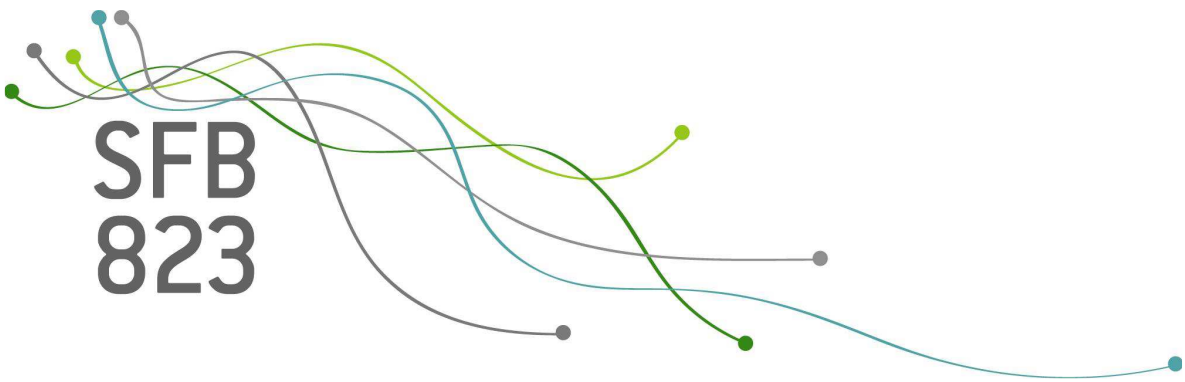
SFB 823

Online monitoring of dynamic networks using flexible multivariate control charts

Jonathan Flossdorf, Roland Fried, Carsten Jentsch

Nr. 33/2021

Discussion Paper



Online Monitoring of Dynamic Networks Using Flexible Multivariate Control Charts

Jonathan Flossdorf, Roland Fried and Carsten Jentsch

***Abstract*—The identification of differences in dynamic networks between various time points is an important task and involves statistical procedures like two-sample tests or change-point detection. Due to the rather complex nature of temporal graphs, the analysis is challenging which is why the complexity is typically reduced to a metric or some sort of a model. This is not only likely to result in a loss of relevant information, but common approaches also use restrictive assumptions and are therefore heavily limited in their usability. We propose an online monitoring approach usable for flexible network structures and able to handle various types of changes. It is based on a sound choice of a set of network characteristics under consideration of their mathematical properties which is crucial in order to cover the relevant information. Subsequently, those metrics are jointly monitored in a suitable multivariate control chart scheme which performs superior to a univariate analysis and enables both parametric and non-parametric usage. The user also benefits from a handy interpretation of the structural reasons for the detected changes which is a crucial advantage in the rather complex field of dynamic networks. Our findings are supported by an extensive simulation study.**

***Index Terms*—Change-point detection, multivariate control chart, network analysis, online monitoring, temporal graphs**

I. Introduction

DYNAMIC networks play an important part in many different application fields nowadays, ranging from biological [1], [2] and social sciences [3], [4] to logistic and transportation processes [5]. Suppose we observe a dynamic network $\mathcal{D} = \{D_t, t = 1, \dots, T\}$ which is a sequence of snapshots of the network of interest at various time points t. Each of those single networks D_t consist of a set of nodes V_t that may be connected through a set of links E_t. It is often of interest to decide whether there are any meaningful differences in the network structure between different time points, e.g. due to a changed consumer behavior in marketing networks, an increased communication in social networks, or a failure of a working machine in a manufacturing process. Other scenarios involve financial market analysis [6], network traffic monitoring [7], or connectomic applications [8]. Relevant statistical analysis procedures for such tasks are two-sample tests and change-point detection. Although parts of our results can be used for classical testing procedures as well, we focus on the latter one in our work.

J. Flossdorf, R. Fried and C. Jentsch are with the Department of Statistics, TU Dortmund University, Dortmund, 44221 Germany (email: flossdorf@statistik.tu-dortmund.de, fried@statistik.tu-dortmund.de, jentsch@statistik.tu-dortmund.de)

This research was financially supported by the Mercator Research Center Ruhr (MERCUR) with project number PR-2019-0019. Partial support from the collaborative research center (SFB 823) "Statistical Modelling of Nonlinear Dynamic Processes" of German Science Foundation (DFG) is gratefully acknowledged.

The goal of change-point detection [9], [10] is to reliably identify time points at which the structure of a dynamic network changes in a meaningful way. Traditionally, there are two perspectives towards this issue. One approach is to observe the whole sample $\mathcal{D}$ of interest first and to decide afterwards if one or more changes have happened (offline change-point detection). Another approach is to monitor the process sequentially in real-time in order to make an immediate decision at each newly observed time point (online monitoring). Our results hold for both scenarios, but for clarity of exposition we mainly focus on the usually more difficult and, in practice, often more relevant task of online monitoring in the following.

Statistical network theory is a quite complex field, since it covers a lot more information compared to normal data (relationships, intensities etc.). Dynamic networks even add another dimension by the consideration of the time component. A direct transfer of traditional monitoring approaches is therefore not feasible. Beside the actual construction of a suitable monitoring procedure, there are, in our view, two main issues: a) the definition of possible changes in network data and b) approaches for complexity reduction to handle the monitoring process.

A. Related work

Regarding problem a), it is not intuitively clear how a change may look like, since there is not only one but many, partly dependent, components which may trigger a change in network structure. A straightforward definition is presented in [11] by assigning a change to time point t, if $|f(D_t) - f(D_{t-1})| > c_0$ and $|f(D_t) - f(D_{t+1})| \leq c_0$ for some scoring function $f : D_t \mapsto \mathbb{R}$ and a threshold c_0. However, this approach is largely limited to the suitability of the applied scoring function as it addresses only those changes, for which $f(.)$ is able to capture the relevant information. In a more comprehensive context, the need of categorizing network changes with respect to their structural levels including nodes, communities or subgraphs is mentioned in [12]. In a former work [13], we presented such a general categorization in order to handle this issue. It covers global as well as local changes of single components (e.g. nodes and links) and addresses their combinations such that more complex structural changes (e.g. in blockmodels) are also considered.

Regarding problem b), there exist various approaches to reduce the complexity of dynamic networks in order to apply traditional monitoring schemes. They can be subdivided in model-based and metric-based approaches. For the former one, a dynamic network model is fitted and the specific parameters or residuals are then monitored with a traditional

control chart. Examples are state space models [14], degree-corrected stochastic blockmodels [15], temporal exponential random graph models [16], or Poisson regression models [17]. However, the model-based approach normally requires strict assumptions like a fixed node set (same nodes for each time point, no node dynamics) or knowledge of the underlying network structure. These assumptions are strongly restrictive, since they can only be used for a small field of applications. Furthermore, they can only detect a limited number of changes while ignoring those which does not affect the fitted model.

