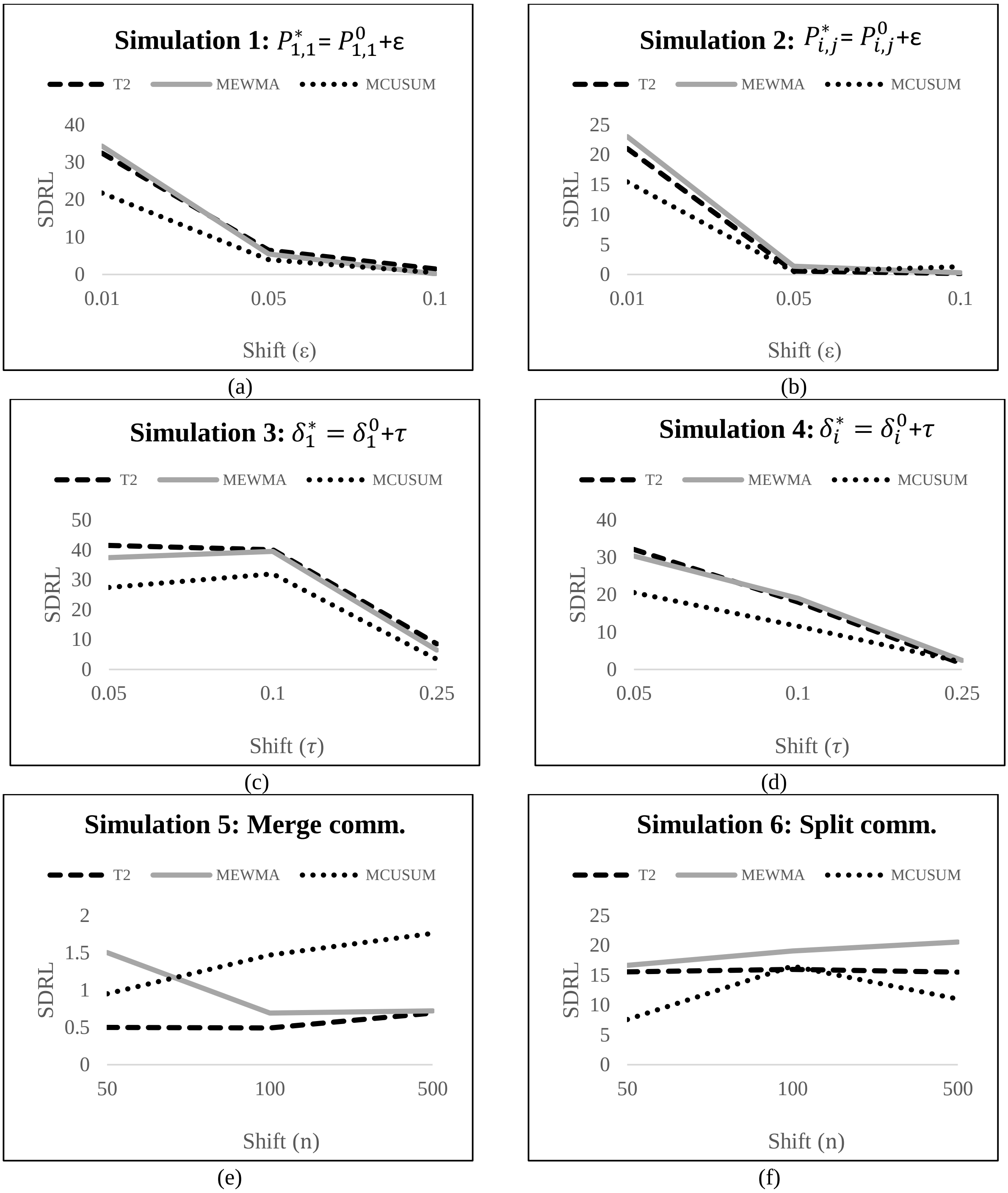

Fig. 2. Comparison of SDRLs among T^2 , MEWMA, and MCUSUM

We extract some results to use a suitable method and investigate them in each case of simulation as stated below:

