



Online monitoring of dynamic networks using flexible multivariate control charts

Jonathan Flossdorf¹ · Roland Fried¹ · Carsten Jentsch¹

Received: 19 January 2023 / Revised: 27 April 2023 / Accepted: 28 April 2023 / Published online: 17 May 2023
© The Author(s) 2023

Abstract

Change-point detection in dynamic networks is a challenging task which is particularly due to the complex nature of temporal graphs. Existing approaches are based on the extraction of a network's information by the reduction to a model or to a single metric. Whereas the former one requires restrictive assumptions and has limited applicability for real-world social networks, the latter one may suffer from a huge information loss. We demonstrate that an extension to a well-balanced multivariate approach that uses multiple metrics jointly to cover the relevant network information can overcome both issues, since it is applicable to arbitrary network shapes and promises to strongly mitigate the information loss. In this context, we give guidelines on the crucial questions of how to properly choose a suitable multivariate metric set together with the choice of a meaningful parametric or nonparametric control chart and show that an improper application may easily lead to unsatisfying results. Furthermore, we identify a solution that achieves reasonable performances in flexible circumstances in order to give a reliably applicable approach for various types of social networks and application fields. Our findings are supported by the use of extensive simulation studies, and its applicability is demonstrated on two real-world data sets from economics and social sciences.

Keywords Change-point detection · Multivariate control chart · Social network analysis · Statistical data analysis · Temporal graphs

1 Introduction

Dynamic networks play an important role in many different application fields nowadays, ranging from biological (Bassett and Sporns 2017; Prill et al. 2005) and social sciences (Sarkar and Moore 2005; Carrington et al. 2005) to logistic and transportation processes (Lee and Dong 2009). 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 these single networks D_t consists of a set of nodes V_t that may be connected through a set of links E_t . Note that we do not only allow the number and positions of edges to differ between different time points, but

also the number of nodes may change. It is often of interest to decide whether there are differences in the (dynamic) stochastic network generating process 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 (Durante and Dunson 2014), network traffic monitoring (Sun et al. 2006), or connectomic applications (Durante et al. 2017). Relevant statistical analysis procedures for such tasks are two-sample tests and change-point detection. Although our results can be employed for classical testing procedures as well, we focus on the latter one in our work.

The main purpose of change-point detection (Basseville and Nikiforov 1993; Montgomery 2012) is to identify time points at which the structure of a dynamic network changes in a meaningful way. Traditionally, there are two perspectives toward this issue. One approach is to observe the whole sample $\mathcal{D} = \{D_1, \dots, D_T\}$ of interest first and to decide afterward 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).

✉ Jonathan Flossdorf
flossdorf@statistik.tu-dortmund.de

Roland Fried
fried@statistik.tu-dortmund.de

Carsten Jentsch
jentsch@statistik.tu-dortmund.de

¹ Department of Statistics, TU Dortmund University, Dortmund 44225, Germany

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.

In classical data setups of statistical process control, a time series of scalar values is monitored which means that the detection is mostly limited to a shift of the mean or variance of the observed measurement. Statistical network data, however, covers a lot more information compared to classical data setups (relationships, intensities etc.). Dynamic networks even add another dimension by the consideration of the time component. This has, on the one hand, the advantage of offering a more informative representation of an underlying system enabling, e.g., a better interpretation of observed changes. On the other hand, the derivation of statistical inference approaches gets more challenging. For the task of online monitoring, there are two main issues in this context. First, it is not clear how a change might look like and, second, a direct transfer of traditional monitoring approaches to dynamic network data is hardly feasible as the process is described by a time series of networks and not by a time series of scalar values anymore. It is therefore crucial to a) be aware of possible changes in network data and b) construct monitoring strategies that are suitable for network data.

1.1 Related work

Regarding 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 Ranshous et al. (2015) 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 $f(\cdot)$ 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 Hewapathirana (2019). Such a general categorization to handle this issue is presented in Flossdorf and Jentsch (2021). 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 b), monitoring strategies for dynamic network data are typically based on a reduction of the complexity of the underlying time series of networks. There exist various methods to do so which can be subdivided in model-based, embedding-based and metric-based approaches. For the model-based methods, 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 (Zou and Li 2017), degree-corrected stochastic blockmodels (Wilson et al. 2019), temporal exponential random graph models (Malinovskaya and Otto 2021), or Poisson regression models (Farahani et al. 2017; Motalebi et al. 2021). However, the model-based approach commonly requires potentially restrictive 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 often too strong, since they can only be used for a small field of applications. Furthermore, they allow to detect only a limited number of changes while ignoring those which does not affect the fitted model.