Metric-based approaches reduce the network by assigning a single metric or a combination of different metrics to each D_t. Hence, they are more flexible as they can be applied to most kind of networks. Exceptions are similarity measures like DeltaCon [18] or Graph Edit Distance [19] which sequentially compare each D_t to a reference network. For those approaches a fixed node set is required. This is not the case for any other network metric that is calculated with the sole information of D_t. Recent works used centrality metrics [20], matrix norms [21] or scan statistics [22]. The application under the consideration of time dependency is discussed in [23]. In a former work [13], we evaluated such metrics in dependent and independent setups and evaluated their individual suitability to detect changes in various situations.

B. Contribution

We expand on our univariate results of [13] and propose a monitoring approach which is based on the usage of a multivariate set of metrics. In our view, the joint monitoring of multiple network characteristics promises to strongly mitigate the information loss of metric-based procedures, because the network information can be covered by various metric types of different sensitivities. Furthermore, as the univariate interpretations still hold, all advantages are maintained which is particularly a broad flexibility and applicability. While a multivariate procedure is a logical extension of the univariate case and is partially mentioned in the univariate literature [13], [20], there does not exist - to the best of our knowledge - a thorough multivariate method able to handle the main challenges that are crucial for a successful application. From our perspective, those are a sound choice of a set of network metrics, the combination with a suitable choice of control charting procedures and the final interpretation of the results. To support the flexibility of our approach, we combine it with both distribution-free and parametric monitoring schemes.

The paper is organized as follows: Section II offers a short recap of the suitability of univariate network metrics for monitoring purposes and extends them to a proper multivariate usage. In Section III, suitable multivariate control chart procedures are defined and adapted to our metric sets. The results are supported by an extensive simulation study in Section IV which underlines the reliability of the procedure in flexible change situations. Section V contains some concluding remarks.

II. Detecting Changes in Dynamic Networks

We start by demonstrating the complexity and challenges to reliably monitor changes in dynamic networks ranging from the basic definition of a change in network data to the actual monitoring strategy. We recap the suitability of univariate network characteristics in different change situations and extend the existing foundations to a more flexible multivariate conception.

TABLE I
Examples and Application Scenarios

Type	Examples
GLC	increased/decreased communication in social networks changed activity in cyber networks (e.g. due to malware)
LLC	formation of new hotspots in disease networks changed route layout in transportation networks
GNC	new advertising strategy in customer networks addition of new destinations in tourism networks
LNC	restructuring of supply chains in logistics networks creation of new leading positions in profession networks

A. Changes in Network Data

The first challenge is to understand which type of changes may happen in network data. Because a dynamic network consists of various structural elements, a simple shift of location or scale parameters like in traditional scenarios does not exist. As explained, we focus on a flexible setup and prefer general type of changes [12] rather than specialized changes and therefore follow the change definition of [13]. In this context, the idea is to consider the influence of each structural network element to obtain a thorough categorization of possible changes. These elements are a) links, b) nodes, and c) extra information that may be put on either nodes or links (i.e. covariates). Each element is assumed to be able to trigger a change either in a global or local manner. A short summary of all scenarios is listed below. Note that we do not consider changes due to covariates here, since their type of occurrence is hugely dependent on the underlying application field.

- **Global Link Change (GLC):** The change is triggered by a significantly increased or decreased link amount. This is assumed to happen globally, i.e. the changed link probability affects each node equally.
- **Local Link Change (LLC):** Similar to GLCs, but the changed link behavior only affects a few nodes which either get more or less influence, i.e. the network structure changes to a more centralized or flat hierarchy
- **Global Node Change (GNC):** The node amount increases or decreases significantly, because new nodes enter the network or existing ones leave it.
- **Local Node Change (LNC):** Only a few influential nodes enter or leave the network which results in a significant impact on the network structure.

Typical examples and relevant application scenarios for these types of changes are given in Table I. While all of those changes may occur individually, it is likely that some of them happen simultaneously, e.g. a global increasement of links may be the consequence of the entry of new nodes in the network. This is referred to as mixed-type changes (MTC), which also enables capturing more complex change scenarios like those in blockmodels or subgraphs.

TABLE II
CONSIDERED NETWORK METRICS

Type	Metric	Definition	Notes
Matrix-based	Frobenius (Fr)	$\sqrt{\sum_{i=1}^{n_t}\sum_{j=1}^{n_t} a_{t,ij}^2}$	
	Spectral (Sp)	$\max_{\|x\|_2=1} \|A_t x\|_2$	corresponds to the largest eigenvalue of A_t for undirected networks
Centralities	Closeness (Cl)	$c_i^C = \frac{1}{\frac{1}{n-1}\sum_{j\in V} \mathrm{d}(i,j)}$	$\mathrm{d}(i,j)$: shortest distance between nodes i and j
	Degree (De)	$c_i^D = \sum_{j=1}^{n_t} a_{t,ij}$	
	Betweenness (Be)	$c_i^B = \sum_{i\neq j\neq l} \frac{\sigma_{jl}(i)}{\sigma_{jl}}$	$\sigma_{jl}(i)$: total amount of shortest paths betw. nodes j and l (which pass through node i)
	Eigenvector (Ei)	$c_i^E = \frac{1}{\lambda}\sum_{j=1}^{n_t} a_{t,ij} c_j^E$	recursive definition, λ: largest eigenvalue of A_t

B. Metric-based Network Monitoring

We already briefly discussed possible complexity reduction procedures in order to monitor a temporal series of graphs which involves model-based and metric-based approaches. The former one is quite restrictive and not applicable ad hoc, e.g. parametric assumptions have to be met. This also affects that only a few model-specific change types can be detected, e.g. LNCs and GNCs are ignored due to the common assumption of a fixed node set. Those restrictions are especially unfavorable, if the structure and behavior of the network of interest is not explicitly known beforehand. Dynamic networks are commonly quite prone to this issue due to their high dimensionality and potentially high dynamics.

The complexity reduction step of the metric-based procedure is even more radical, but those approaches provide a more flexible monitoring tool without restrictions for a broad application field. Consider that each network D_t of the dynamic network $\mathcal{D}$ is reduced to a scalar $f(D_t) = s_t$. Thus, the vector $\mathbf{s} = (s_1, \ldots, s_T)$ contains the captured information of the applied metric to $\mathcal{D}$. See Table II for an overview of the used metrics in this work which are also common choices when dealing with network monitoring [20], [21]. Note that the matrix norms are calculated based on the temporal series of adjacency matrices $A_t \in \mathbb{R}^{n_t \times n_t}$ of the networks D_t, where n_t is the number of nodes at time point t. For the sake of simplicity, we mainly focus on undirected and unweighted networks which means that a matrix entry $a_{t,ij} = 1$, if a link between nodes i and j exists, and $a_{t,ij} = 0$ otherwise.