- Simulation 1: In this state, varying $\hat{p}_{1,1}$ causes local changes. By increasing shifts in this condition, all four methods improve and detect change more quickly than small changes. However, the function of T^2 and MEWMA are very similar. MCUSUM shows better performance in ARLs but with increasing the shifts, MEWMA represents good results in SDRLs.
- Simulation 2: There are global changes by shifting $\hat{p}_{1,2}$. ARLs for T^2 and MEWMA are alike here as well. SDRLs of these two methods are converging to a point with larger shifts, but we can't ignore the relative large SDRL of MCUSUM in the last shift.

- Simulation 3: local variability (δ_1^0) occurs, MCUSUM displays better performance in all shifts. Also in medium shifts, different behaviour of MEWMA and MCUSUM from a normal trend is clear by their ARLS and SDRLs.
- Simulation 4: Applying global changes among all nodes in the network by shifting δ_1^0 and δ_2^0, MCUSUM gives the good performance of ARLs, also T^2 has suitable performance in large shifts by SDRL.
- Simulation 5: By merging communities, community structure changes and $\hat{p}_{1,2}$ should be strongly affected. In this condition, changes are detected so quickly and the values are very close to each other. Shewhart and T^2 both are good in ARLs and T^2 indicates preferable results by SDRLs, too. MCUSUM has not performed well at all in this mode.
- Simulation 6: Same as simulation 5, change in community structure happens. MCUSUM is good at ARLs. Also, in a medium shift, T^2 overcomes other methods by SDRL.

Generally, we find out the results of multivariate control charts (T^2, MEWMA, and MCUSUM) in monitoring the DCSBM by estimated parameters are better than Shewhart. However, as can be expected, Shewhart is useful to detect for some large shifts. But among all methods, MCUSUM detects signals as soon as others in most cases (except simulations 1 and 2 in large shifts and 5), especially in small shifts. In simulation 5, it seems the shift impacts on $\hat{p}_{1,2}$ more than other variables, the change is merging communities, then T^2 shows better performance and Shewhart is in the next stage. It should be noted when we try to control $\hat{p}_{1,2}$, MCUSUM doesn't reveal better performance, this is especially true in medium and large changes. On the other hand, T^2 and MEWMA have similar functions in most ARLs. But according to comparing SDRLs of methods, T^2 obtains proper SDRLs in large shifts in simulations 2, 4, and all shifts in 5. In the last two simulations (5 and 6), we can't observe a determined trend in ARLs and SDRLs of all multivariate methods. Also, using T^2 can be valuable when large global changes arise in the report of SDRLs. Overall, ranking the performance of multivariate approaches with estimated parameters is MCUSUM, T^2, and MEWMA, respectively.

Managerial insights

Multivariate control charts are widely used in the service process. Our results help to reduce the time of process and labour costs, and process defects are quickly detected. These methods can monitor the different parameters of the network with fewer charts, simultaneously. Actually, the abnormal behaviours identified in each parameter can assist a manufacturer to take corrective actions as soon as possible. Even in some other areas such as the health care system, finding abnormal shifts to the right time can lead to prevent possible errors or hazardous events for patient services offered by treatment staff. In sociological or psychological studies, we can detect an unusual case of human behaviour. However, discovering these signals is useful for predicting the future of a social, economic, political, industrial, or some other types of process and finding the source and reason of the signals to avoid the consequence of that.

Conclusions and suggestions for future research

The focus of this study is to apply three famous multivariate approaches (Hotelling's T^2, MEWMA, and MCUSUM) to detect the local and global changes in the dynamic DCSBM. Since we use parameter estimation to evaluate the control chart performance, the minimum number of Phase I samples, ARL, and SDRL are defined. Based on simulation results and ARL metric, MCUSUM performs better than other methods, both univariate and multivariate control charts except in the change of propensity of connection between nodes in merging communities. It is specified that MCUSUM meets the objectives of quality control charts to detect signals quickly in a random dynamic network by estimated parameters more than other methods. In order to decrease the effect of parameter estimation, other methods for improving the monitoring process in Phase I and Phase II can be performed for future research. Although our methods protect all parameters of the DCSBM, investigating the impact of adding node attributes is recommended for monitoring in the future.