A related approach is the usage of embedding techniques. In a nutshell, the main goal of network embedding is to learn a mapping function in order to map each single node to a lower-dimensional vector. This results in a latent lower-dimensional feature representation of a graph that reduces noise and redundant information, but still aims to maintain important structural information (Cui et al. 2018). There exist various of these network representation methods such as node2vec (Grover and Leskovec 2016) and DeepWalk (Perozzi et al. 2014) that are developed based on random walks. Other approaches use matrix factorization (Belkin and Niyogi 2001; Ou et al. 2016). Whereas these are specifically designed for static networks, there also exist versions for dynamic networks like temporalnode2vec (Haddad et al. 2020) and DynamicTriad (Zhou et al. 2018). A thorough analysis of the stability of such embedding approaches can be found in Gürsoy et al. (2021), where also the importance for ensuring alignment is stressed as misaligned embeddings might negatively impact the performance for dynamic network inference tasks such as change detection. For the usage of change detection, embedding methods are particularly used to detect vertex-based changes in a time series of graphs like, e.g., detecting vandal users in online social networks using a vector autoregression approach (Li et al. 2019). Moreover, Hewapathirana et al. (2020) propose a spectral embedding method that particularly addresses sparsity and degree heterogeneity. Lin et al. (2022) construct an embedding change detection approach with a focus on node coordinates in a latent space that are used to model edge dependencies. Further approaches involve Sun and Liu (2018), Grattarola et al. (2019), Duan et al. (2020), and Xie et al. (2023).

Lastly, metric-based approaches summarize the network information 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 arbitrary types of networks. Exceptions are similarity measures like DeltaCon (Koutra et al. 2016) or Graph Edit Distance (Bunke et al. 2007) which sequentially compare each D_t to a reference network. For

those approaches, a fixed node set is obviously 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 (McCulloh and Carley 2011), matrix characteristics (Barnett and Onnela 2016; Hazrati-Marangaloo and Noorossana 2021), scan statistics (Neil et al. 2013), or further network metrics like the clustering coefficient (Kendrick et al. 2018). The application under the consideration of time dependency is discussed in Ofori-Boateng et al. (2021). Obviously though, univariate metric-based methods tend to lead to flexibility issues. A single metric is only able to capture some, but not all information that might be relevant for change detection. Therefore, in Flossdorf and Jentsch (2021), such metrics are analyzed in dependent and independent setups and evaluated with respect to their individual suitability in various situations. A multivariate usage of three network centrality metrics and the network density in a multivariate EWMA chart is studied in Salmasnia et al. (2020). However, a fixed node set as well as a Gaussian distribution for each involved metric is assumed which noticeably reduces the flexibility advantage that metric-based approaches have compared to the other mentioned categories of network change detection.

1.2 Contribution

We expand on the univariate results of the metric-based approaches and study a monitoring method that uses a multivariate set of metrics. We illustrate that such a multivariate approach overcomes the mentioned issues of model-based and other metric-based methods, since it is ad hoc applicable without restrictive assumptions and provides flexibility gains as the relevant information is now simultaneously captured by multiple metrics of various types. We base our analysis on theoretical considerations, simulation-based evidences and practical applications. The need of such a multivariate analysis is mentioned in the univariate literature (McCulloh and Carley 2011; Flossdorf and Jentsch 2021). Contrary to the multivariate investigations in Salmasnia et al. (2020), we do not limit our analysis on a certain set of metrics or control chart and also allow for arbitrary network structures including a dynamic node set and a nonparametric setup. Our primary objective is to present a solution on how exactly a general multivariate procedure shall be implemented and performed. In this context, we identify three challenges that are crucial for a successful application: a) a sound choice of a set of network metrics, b) their combination with a suitable choice of control charting procedures, and c) the final interpretation of the results. We introduce strategical guidelines to solve these challenges and demonstrate that ignoring them could easily lead to erroneous conclusions and unsatisfying results. We identify a balanced solution that achieves reliable performances in various situations as well

as propose solutions that are specialized for more specific scenarios. To support the flexibility of this study, we study both distribution-free and parametric monitoring schemes.