Whereas the matrix norms are global metrics for the whole network D_t, the centrality scores are locally defined for each node. To transform them into a global network metric, we consider two approaches. This involves the average score over all nodes, i.e.

$$c_{\text{avg}} = \frac{1}{n}\sum_{i\in V} c_i,$$

and some scale metric by taking the deviation to the largest observed score

$$c_{\text{dev}} = \sum_{i\in V} (\max_{j\in V}(c_j) - c_i).$$

We specify the used version in the following by noting an Av- and De- prefix (e.g. AvCl for Average Closeness). Note that for the eigenvector centrality, both versions are affine linear transformations to each other and therefore achieve equal monitoring performances. Thus, we only consider AvEi for this centrality and, hence, nine metrics in total. It is possible to use any other network metric as well, but many are strongly related to the proposed ones and do not contribute added value (e.g. Average Path Length, Density) [13].

C. Extensions to the Multivariate Scenario

The most crucial step in the metric-based monitoring setup is to choose a suitable metric which covers as much of the relevant information of D_t as possible. Because the complex network structure gets reduced to a single scalar value, the information loss is typically quite high in comparison to a well-fitted model-based approach. It is therefore of utmost importance to be aware of the information a metric is able to capture in order to understand which type of changes it is able to detect. This topic is extensively discussed in our related work [13] which serves as a theoretical foundation for the derivations in the following. See Table III for a short summary of the suitability of the considered metrics in the presented change scenarios.

Due to the described information loss, it is not too surprising that no single metric is able to perform well in every scenario. However, for each type of change there exist multiple metrics that work reasonably. Hence, it is a logical step to use multiple metrics jointly in order to capture various pieces of information to mitigate the loss. Formally, for each D_t a vector $\mathbf{s}_t = (s_{t1}, ..., s_{tp})$ of p different scores is calculated at each time point t. This promises to result in a more flexible monitoring procedure which has an improved performance for flexible change types. Furthermore, it still maintains the advantages of the univariate setup, since it is directly applicable as it does not require restrictions regarding the network data. Hence, the interpretation of a detected change keeps handy, because the univariate interpretations still hold.

In the practical examination, there are a few challenges including a) the determination of the number of considered

TABLE III
PERFORMANCES OF THE METRICS IN VARIOUS SITUATIONS.

Change Type	Fr	Sp	AvCl	DeCl	AvDe	DeDe	AvBe	DeBe	AvEi
GLC	++	++	+	−	++	−	+	−	o
LLC	+	o	−	+	o	+	o	+	++
GNC	−−	++	+	−	+	−	+	−	o
LNC	−−	−	o	+	−	+	−	+	++

++ = suitable, + = mostly suitable, o = moderate performance/dependent on other circumstances, − = rather not suitable, −− = not suitable

metrics p and b) the choice of a suitable set of metrics. Regarding a), we would like to note that a higher value of p is in theory helpful for capturing more information, but is also more likely to generate flexibility and interpretation issues. Moreover, some metrics might be highly correlated due to a similar definition which would make the monitoring procedure (see Section III) more difficult for higher p. Hence, we set $p = 3$ in this work as this seems to be a good trade-off between information capturing and maintaining flexibility.

Regarding b), it is crucial to be aware of the suitability of each metric in each situation. In this context, the main statement of Table III is that most metrics either perform reasonable in change situations that affect the network globally (i.e. GLC, GNC) or in local change scenarios (i.e. LLC, LNC). Based on this, we may classify most metrics into two different performance groups A = {Sp, AvCl, AvDe, AvBe}, which perform well in global change scenarios, and group B = {DeCl, DeDe, DeBe} that perform superior in local setups. Remaining are Fr, which can handle link changes but ignores node changes, and AvEi that is theoretically affected by all change types but sometimes to a lesser extent.

The final choice of a suitable set of metrics is dependent on the goal and the expectations of the application. We propose the consideration of six different sets which are denoted in Table IV. The idea behind Set 1 is to use one metric out of both classes A and B, in order to capture various types of information. The usage of the average eigenvector then provides some neutral perspective. This balanced setup seems to be a promising candidate for an ad-hoc application in which users do not know what to expect and how a change might look like (e.g. networks with high dynamics like social networks). The other sets offer some more unbalanced setups by 2 vs. 1 and 3 vs. 0 combinations. These are constructed for more specialized cases where the user might be interested in detecting some particular changes which frequently occur in the corresponding application (see examples in Table I). Finally, Set 6 emphasizes link changes more by taking the Frobenius norm into account.

III. MULTIVARIATE CONTROL CHARTS FOR NETWORK DATA

We now move on to the practical implementation of a monitoring setup for network data by combining and adapting traditional control chart schemes with an intelligent choice of a set of the presented network metrics. We shortly recap the required theory in the traditional scenario first and explain the expansion to the multivariate network scenario afterwards.

TABLE IV
USED SETUP CHOICES IN THIS WORK

No.	Name	Set of Metrics		
1	SBE	Sp	DeBe	AvEi
2	SDC	Sp	AvDe	DeCl
3	BDS	DeBe	DeDe	Sp
4	CDS	AvCl	AvDe	Sp
5	BCD	DeBe	DeCl	DeDe
6	FSB	Fr	Sp	DeBe

A. General Online Monitoring Procedure

The main goal of online monitoring is to detect anomalies in a process as soon as possible after their occurrence. Typically, the process of interest is subdivided into two phases. In Phase I, it is assumed that the process is somewhat stable and reliably represents the typical state of the underlying system without meaningful deviations. The system is then called to be in-control. In Phase II on the other hand, the actual monitoring takes place by deciding if an incoming signal sufficiently matches the in-control state. An alarm is triggered if this is not the case and the signal is classified as out-of-control [9].

Consider $\{Y_t, t = 1, \dots, T\}$ to be a sequence of a random variable of interest with conditional density $P_\theta(Y_t|Y_{t-1}, ..., Y_1)$ and τ to be the unknown change time. If $\tau > t$, then the conditional density parameter θ is constant with $\theta = \theta_1$. For $\tau \leq t$, it applies $\theta = \theta_2$. The goal is to detect the anomaly as soon as possible with a fixed rate of false alarms before τ. An estimation of θ_1 and θ_2 is often not necessary, but might be useful for interpretation purposes regarding possible reasons for a change.