References

[1] Sengupta S, Woodall WH. Discussion of "Statistical methods for network surveillance". Appl Stochastic Models Bus Ind. 2018;34 (4):446-448. doi:10.1002/asmb.2354.

[2] Holme P, Saramäki J. Temporal networks. Phys Rep. 2012;519(3):97-125. doi:10.1016/j.physrep.2012.03.001.

[3] Wilson JD, Stevens NT, Woodall WH. Modeling and detecting change in temporal networks via the degree corrected stochastic block model. Qual Reliab Eng Int. 2019;35(5):1363-1378. doi:10.1002/qre.2520.

[4] Savage D, Zhang X, Yu X, et al. Anomaly detection in online social networks. Soc Networks. 2014;39:62-70. doi:10.1016/j.socnet.2014.05.002.

[5] Woodall WH, Spitzner DJ, Montgomery DC, et al. Using control charts to monitor process and product quality profiles. J Qual Technol. 2004;36(3):309-320. doi:10.1080/00224065.2004.11980276.

[6] Yazdi AA, Hamadani AZ, Amiri A. Phase II monitoring of multivariate simple linear profiles with estimated parameters. J Ind Eng Int. 2019;15(4):557-570. doi:10.1007/s40092-019-0305-y.

[7] Saleh NA, Mahmoud MA, Keefe MJ, et al. The difficulty in designing Shewhart X and X control charts with estimated parameters. J Qual Technol. 2015;47(2):127-138. doi:10.1080/00224065.2015.11918120.

[8] Mahmoud MA, Maravelakis PE. The performance of multivariate CUSUM control charts with estimated parameters. J Stat Comput Simul. 2013;83(4):721-738. doi: 10.1080/00949655.2011.633910.

[9] Javaheri A, Houshmand AA. Average run length comparison of multivariate control charts. J Stat Comput Simul. 2001;69(2):125-140. doi: 10.1080/00949650108812086.

[10] Maleki MR, Castagliola P, Amiri A, et al. The effect of parameter estimation on phase II monitoring of poisson regression profiles. Commun Stat- Simul Comput. 2019;48(7):1964-1978. doi:10.1080/03610918.2018.1429619.

[11] Najafi H, Saghaei A. Statistical monitoring for change detection of interactions between nodes in networks: With a case study in financial interactions network. Commun Stat-Theory Methods. 2020. doi:10.1080/03610926.2020.1725830.

[12] Woodall WH, Zhao MJ, Paynabar K, et al. An overview and perspective on social network monitoring. IISE Trans. 2017;49(3):354-365. doi:10.1080/0740817X.2016.1213468.

[13] McCulloh I. Detecting changes in a dynamic social network. Institute for Software Research School of Computer Science Carnegie Mellon University; 2009.

[14] McCulloh I, Carley KM. Detecting change in longitudinal social networks. Military Academy West Point NY Network Science Center (NSC); 2011. doi:10.21307/joss-2019-031.

[15] Perry MB. An EWMA control chart for categorical processes with applications to social network monitoring. J Qual Technol. 2019. doi:10.1080/00224065.2019.1571343.

[16] Taboran R, Sukparungsee S, Areepong Y. Mixed Moving Average-Exponentially Weighted Moving Average Control Charts for Monitoring of Parameter Change. In: Proceeding of the International MultiConference of Engineers and Computer Scientists. 2019;13-15.

[17] Aly AA, Saleh NA, Mahmoud MA. An adaptive EWMA control chart for monitoring zero-inflated Poisson processes. Commun Stat Simul Comput. 2019. doi:10.1080/03610918.2019.1676437.

[18] Aparisi F, Haro CL. A comparison of T 2 control charts with variable sampling schemes as opposed to MEWMA chart. Int J Prod Res. 2003;41(10):2169-2182. doi:10.1080/0020754031000138655.