The paper is organized as follows: Sect. 2 offers a short recap of the necessary theory regarding change detection in network data and multivariate control charts. In Sect. 3, we formulate the multivariate network change detection procedure in details and study the suitability of different metric sets, various control charts, and the interaction of both. The results are supported by an extensive simulation study in Sect. 4 which confirms the reliability in various change situations and shows superior performance compared to univariate approaches. In Sect. 5, the procedure is applied to two real-world social network data sets. Section 6 contains some concluding remarks.

2 Existing foundations

In this section, we recap the existing concepts of change detection in network data as well as of traditional statistical process control that are necessary for the introduction of the methodology in Sect. 3.

2.1 Changes in network data

The first challenge is to understand which types 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 above, we focus on a flexible setup and prefer general types of changes (Hewapathirana 2019) rather than specialized changes and therefore follow the change definition of Flossdorf and Jentsch (2021). 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, attributes). 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 caused by 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

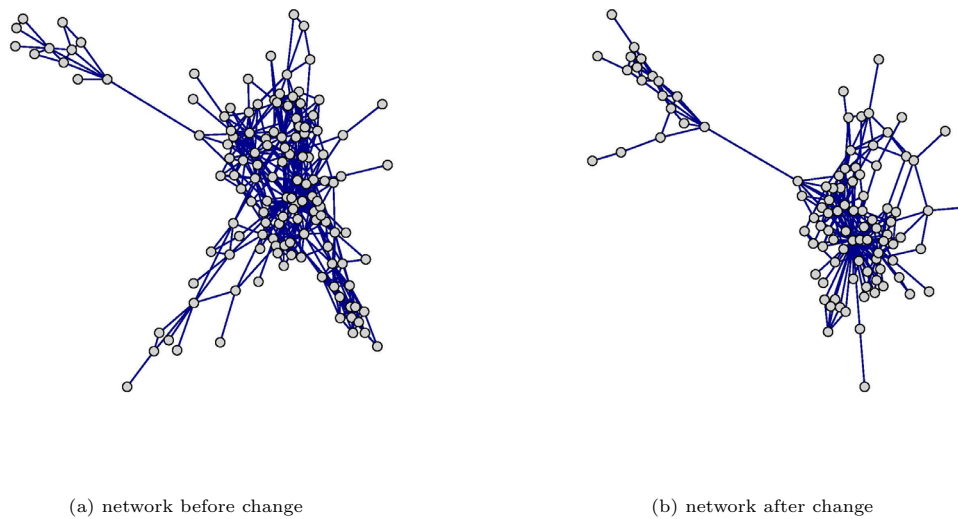


Fig. 1 Real-world example for a mixed-type change (MTC): Enron E-mail communication

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.

While all of these changes may occur individually, it is likely that some of them happen simultaneously, e.g., a global increase 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. A real-world social network example for MTCs is depicted in Fig. 1. It shows the e-mail communication between employees of the former US company Enron. It is quite obvious that the overall structure of the network has not changed drastically between the both observed time points. Hence, we have no hint to local changes. However, we can observe a clearly reduced node and link amount in Fig. 1 compared to Fig. 1. The network hence got more sparse both in terms of nodes and links which hints to a mixed-type change of GLC and GNC. Further typical examples and relevant application scenarios for the different types of change are given in Table 1.

2.2 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

Table 1 Examples and application scenarios

Type	Examples
GLC	Increased/decreased communication in communication 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 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 $\mathbf{s} = (s_1, \dots, s_T)$ contains the captured information of the applied metric to \mathcal{D} . Below, we give a summary of typical metrics that are common choices when dealing with network monitoring. For instance, the Frobenius norm is studied in Barnett and Onnela (2016), the Spectral norm in Chen et al. (2021) and as it is equal to the largest eigenvalue for undirected networks also in parts in Hazrati-Marangaloo and Noorossana (2021), the centrality metrics in McCulloh and Carley (2011), Salmasnia et al. (2020), and Ofori-Boateng et al. (2021). 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.