In terms of the practical usage, so-called control charts are applied. First, a metric x_t is chosen which a) can be calculated for each time point t, b) covers and represents most relevant aspects of the behavior of the system (i.e. Y_t), and c) is able to identify all considered changes by a sensitive reaction to them. This metric serves as the main input for the control statistic z_t which is calculated for the process at each time point t. Depending on the setup, z_t can be the metric itself (e.g. in memory-free Shewhart charts) or some sort of transformation $z_t = g(x_t)$ (e.g. in memory-based EWMA or CUSUM charts). In Phase I, z_t is expected to represent the in-control state of the system and, hence, to be a stable process without meaningful deviations. This information is used to define the upper and lower control limits h_u and h_l. Those limits are chosen under the consideration of the desired rate of false alarms and can be derived by parametric or distribution-free procedures. In

Phase II, if we observe $h_l \leq z_t \leq h_u$, the process is deemed in-control. Otherwise an alarm is triggered at time point t to signalize a detected change.

B. Control Charts for Traditional Multivariate Data

In practice, it is customary to improve the monitoring procedure by taking more process metrics into account. Their independent usage in a univariate manner is possible but not recommended, since it is inefficient and may result in erroneous conclusions [10]. Hence, the construction of multivariate control charts, which consider the metrics jointly, are of interest.

In this context, the most basic multivariate chart is the Hotelling T^2 chart. Suppose we observe a vector $\mathbf{x_t} = (x_{t1}, ..., x_{tp})$ of p different process metrics at each time point t, then the corresponding control statistic is calculated by

$$z_t = (\mathbf{x_t} - \bar{\mathbf{x}})' \mathbf{S}^{-1} (\mathbf{x_t} - \bar{\mathbf{x}}),$$

where $\bar{\mathbf{x}}$ and $\mathbf{S}$ are the sample mean vector and covariance matrix of the underlying observations. Since z_t mainly takes the squared deviation to the sample mean into account, it is non-negative and we expect values near zero if the process is in-control. Therefore, only an upper control limit has to be derived. This can be done under parametric assumptions with an approximation via the F-distribution which yields

$$h_u = \frac{p(n+1)(n-1)}{n^2 - np} F_{\alpha,p,n-p},$$

where n is the number of observations in Phase I and α the desired false alarm rate.

In many practical applications the met distributional assumption might be unjustified which can have a strong negative impact on the monitoring quality. To avoid such issues, the usage of non-parametric techniques seems promising. In this context, a bootstrap approach was proposed [24], which is able to efficiently handle the monitoring process, even if the observed data is non-normal or unknown. It works as follows. First, the statistic z_t is calculated for all T observations of Phase I as before, which yields the vector $\mathbf{z} = (z_1, ..., z_T)$. Subsequently, B Bootstrap samples with sample size T are drawn from $\mathbf{z}$ and for each of those samples the $(1-\alpha)$-quantile is calculated. The upper control limit h_u is then determined by taking the average over those values.

The Hotelling T^2 charts are multivariate extensions of a univariate Shewhart-type control chart, because they only use information of the current observation which makes them rather insensitive to small shifts. Memory-based control charts like exponential weighted moving average charts (EWMA) overcome this issue and it also exists a multivariate version (MEWMA) [25], for which the control statistic is defined as

$$z_t = \mathbf{m}_t' \mathbf{S_t}^{-1} \mathbf{m}_t,$$

where $\mathbf{m_t}$ is recursively defined as

$$\mathbf{m}_t = \lambda(\mathbf{x}_t - \bar{\mathbf{x}}) + (1-\lambda)\mathbf{m}_{t-1}.$$

The estimation of the covariance matrix is given by

$$\mathbf{S}_t = \frac{\lambda}{2-\lambda}[1 - (1-\lambda)^{2t}]\mathbf{S},$$

where $\mathbf{S}$ is the estimated covariance matrix given all observations from Phase I. While the formula of the control statistic is quite similar to the Hotelling T^2 chart, the main difference lies in the intermediate step of calculating $\mathbf{m}_t$, where the smoothing parameter $\lambda \in [0, 1]$ serves as a factor for providing weights to past observations and the current one. For $\lambda = 1$, the MEWMA setup corresponds to the Hotelling T^2 chart. Optimal control limits depending on λ, the number of variables p and the desired false alarm rate can be found in several works [26], [27]. In general, MEWMA charts with small values of λ are rather robust to the normal assumption yielding satisfying results for different distributions of the underlying data [10].

C. Adaption to Dynamic Network Data

After the recap and the described extensions of the theoretical background in Sections II and III, we now move on to the combination of traditional control chart schemes with the proposed multivariate metric sets. Our recommended procedure is listed step by step below.

1) Selection of p and a suitable set of metrics
2) Univariate calculation of each metric
3) Selection of a suitable control chart procedure
4) Calculation of the corresponding control statistic and control limits (Phase I)
5) Monitoring of new observations (Phase II)
6) Stop at a detected change
7) Interpretation of the change with the help of univariate behavior

While 2), 4), 5), and 6) are relatively straightforward steps of common statistical process control, the quality of the procedure is hugely dependent on a careful execution of steps 1), 3), and 7).

We already discussed our recommended procedure for step 1) in Section II-C, where we provided multivariate metric sets for various situations under consideration of the suitability of the univariate metrics in various change situations.

Regarding step 3), the choice of a suitable control chart setup is as important as the choice of a metric set. The monitoring performance of the parametric Hotelling T^2 chart is dependent on the quality of the fit of the applied F-distribution. To the best of our knowledge, no complete asymptotic inference was yet derived for the considered network metrics due to their complex nature. Consequently, putting parametric assumptions on their joint distribution seems rather implausible. The parametric Hotelling T^2 chart is known to react rather sensitive to violations of its distributional assumption [28] and might suffer from reliability issues in this context. We expect that this is especially the case for rather unbalanced multivariate sets of metrics, since their marginal distributions tend to be more similar to each other and are sensitive to the same impact factors, which may result in a more skewed joint distribution. For more balanced setups, this effect is likely to be weakened, because different sensitivities are involved. Furthermore, we expect a worse performance of the Hotelling T^2 chart for lower false alarm rates α, because the corresponding control limit h_u is dependent on the $(1-\alpha)$-quantile of the applied distributional assumption.

The higher quantile is likely to be a bad approximation for the corresponding quantile of the empirical distribution, which is - for very low α - sensitive to the observed extreme values that might especially play a role for rather unstable and high-dynamic processes like networks. Overall, this effect tends to be smoothened for larger values of α as the quantile of interest is shifted more towards the center of the distribution. The explained impacts affect the MEWMA chart as well, but to a lesser extent. Due to the smoothing of the control statistic that involves the consideration of past observations, the chart is noticeably more robust against non-normal behavior. This is particularly the case for lower values of the smoothing parameter λ which weakens the individual influence of the current observation. However, note that inertia issues might be a consequence of this [25]. The non-parametric Hotelling T^2 chart might be the safest choice for a reliably constructed control chart when using the considered metric sets. As the bootstrap procedure is directly dependent on the empirical distributions, we expect the chart to be more robust against various type of metric sets, i.e. to perform on the same level for all sets. Obviously, its quality increases for larger sample sizes (i.e. longer in-control phase), and the bootstrap procedure ensures that it works reasonable in most cases of lower sample sizes as well. However, in the latter cases its performance might not be superior to the parametric candidates anymore as we will see in Section IV.