[19] Aparisi F, Avendaño G, Sanz J. Techniques to interpret T^2 control chart signals. IIE Trans. 2006;38(8):647-657. doi:10.1080/07408170600692200.

[20] Mukherjee CS, Peng P, Zhang J. Recovering Unbalanced Communities in the Stochastic Block Model With Application to Clustering with a Faulty Oracle. 2022. arXiv preprint arXiv:2202.08522.

[21] Alaluusua K, Leskelä L. Consistent Bayesian community recovery in multilayer networks. 2022. arXiv preprint arXiv:2202.05823.

[22] Seif M, Nguyen D, Vullikanti A, Tandon R. Differentially Private Community Detection for Stochastic Block Models. 2022. arXiv preprint arXiv:2202.00636.

[23] Qing H, Wang J. Impact of regularization on spectral clustering under the mixed membership stochastic block model. 2021. arXiv preprint arXiv:2107.14705.

[24] Vaca-Ramírez F, Peixoto TP. Systematic assessment of the quality of fit of the stochastic block model for empirical networks. 2022. arXiv preprint arXiv:2201.01658.

[25] Yu L, Woodall WH, Tsui KL. Detecting node propensity changes in the dynamic degree corrected stochastic block model. Soc Networks. 2018;54:209-227. doi:10.1016/j.socnet.2018.03.004.

[26] Yu L, Zwetsloot IM, Stevens NT, Wilson JD, Tsui KL. Monitoring dynamic networks: A simulation- based strategy for comparing monitoring methods and a comparative study. Quality and Reliability Engineering International. 2021.

[27] Herlau T, Schmidt MN, Mørup M. Infinite-degree-corrected stochastic block model. Phys Rev E. 2014;90 (3):032819. doi:10.1103/PhysRevE.90.032819.

[28] Zhang X, Wang X, Zhao C, et al. Degree-corrected stochastic block models and reliability in networks. Physica A. 2014;393:553-559. doi:10.1016/j.physa.2013.08.061.

[29] Gulikers L, Lelarge M, Massoulié L. An impossibility result for reconstruction in the degree-corrected stochastic block model. Ann Appl Probab. 2018;28(5):3002-3027. doi: 10.1214/18-AAP1381.

[30] Yan X, Shalizi C, Jensen JE, et al. Model selection for degree-corrected block models. J Stat Mech.: Theory Exp. 2014;(5):P05007. doi:10.1088/1742-5468/2014/05/P05007.

[31] Chen Y, Li X, Xu J. Convexified modularity maximization for degree-corrected stochastic block models. Ann Stat. 2018;46(4):1573-1602. doi:10.1214/17-AOS1595.

[32] Zhao Y, Levina E, Zhu J. Consistency of community detection in networks under degree-corrected stochastic block models. Ann Stat. 2012;40(4):2266-2292. doi:10.1214/12-AOS1036.

[33] Gulikers L, Lelarge M, Massoulié L. A spectral method for community detection in moderately sparse degree-corrected stochastic block models. Adv Appl Probab. 2017;49(3):686-721. doi:10.1017/apr.2017.18.

[34] Wang T, Dai W, Jiao P, et al. Identifying influential nodes in dynamic social networks based on degree-corrected stochastic block model. Int J Mod Phys B. 2016;30(16):1650092. doi:10.1142/S0217979216500922.

[35] Fasino D, Tudisco F. The expected adjacency and modularity matrices in the degree corrected stochastic block model. Special Matrices. 2018;6(1):110-121. doi:10.1515/spma-2018-0010.

[36] Lei J, Zhu L. Generic Sample Splitting for Refined Community Recovery in Degree Corrected Stochastic Block Models. Stat Sin. 2017. doi:10.5705/ss.202015.0279.

[37] Ragain S. Community Detection via Discriminant functions for Random Walks in the degree-corrected Stochastic Block Model. 2017.

[38] Qian X, Chen Y, Minca A. Clustering Degree-Corrected Stochastic Block Model with Outliers. 2019. arXiv Preprint arXiv:1906.03305.