- Frobenius norm (F): $\sqrt{\sum_{i=1}^{n_t} \sum_{j=1}^{n_t} a_{t,ij}^2}$.
- Spectral norm (S): $\max_{\|x\|_2=1} \|A_t x\|_2$.
- Closeness centrality (C): $c_i^C = (\frac{1}{n-1} \sum_{j \in V} d(i,j))^{-1}$.
- Degree centrality (D): $c_i^D = \sum_{j=1}^{n_t} a_{t,ij}$.
- Betweenness centrality (B): $c_i^B = \sum_{i \neq j \neq l} \frac{\sigma_{jl}(i)}{\sigma_{jl}}$.
- Eigenvector centrality (E): $c_i^E = \frac{1}{\lambda} \sum_{j=1}^{n_t} a_{t,ij} c_j^E$.

In this context, $d(i, j)$ denotes the shortest path between two nodes i and j , $\sigma_{jl}(i)$ the number of shortest path between two nodes j and l (that pass through node i), and λ the largest eigenvalue of A_t .

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. On the one hand, this involves the average score over all nodes, i.e.,

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

On the other hand, we can use a 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 m or d index (e.g., C_m for mean Closeness and C_d for

Closeness deviation). Note that for the eigenvector centrality, both versions are affine linear transformations to each other and therefore achieve equal monitoring performances. It is possible to use any other network metric as well, e.g., the Average Path Length or the Clustering Coefficient like in Kendrick et al. (2018), but many are strongly related to the presented ones and do not contribute added value (Flossdorf and Jentsch 2021).

2.3 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, 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* (Basseville and Nikiforov 1993).

Consider $\{Y_t, t = 1, \dots, T\}$ to be a sequence of a random variable of interest with conditional density $P_\theta(Y_t | Y_{t-1}, \dots, Y_1)$ 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, 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 signal a detected change.

2.4 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 Montgomery (2012). Hence, the construction of multivariate control charts, which consider the metrics jointly, is of interest.

In this context, the most basic multivariate chart is the Hotelling T^2 chart. Suppose we observe a vector $\mathbf{x}_t = (x_{t1}, \dots, x_{tp})$ 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 not be justified which can have a strong negative impact on the monitoring performance. To avoid such issues, the usage of nonparametric techniques seems promising. In this context, a bootstrap approach was proposed in Phaladiganon et al. (2011), which is able to efficiently handle the monitoring process, even if the observed data is non-Gaussian 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 $\mathbf{z} = (z_1, \dots, z_T)$. 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, the control statistic of a multivariate version (MEWMA) (Lowry et al. 1992) 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 (Prabhu and Runger 1997; Knoth 2017). 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 (Montgomery 2012).

3 Multivariate metric-based monitoring solutions

After the recap of the theoretical background of the existing foundations in Sect. 2, we now move on to the practical implementation of a monitoring setup for network data by combining and adapting multivariate control chart schemes with an intelligent choice of a set of network metrics. Making use of network metrics for the monitoring of networks, first, the practitioner has to choose the (multivariate) set of metrics in conjunction with a suitable control chart procedure. While there is a whole variety of (and even more combinations of) networks metrics that could be used, it is generally unclear which combination of such network metrics should be used in which dynamic network scenario to detect deviations from the control state as reliable as possible. As the further steps are rather straightforward, that is, the calculation of the corresponding control statistic and control limits (Phase I), the monitoring of new observations (Phase II) and how to stop at a detected change, the interpretation of the change is also not straightforward. In this context, note that this work aims to present a general procedure and particularly focuses on developing an ad hoc solution that is applicable to arbitrary network shapes and application fields and enables a reliable change analysis in various situations. Therefore, we present general guidelines for different categories of application scenarios and identify the most flexible solution that performs reliably even in unknown circumstances.

Table 2 Performances of the metrics in various situations

Change type	F	S	C_m	C_d	D_m	D_d	B_m	B_d	E
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

3.1 Selection of a suitable set of metrics

We use the metrics that were presented in Sect. 2.2 as these are the common choices for extracting the information of a network (McCulloh and Carley 2011; Hazrati-Marangaloo and Noorossana 2021; Chen et al. 2021; Barnett and Onnela 2016; Salmasnia et al. 2020; Ofori-Boateng et al. 2021). Because the information of a network gets reduced to a single scalar value, the resulting information loss is typically quite large. It is therefore crucial to be aware of the information a metric is able to capture in order to understand which type of change it is able to detect which was discussed in Flossdorf and Jentsch (2021). See Table 2 for a short summary of the individual 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 all scenarios. However, for each type of change, there exist multiple metrics that work reasonably well. Hence, it is a natural approach to use multiple metrics jointly in order to capture various pieces of information to mitigate the loss. Formally, for each D_t , a vector $s_t = (s_{t1}, \dots, s_{tp})$ of p different scores is calculated at each time point t . In this work, we mainly focus on sets with $p = 3$.