Regarding step 7), note that we aim to monitor a temporal series of networks $\mathcal{D} = \{D_t, t = 1, \dots, T\}$ instead of a simple process variable Y_t. Hence, the change parameter θ gets more complex (see Section III-A), which makes its interpretation in a change situation all the more important. This especially concerns the purpose of maintaining transparency and reliability of the monitoring tool. In practice, we propose to stop after a detected change and to analyze the corresponding network D_t carefully. This can be handily done by descriptively analyzing the univariate values of the used metrics and applying their interpretations given in [13]. While focussing on time-independent setups and the related challenges in this paper, the presented procedure can be handily extended to allow also for time-dependency. For instance, this can be done by adding an intermediate step between 2) and 3), where an ARIMA model is fitted to the multivariate series of metrics in order to monitor its residuals similar to [13], [23], [29].

IV. Simulation Study

To underline our findings, we execute an extensive simulation study in the following. We generate numerous example situations of each described change type and analyze the performances of all proposed metric sets in combination with the described control chart procedures. We compare their results with the univariate approach and demonstrate performance differences. Recall that a classical model-based approach is not feasible here, because we aim to detect flexible changes and set no assumptions to the network structure (e.g. no fixed node set). In a second part, we extend the study by analyzing more practical relevant mixed-type changes in different situations of stochastic blockmodels.

TABLE V
Data generation processes using Erdös-Renyi (ER) graphs.

Type	Data Generation
GLC	In-Control: ER graph with probability p and n nodes Out of Control: ER graph with probability $\tilde{p}$ and n nodes
LLC	In-Control: ER graph with probability p and n nodes Out of Control: heterogenous ER graph with a changed probability $\tilde{p}$ for k central nodes
GNC	In-Control: ER graph with at each time point a randomly chosen node size in $[n - n \cdot d, n + n \cdot d]$ and a randomly chosen link amount in $[m - m \cdot d, m + m \cdot d]$. Out of Control: ER graph with at each time point a randomly chosen node size in $[\tilde{n} - \tilde{n} \cdot d, \tilde{n} + \tilde{n} \cdot d]$ and a randomly chosen link amount in $[m - m \cdot d, m + m \cdot d]$.
LNC	In-Control: ER graph with probability p and at each time point a random node size $n_{\text{flex}} \in [n - n \cdot d, n + n \cdot d]$. Out of Control: heterogenous ER graph with at each time point a random node size $\tilde{n}_{\text{flex}} \in [\tilde{n} - \tilde{n} \cdot d, \tilde{n} + \tilde{n} \cdot d]$ and changed probability $\tilde{p}$ for k central nodes and a probability of $\frac{pn_{\text{flex}} - \tilde{p}k}{\tilde{n}_{\text{flex}} - k}$ for all other nodes.

TABLE VI
Parameter setup for each situation according to Table V with $n = 100$.

Type	Intensity	Parameter Setup
GLC	small	$p = 0.3$, $\tilde{p} = 0.31$
	moderate	$p = 0.7$, $\tilde{p} = 0.68$
	heavy	$p = 0.4$, $\tilde{p} = 0.45$
LLC	small	$p = 0.2$, $\tilde{p} = 0.3$, $k = 5$
	moderate	$p = 0.4$, $\tilde{p} = 0.55$, $k = 3$
	heavy	$p = 0.3$, $\tilde{p} = 0.6$, $k = 1$
GNC	small	$d = 0.05$, $m = 1500$, $\tilde{n} = 115$
	moderate	$d = 0.05$, $m = 1500$, $\tilde{n} = 130$
	heavy	$d = 0.05$, $m = 1500$, $\tilde{n} = 150$
LNC	small	$d = 0.03$, $p = 0.2$, $\tilde{n} = 105$, $\tilde{p} = 0.3$, $k = 5$
	moderate	$d = 0.03$, $p = 0.7$, $\tilde{n} = 102$, $\tilde{p} = 0.85$, $k = 2$
	heavy	$d = 0.03$, $p = 0.3$, $\tilde{n} = 101$, $\tilde{p} = 0.7$, $k = 1$

A. Setup

We generate sampling data for each of the four described change-types of Section II-A (GLC, LLC, GNC, and LNC). For each of those, we execute three sub-situations with varying change intensities (small, moderate, and heavy). This results in 12 scenarios in total, which are repeated 1000 times to obtain a reliable performance analysis. The data generating processes as well as their different parameter setups to control the intensity are given in Tables V and VI, respectively. Note that those are similar to our former study [13] in order to maintain comparability across both works.

For each situation, we simulate dynamic networks of length $T = 1400$ and use the first 1000 observations as Phase I in order to reliably calibrate the control chart. The change time is set to happen at time point $\tau = 1050$. The control limits of the control charts are set with a false alarm rate of $\alpha = 0.5\%$. While the rather large number of observations in Phase I is interesting to obtain meaningful findings about the theoretical performances, we are aware that, from a practical point of view, those numbers are hard to provide in some applications

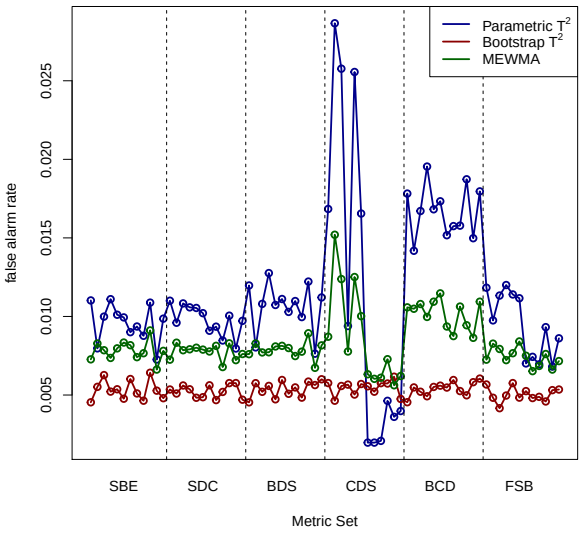


Fig. 1. Average empirical false alarm rates for all simulated situations.