[39] Ma S, Su L, Zhang Y. Determining the Number of Communities in Degree-corrected Stochastic Block Models. J Mach Learn Res. 2021;22(69):1-63. doi:1809.01028.

[40] Passino FS, Heard NA, Rubin-Delanchy P. Spectral clustering on spherical coordinates under the degree-corrected stochastic blockmodel. Technometrics. 2022; 1-12. doi: 10.1080/00401706.2021.2008503.

[41] Hosseini SS, Noorossana R. Performance evaluation of EWMA and CUSUM control charts to detect anomalies in social networks using average and standard deviation of degree measures. Qual Reliab Eng Int. 2018;34(4):477-500. doi:10.1002/qre.2267.
[42] Noorossana R, Hosseini SS, Heydarzade A. An overview of dynamic anomaly detection in social networks via control charts. Qual Reliab Eng Int. 2018;34(4):641-648. doi:10.1002/qre.2278.
[43] Mazrae Farahani E, Baradaran Kazemzadeh R. Phase I monitoring of social network with baseline periods using poisson regression. Commun Stat- Theory Methods. 2019;48(2):311-331. doi:10.1080/03610926.2017.1408836.
[44] Fotuhi H, Amiri A, Taheriyoun AR. A novel approach based on multiple correspondence analysis for monitoring social networks with categorical attributed data. J Stat Comput Simul. 2019;89(16):3137-3164. doi:10.1080/00949655.2019.1657429.
[45] Farahani EM, Baradaran Kazemzadeh R, Noorossana R, et al. A statistical approach to social network monitoring. Commun Stat- Theory Methods. 2017;46(22):11272-11288. doi:10.1080/03610926.2016.1263741.
[46] Azarnoush B, Paynabar K, Bekki J, et al. Monitoring temporal homogeneity in attributed network streams. J Qual Technol. 2016;48(1):28-43. doi:10.1080/00224065.2016.11918149.
[47] Miller BA, Arcolano N, Bliss NT. Efficient anomaly detection in dynamic, attributed graphs: Emerging phenomena and big data. In: 2013 IEEE International Conference on Intelligence and Security Informatics. 2013;179-184. doi:10.1109/ISI.2013.6578815.
[48] Fotuhi H, Amiri A, Maleki MR. Phase I monitoring of social networks based on Poisson regression profiles. Qual Reliab Eng Int. 2018;34(4):572-588. doi:10.1002/qre.2273.
[49] Hazrati- Marangaloo H, Noorossana R. Detecting outbreaks in temporally dependent networks. Qual Reliab Eng Int. 2019;35(6):1753-1765. doi:10.1002/qre.2473.
[50] Zou N, Li J. Modeling and change detection of dynamic network data by a network state space model. IISE Trans. 2017;49(1):45-57. doi:10.1080/0740817X.2016.1198065.
[51] Aiello W, Chung F, Lu L. A random graph model for massive graphs. In: Proceedings of the thirty-second annual ACM symposium on Theory of computing. 2000;171-180. doi:10.1145/335305.335326.
[52] Holland PW, Laskey KB, Leinhardt S. Stochastic blockmodels: First steps. Soc Networks. 1983;5(2):109-137. doi:10.1016/0378-8733(83)90021-7.
[53] Karrer B, Newman ME. Stochastic blockmodels and community structure in networks. Phys Rev E. 2011;83(1):016107. doi:10.1103/PhysRevE.83.016107.
[54] Vanhatalo E, Kulahci M. The effect of autocorrelation on the Hotelling T^2 control chart. Qual Reliab Eng Int. 2015;31(8):1779-1796. doi:10.1002/qre.1717.
[55] Mason RL, Tracy ND, Young JC. Decomposition of T 2 for multivariate control chart interpretation. J Qual Technol. 1995;27(2):99-108. doi:10.1080/00224065.1995.11979573.
[56] Santos-Fernández E. Multivariate statistical quality control using R.Vol. 14. Springer Science & Business Media; 2012. doi:10.1007/978-1-4614-5453-3.