In this context, the main challenge is the choice of a suitable set of metrics. The main statement of Table 2 is that most metrics either perform reasonably well in change situations that affect either 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 $A = \{S, C_m, D_m, B_m\}$, which perform well in global change scenarios, and group $B = \{C_d, D_d, D_b\}$ that perform superior in local setups. Remaining are the Frobenius norm, which can handle link changes but ignores node changes, and average Eigenvector Centrality 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. Based on the described univariate behavior of the presented metrics, we propose to consider the following multivariate monitoring strategies.

- *I - Balanced Setups*: This category represents the most flexible monitoring strategy. The goal is to achieve a reliable performance for as many as possible change types. Particularly, changes of moderate to high intensity should be detected. Corresponding sets include one metric from each of the performance groups A and B. The third metric is the average eigenvector centrality. This balanced setup is a promising candidate for ad hoc applications 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). Example setup: SB_dE .
- *Ila - Balanced Setups with a focus on global changes*: In this category, it is still of interest to be sensitive to as many change types as possible, but with a clear focus on the detection of global changes. This category involves all 2 vs. 1 combinations (i.e., two metrics out of group A and one out of group B). Example setup: SD_mC_d .
- *Ilb - Balanced Setups with a focus on local changes*: The same as Ila, just with a major focus on local changes (i.e., one metric out of group A and two out of group B). Example setup: B_dD_dS .
- *IIla - Unbalanced setups for global changes*: It is only of interest to detect global changes in the network - even those which are characterized by a small change intensity. The detection of other change types is not relevant. The corresponding metric sets are constructed with three metrics out of group A. Example setup: C_mD_mS .
- *IIlb - Unbalanced setups for local changes*: The same as IIIa, just for local changes. The metric sets are constructed with three metrics out of group B. Example setup: $C_dD_dB_d$.
- *IV - Setups with a particular focus on link changes*: This category represents all metric sets in which the Frobenius norm is part of, since this metric is specialized to detect link changes. Changes purely triggered by nodes can also be emphasized by combining the Frobenius norm with other metrics that are sensitive to it. Example setup: FSB_d .

In a nutshell, the idea behind Category I 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 Categories 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 1). Finally, Category VI emphasizes link changes more by taking the Frobenius norm into account.

As mentioned, we used $p = 3$ as the number of considered metrics for the development of the categorization scheme above. The usage of other values is of course possible. However, we would like to note that a higher value of p is in theory helpful for capturing more information, but is also prone to more uncertainty picked up by the monitoring procedure, which might result in power losses. Moreover, some of the metrics might be highly correlated due to a similar definition which would make the monitoring procedure less efficient for higher p and might even lead to erroneous conclusions. Moreover, values of $p > 3$ do likely not contribute added value as there are the two explained performance groups. Considering this, $p = 3$ seems to be a good trade-off between information capturing and maintaining flexibility. This is also validated by our simulation study in Sect. 4.

A particular metric set with a value of $p = 4$ is considered in Salmasnia et al. (2020). For the remainder of this paper, we denote this set by SAL. The included metrics are deviation metrics for the Betweenness, Degree, and Closeness centralities with a similar definition to B_d , D_d , and C_d . Additionally, the network density is used as a fourth metric. Overall, this metric set can be classified in our Category IIIb as it uses three metrics that are specifically sensitive to detect local changes. As we will see in the simulation study, the additional consideration of the density does not change much compared to the $B_d D_d C_d$ set as these metrics are highly correlated and quite dominant in their multivariate combination.

3.2 Selection of a suitable multivariate control chart

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 has yet been 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 sensitively to violations of its distributional assumption (Stoumbos and Sullivan 2002) 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. This behavior can particularly be observed for sets of Categories IIIa and IIIb and also for the presented one of Salmasnia et al. (2020). 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 depends on the $(1 - \alpha)$ -quantile of the applied distributional assumption. Higher quantiles are likely to be bad approximations for the corresponding quantiles 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 less pronounced for larger values of α as the quantile of interest is shifted more toward 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-Gaussian 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 (Lowry et al. 1992).