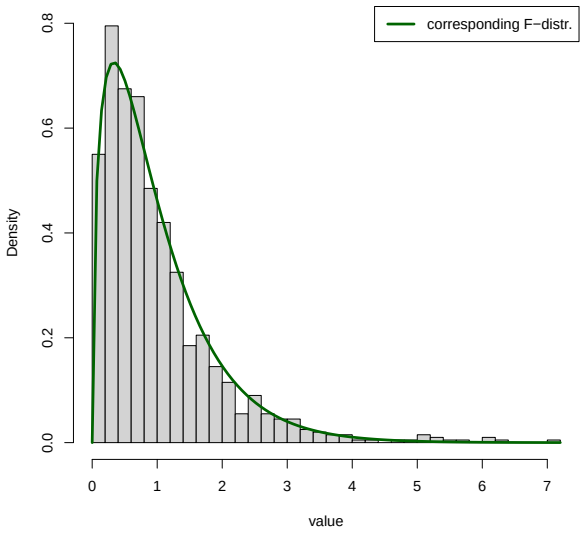


Fig. 2. Example situation for a comparison of the simulated empirical distribution with the corresponding F-approximation of the parametric Hotelling T^2 chart.

with rather high dynamics. To this purpose, we examine the performances in a more practical setting in Section IV-C.

For all scenarios, we evaluate the performances of the proposed multivariate metric sets in combination with the three presented control chart procedures. We compare their results with the univariate approach and use ARL_0 and ARL_1 as performance measures. The ARL_0 is defined as the in-control average run length which can reach an optimal value of $\frac{1}{\alpha} = 200$ in our setup. On the other hand, ARL_1 calculates the post-change average run length which measures the delay to detection, i.e. the number of time points an alarm is sent after the actual change has occurred.

To maintain comparability across the univariate and multivariate setups, we compare memory-free settings and memory-based procedures separately. Hence, the parametric and non-parametric Hotellings T^2 charts are compared with the Shewhart chart and the MEWMA chart with the EWMA procedure. For the latter ones, we try different smoothing parameters $\lambda \in \{0.1, 0.2, ..., 0.7\}$ and report the ones with the best results.

B. Performances

We begin with the evaluation of the in-control state. Fig. 1 illustrates the empirical false alarm rates for all considered multivariate control charts in each of the 72 examined scenarios. The results largely meet our expectations as the parametric Hotelling T^2 procedure tends to yield relatively low control limits. This particularly seems to be the case for more unbalanced setups (e.g. CDS, BCD) which tend to generate more skewed joint distributions as explained in Section III-C. Overall, however, the desired fit is not reached for more balanced sets as well, since their false alarm rates lie above the desired α and above the ones of the other two procedures. See Fig. 2 for an example of the quality of the fit where the deviation of the empirical distribution to the assumed F assumption is clearly visible, particularly at the tails of the distribution. Regarding the other control charts, the results support the statement that MEWMA is more robust against possible parametric violations. However, the non-parametric bootstrap approach yields the most reliable in-control results for the rather long in-control phase and the small value of α.

We saw that the charts produce quite different ARL_0 values, although they were designed to hold a fixed false alarm rate. Obviously, charts with a lower ARL_0 will produce lower values of ARL_1 on the same data set, since the control limit h_u is lower. For the ARL_1 analysis, we therefore only report the ones of the bootstrap chart and concentrate on the performance differences of the applied metric sets. See Table VII for an overview of the results.

Overall, the results meet our expectations. The univariate metrics might perform reasonable in special scenarios, but are having clear weaknesses in others. The multivariate procedures, on the other hand, perform clearly more reliable over all situations and are more robust to the underlying change-type which underlines the improved flexiblity compared to the univariate approach. On a further note, the results are handily interpretable. We can e.g. take a closer look at the performance of the most balanced set SBE (consisting of Sp, BeDe, and AvEi). Regarding the involved univariate performances, Sp is able to handle global changes, but is having problems with LLCs and especially LNCs. For BeDe the behavior is vice versa. Their joint monitoring, however, leads to promising results for all scenarios, since their effects are combined. The additional consideration of AvEi, which is in theory sensitive to all types but to a lesser extent, serves slightly supportive to all impacts and as an overall smoothing factor. The behavior for more unbalanced multivariate setups is similar as they provide more flexibility overall compared to the involved univariate candidates. Particularly, the ARL_1 for the "non-

TABLE VII
ARL_1 FOR UNIVARIATE AND MULTIVARIATE PERFORMANCES. BEST ARL_1 FOR EACH SCENARIO IS PRINTED IN BOLD.

		Univariate Setup									Multivariate Setup					
Type	Intensity	Fr	Sp	AvCl	DeCl	AvDe	DeDe	AvBe	DeBe	AvEi	SBE	SDC	BDS	CDS	BCD	FSB
GLC	small	9.68	10.23	9.65	132.52	9.68	143.33	**9.61**	121.84	120.27	33.01	25.24	38.98	44.49	55.93	51.67
	moderate	1.77	**1.69**	1.76	136.07	1.77	132.93	1.79	104.42	70.21	3.03	3.29	2.88	1.76	17.13	3.10
	heavy	**1.00**	**1.00**	**1.00**	126.29	**1.00**	162.48	**1.00**	80.42	22.21	**1.00**	**1.00**	**1.00**	**1.00**	1.93	**1.00**
LLC	small	6.31	**4.14**	6.99	6.19	7.47	9.35	6.31	5.57	11.48	4.18	4.54	4.85	8.28	4.97	7.88
	moderate	16.50	11.04	14.55	2.77	16.50	3.84	16.50	2.92	3.88	2.47	2.50	5.75	12.61	2.18	**2.06**
	heavy	33.14	14.06	29.06	**1.01**	33.63	1.02	33.14	1.02	1.02	**1.01**	**1.01**	1.05	13.36	**1.01**	**1.01**
GNC	small	313.74	1.26	1.01	21.63	**1.00**	83.32	1.26	41.96	37.16	3.07	34.12	1.45	1.38	3.91	19.87
	moderate	312.89	**1.00**	**1.00**	8.89	**1.00**	27.17	**1.00**	8.08	8.54	**1.00**	7.67	**1.00**	**1.00**	1.05	2.01
	heavy	311.25	**1.00**	**1.00**	103.87	**1.00**	15.16	**1.00**	2.31	2.87	**1.00**	2.89	**1.00**	**1.00**	**1.00**	**1.00**
LNC	small	139.98	41.72	1.85	1.81	1.06	1.75	11.42	2.06	1.77	1.51	1.52	**1.02**	1.12	1.60	1.50
	moderate	143.48	11.58	1.07	1.30	1.13	1.12	10.07	1.21	1.13	**1.00**	7.67	**1.00**	**1.00**	1.05	2.01
	heavy	149.93	127.78	65.00	**1.00**	14.97	**1.00**	115.68	**1.00**	**1.00**	**1.00**	**1.00**	**1.00**	1.04	**1.00**	**1.00**

specialized" cases (i.e. the cases, for which all involved metrics are not really suitable) improved as the small sensitivities of the single metrics have a larger impact if they are considered jointly, see e.g. the BCD performance for GNCs or the CDS performance for LLCs. Despite the improved performance compared to the univariate metrics, the values are obviously higher than those of more balanced setups in these situations. Moreover, it is somewhat surprising that they also do not clearly outperform the more balanced sets in "specialized" cases, for which they are mainly constructed.

C. Extension to mixed-type changes

While this examination under rather rigid settings gave us crucial insights on the theoretical performance limits of the considered monitoring applications, we now examine if the studied behaviors still hold in more practical application examples. For this purpose, we consider mixed-type changes (MTC) in the popular scenario of community changes in stochastic blockmodels (SBM) [30]. The explicit setups can be found in Table VIII, where K represents the number of communities, p_i the intra-group link probability of group $i \in \{1, ..., K\}$, p_{ij} the inter-group link probability between groups i and j, and n_i the number of nodes in group i. Furthermore, we reduce the length of the in-control phase to 100 in order to examine the performance of the charts in more data-restrictive circumstances. Due to the shorter in-control length, we set the false alarm rate to $\alpha = 5\%$.

The in-control results shown in Fig. 3 are different to before. Whereas the non-parametric version clearly outperformed the parametric control charts in Section IV-B, the performances are more equal now. The non-parametric approach suffers from the shorter in-control length, because the estimation of the theoretical distribution becomes more unreliable by applying a smaller data sample to the bootstrap procedure. However, the chart still performs reasonable as it only lies approx. 0.5% above the desired α. Another advantage is the robustness against different metric sets. Overall, the parametric control charts perform better than before and reach similar performances for more balanced metric sets to the non-parametric candidate. An explanation is the higher value of α, for which the charts are designed, as explained in Section III-C. However, the sensitivity to the applied metric set still holds as the performance gets quite unstable for unbalanced sets.

TABLE VIII
SBM PARAMTER SETUPS FOR THE COMMUNITY CHANGES

No.	In-Control	Changes
1	$K = 3, p_1 = 0.7, p_2 = 0.6,$ $p_3 = 0.8, p_{ij} = 0.1,$ $n_1 = n_2 = 33, n_3 = 34$	$p_1 = 0.9, p_2 = 0.7,$ $p_3 = 0.9, p_{ij} = 0.2,$ $n_1 = n_2 = n_3 = 40$
2	$K = 3, p_1 = 0.7, p_2 = 0.6,$ $p_3 = 0.8, p_{ij} = 0.1,$ $n_1 = n_2 = 33, n_3 = 34$	$p_1 = 0.4, p_2 = 0.9$
3	$K = 3, p_1 = 0.7, p_2 = 0.6,$ $p_3 = 0.3, p_{ij} = 0.1,$ $n_1 = 30, n_2 = 20,$ $n_3 = 50$	$K = 4, p_4 = 0.3,$ $n_3 = 30, n_4 = 20$
4	$K = 3, p_1 = 0.7, p_2 = 0.6,$ $p_3 = 0.3, p_{ij} = 0.1,$ $n_1 = 30, n_2 = 20,$ $n_3 = 50$	$K = 4, p_3 = p_4 = 0.49$ $n_3 = 30, n_4 = 20$

The ARL_1 results in Table IX can be interpreted similarly to before. Siuation 1 describes a change with an increased popularity of the whole network with an increased inter- and intra-communication (link amount) and the arrival of new members which can be seen as a MTC of GNC and GLC. Apart from some univariate deviation centralities, all considered metric sets perform well. The second scenario addresses changes, where the importance between two groups is shifted, which results in an increased communication of one group and a decreased communication in the other. While most univariate metrics are not able to handle this change type as their ARL_1 does match their set ARL_0, the multivariate sets perform reasonable, particularly the more balanced ones. The last two situations address changes in the number of communities, e.g a split of one group into two different ones. For the third situation, the link probability in both new groups stays the same as before which results in an overall decreased link amount due to decreased link probability of those nodes, which were in one group before and are in different ones now. Hence, it is an relatively easy GLC situation and all applied metrics achieve a satisfactory performance. We designed the last scenario such that the overall link probability

TABLE IX
ARL_1 FOR THE COMMUNITY CHANGES. BEST ARL_1 FOR EACH SCENARIO IS PRINTED IN BOLD.

	Univariate Setup									Multivariate Setup					
Case	Fr	Sp	AvCl	DeCl	AvDe	DeDe	AvBe	DeBe	AvEi	SBE	SDC	BDS	CDS	BCD	FSB
1	**1.00**	**1.00**	**1.00**	19.50	**1.00**	14.16	**1.00**	1.89	1.08	**1.00**	**1.00**	**1.00**	**1.00**	**1.00**	**1.00**
2	22.13	10.04	20.24	26.30	22.17	24.56	20.09	12.20	**1.58**	1.61	3.43	9.63	2.08	5.76	3.49
3	1.01	2.78	1.01	1.86	1.01	3.06	1.01	3.09	1.48	1.39	**1.00**	1.94	**1.00**	1.42	**1.00**
4	20.30	18.44	7.97	15.02	21.30	22.22	7.68	14.89	21.38	4.94	13.85	14.38	**3.90**	5.92	11.88

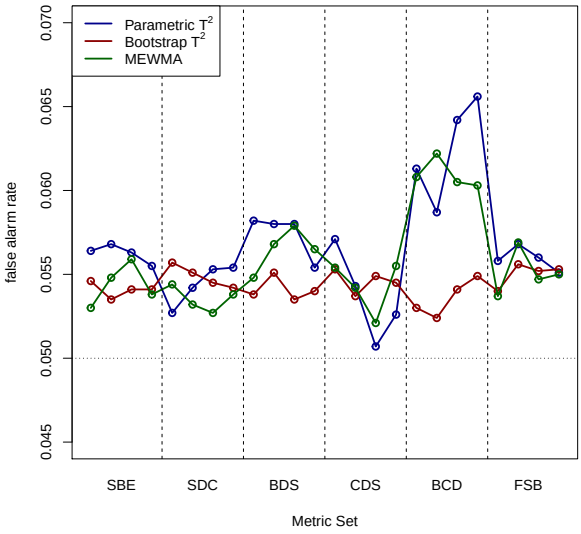


Fig. 3. Average empirical false alarm rates for all simulated situations.

stays the same after the split. This makes the detection more challenging which results in higher ARL_1 values. However, the multivariate sets again underline their superior flexibility as they achieve better performances in this situation than their univariate counterparts.