Finally, the nonparametric 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 types of metric sets, i.e., to perform on a similar 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 reasonably 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 Sect. 4.

3.3 Interpretation of the results and further analyses

To conclude, we recommend using a balanced metric set like $SB_d E$ together with the nonparametric Hotellings T^2 chart for the most reliable solution in flexible scenarios. Regarding the interpretation of the monitoring results, note that we 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 Sect. 2.3) making 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 Table 2 in order to identify the underlying change type. Precise interpretation examples are given in Sect. 5.

While focussing on temporally 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 fitting an ARIMA model to the multivariate series of metrics in order to monitor its residuals similar to Ofori-Boateng et al. (2021); Flossdorf and Jentsch (2021); Pincombe (2005).

4 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 the proposed multivariate metric strategies in combination with the described control chart procedures. Additionally, we compare their results with the corresponding univariate metric performances as studied in McCulloh and Carley (2011), Ofori-Boateng et al. (2021), or Barnett and Onnela (2016) and with the multivariate approach SAL of Salmasnia et al. (2020). Recall that a comparison with typical model-based and embedding-based approaches is not feasible here in a meaningful way, as our study is designed to detect flexible changes without imposing any high-level assumptions on the network structure (e.g., fixed node set, model assumptions). In a second part, we extend the study by analyzing more practical relevant mixed-type changes in different situations of stochastic blockmodels. The code for the simulation study can be found in a GitHub repository¹ created by the authors.

4.1 Setup

We generate sampling data for each of the four described change types of Sect. 2.1 (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 3 and 4.

For each situation, we simulate dynamic networks of length $T = 1400$ and use the first 1000 observations as Phase

Table 3 Data generation processes using Erdős-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 $n_{flex} \in [n - n \cdot d, n + n \cdot d]$. Out of Control: heterogenous ER graph with at each time point a random node size $\tilde{n}_{flex} \in [\tilde{n} - \tilde{n} \cdot d, \tilde{n} + \tilde{n} \cdot d]$ and changed probability \tilde{p} for k central nodes and a probability of $\frac{pn_{flex} - \tilde{p}k}{\tilde{n}_{flex} - k}$ for all other nodes.

In-Control: Setup for the in-control behavior of the dynamic network (Phase I)

Out of Control: Setup for the out of control behavior of the dynamic network (Phase II)

Table 4 Parameter setup for each situation according to Table 3 with $n = 50$

Type	Intensity	Parameter setup
GLC	Small	$p = 0.4, \tilde{p} = 0.43$
	Moderate	$p = 0.4, \tilde{p} = 0.45$
	Heavy	$p = 0.4, \tilde{p} = 0.5$
LLC	Small	$p = 0.35, \tilde{p} = 0.45, k = 5$
	Moderate	$p = 0.4, \tilde{p} = 0.6, k = 2$
	Heavy	$p = 0.35, \tilde{p} = 0.65, k = 1$
GNC	Small	$d = 0.05, m = 800, \tilde{n} = 55$
	Moderate	$d = 0.05, m = 800, \tilde{n} = 60$
	Heavy	$d = 0.05, m = 800, \tilde{n} = 75$
LNC	Small	$d = 0.15, p = 0.4, \tilde{n} = 54, \tilde{p} = 0.6, k = 4$
	Moderate	$d = 0.15, p = 0.4, \tilde{n} = 52, \tilde{p} = 0.7, k = 2$
	Heavy	$d = 0.15, p = 0.4, \tilde{n} = 51, \tilde{p} = 0.8, k = 1$

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 $\alpha = 1\%$. While the rather large number of observations in Phase I is interesting to obtain meaningful findings about the theoretical performances and the suitability of the metric sets, we are aware that, from a practical point of view, those numbers are usually hard to provide in real-world applications. To this purpose, we examine the performances in