V. CONCLUSION

The detection of temporal differences in a time series of graphs is a rather challenging task due to the complex nature of dynamic networks. We proposed an extension of a metric-based approach to a multivariate setup and its combination with suitable control charting procedures involving parametric as well as non-parametric setups. We explicitly explained the challenges of such a multivariate design and presented recommendations including a sound choice of a suitable set of metrics, its combination with a suitable control chart, and the final interpretation of the results. We further underlined our statements with the help of a simulation study in which a thoroughly designed multivariate approach outperforms the univariate procedure by offering a more flexible solution to the problem of change detection in dynamic networks.

REFERENCES

[1] D. S. Bassett and O. Sporns, "Network neuroscience," *Nature neuroscience*, vol. 20, no. 3, pp. 353–364, 2017.

[2] R. J. Prill, P. A. Iglesias, and A. Levchenko, "Dynamic properties of network motifs contribute to biological network organization," *PLoS biology*, vol. 3, no. 11, p. e343, 2005.

[3] P. Sarkar and A. W. Moore, "Dynamic social network analysis using latent space models," *Acm Sigkdd Explorations Newsletter*, vol. 7, no. 2, pp. 31–40, 2005.

[4] P. J. Carrington, J. Scott, and S. Wasserman, *Models and methods in social network analysis*. Cambridge university press, 2005, vol. 28.

[5] D.-H. Lee and M. Dong, "Dynamic network design for reverse logistics operations under uncertainty," *Transportation research part E: logistics and transportation review*, vol. 45, no. 1, pp. 61–71, 2009.

[6] D. Durante and D. B. Dunson, "Bayesian dynamic financial networks with time-varying predictors," *Statistics & Probability Letters*, vol. 93, pp. 19–26, 2014.

[7] J. Sun, D. Tao, and C. Faloutsos, "Beyond streams and graphs: dynamic tensor analysis," in *Proceedings of the 12th ACM SIGKDD international conference on Knowledge discovery and data mining*, 2006, pp. 374–383.

[8] D. Durante, D. B. Dunson, and J. T. Vogelstein, "Nonparametric bayes modeling of populations of networks," *Journal of the American Statistical Association*, vol. 112, no. 520, pp. 1516–1530, 2017.

[9] M. Basseville and I. V. Nikiforov, *Detection of abrupt changes: theory and application*. Prentice Hall Englewood Cliffs, 1993, vol. 104.

[10] D. C. Montgomery, *Statistical quality control*. Wiley Global Education, 2012.

[11] S. Ranshous, S. Shen, D. Koutra, S. Harenberg, C. Faloutsos, and N. F. Samatova, "Anomaly detection in dynamic networks: a survey," *Wiley Interdisciplinary Reviews: Computational Statistics*, vol. 7, no. 3, pp. 223–247, 2015.

[12] I. U. Hewapathirana, "Change detection in dynamic attributed networks," *Wiley Interdisciplinary Reviews: Data Mining and Knowledge Discovery*, vol. 9, no. 3, pp. 1286–1306, 2019.

[13] J. Flossdorf and C. Jentsch, "Change detection in dynamic networks using network characteristics," *IEEE Transactions on Signal and Information Processing over Networks*, vol. 7, pp. 451–464, 2021.

[14] N. Zou and J. Li, "Modeling and change detection of dynamic network data by a network state space model," *IISE Transactions*, vol. 49, no. 1, pp. 45–57, 2017.

[15] J. D. Wilson, N. T. Stevens, and W. H. Woodall, "Modeling and detecting change in temporal networks via the degree corrected stochastic block model," *Quality and Reliability Engineering International*, vol. 35, no. 5, pp. 1363–1378, 2019.

[16] A. Malinovskaya and P. Otto, "Online network monitoring," *Statistical Methods & Applications*, vol. 30, no. 5, pp. 1337–1364, 2021.

[17] E. M. Farahani, R. Baradaran Kazemzadeh, R. Noorossana, and G. Rahimian, "A statistical approach to social network monitoring," *Communications in Statistics-Theory and Methods*, vol. 46, no. 22, pp. 11 272–11 288, 2017.

[18] D. Koutra, N. Shah, J. T. Vogelstein, B. Gallagher, and C. Faloutsos, "Deltacon: Principled massive-graph similarity function with attribution," *ACM Transactions on Knowledge Discovery from Data (TKDD)*, vol. 10, no. 3, pp. 1–43, 2016.

[19] H. Bunke, P. J. Dickinson, M. Kraetzl, and W. D. Wallis, *A graph-theoretic approach to enterprise network dynamics*. Springer Science & Business Media, 2007, vol. 24.

[20] I. McCulloh and K. M. Carley, "Detecting change in longitudinal social networks," *Military Academy West Point NY Network Science Center (NSC)*, 2011.

[21] I. Barnett and J.-P. Onnela, "Change point detection in correlation networks," *Scientific reports*, vol. 6, no. 1, pp. 1–11, 2016.

[22] J. Neil, C. Hash, A. Brugh, M. Fisk, and C. B. Storlie, "Scan statistics for the online detection of locally anomalous subgraphs," *Technometrics*, vol. 55, no. 4, pp. 403–414, 2013.

[23] D. Ofori-Boateng, Y. R. Gel, and I. Cribben, "Nonparametric anomaly detection on time series of graphs," *Journal of Computational and Graphical Statistics*, vol. 30, no. 3, pp. 756–767, 2021.
[24] P. Phaladiganon, S. B. Kim, V. C. Chen, J.-G. Baek, and S.-K. Park, "Bootstrap-based t2 multivariate control charts," *Communications in Statistics - Simulation and Computation*, vol. 40, no. 5, pp. 645–662, 2011.
[25] C. A. Lowry, W. H. Woodall, C. W. Champ, and S. E. Rigdon, "A multivariate exponentially weighted moving average control chart," *Technometrics*, vol. 34, no. 1, pp. 46–53, 1992.
[26] S. S. Prabhu and G. C. Runger, "Designing a multivariate ewma control chart," *Journal of Quality Technology*, vol. 29, no. 1, pp. 8–15, 1997.
[27] S. Knoth, "Arl numerics for mewma charts," *Journal of Quality Technology*, vol. 49, no. 1, pp. 78–89, 2017.
[28] Z. G. Stoumbos and J. H. Sullivan, "Robustness to non-normality of the multivariate ewma control chart," *Journal of Quality Technology*, vol. 34, no. 3, pp. 260–276, 2002.
[29] B. Pincombe, "Anomaly detection in time series of graphs using arma processes," *Asor Bulletin*, vol. 24, no. 4, p. 2, 2005.
[30] P. W. Holland, K. B. Laskey, and S. Leinhardt, "Stochastic blockmodels: First steps," *Social networks*, vol. 5, no. 2, pp. 109–137, 1983.