¹ <https://github.com/jonathanFlossdorf/NetworkMonitoring>

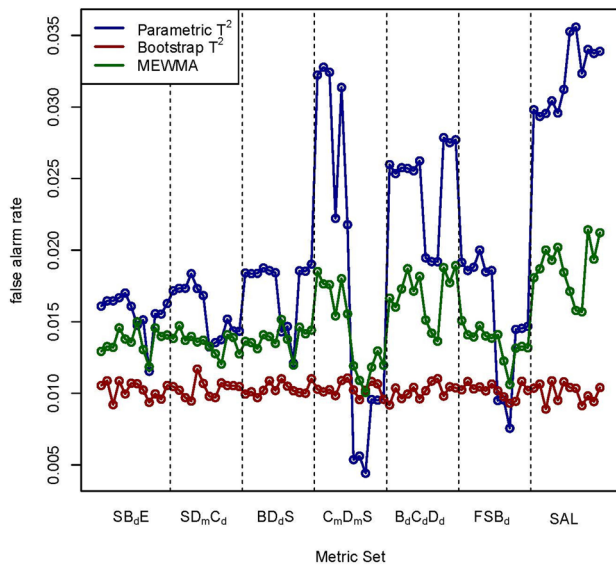


Fig. 2 Average empirical false alarm rates for all simulated situations. The desired value for α is 0.01

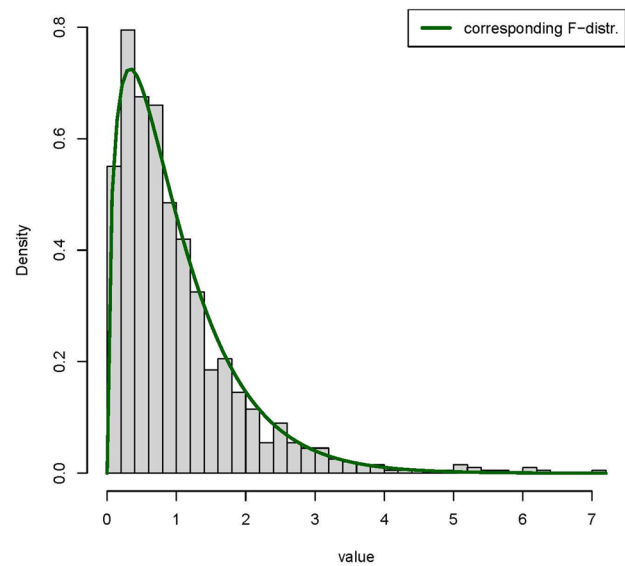


Fig. 3 Example situation for a comparison of the simulated empirical distribution with the corresponding F -approximation of the parametric Hotelling T^2 chart

practically more relevant simulation settings in Sect. 4.4 and analyze real-world networks in Sect. 5.

For all scenarios, we evaluate the performances of all triple combinations of the presented metrics and particularly focus on the results of the multivariate metric set categories that were derived in Sect. 3.1. We evaluate their performance in combination with all three presented control chart procedures and compare the results with the univariate approaches. As performance measures, we use ARL_0 and ARL_1 . The ARL_0 is defined as the in-control average run length which can reach an optimal value of $\frac{1}{\alpha} = 100$ 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 nonparametric 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 $\lambda \in \{0.1, 0.2, \dots, 0.7\}$ and report the ones with the best results.

4.2 Phase I performances

We begin with the evaluation of the in-control state. Figure 2 illustrates the empirical false alarm rates for all considered multivariate control charts in each of the 72 examined scenarios for all three applied control chart types.

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., C_mD_mS , $B_dC_dD_d$, SAL) which tend to generate more skewed joint distributions as explained in Sect. 3.2. 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. 3 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 nonparametric bootstrap approach clearly yields the most reliable in-control results for the rather long in-control phase and the small value of α .

4.3 Phase II performances

We saw that the charts produce quite different ARL_0 values, although they were designed to hold a common 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. Hence, the ARL_1 results should only be compared between the same applied control chart type. As the bootstrap chart achieved the most reliable Phase I performances, we therefore report the Phase II results only for the bootstrap chart and concentrate on the performance differences of the applied metric sets. See Table 5 for an

Table 5 ARL_1 for the studied change types. Best ARL_1 for each scenario across all setups (univariate and multivariate) is printed in bold

(a) Univariate setup performances										
Type	Intensity	F	S	C_m	C_d	D_m	D_d	B_m	B_d	E
GLC	small	3.12	3.24	3.03	94.78	3.12	109.89	3.17	55.63	50.61
	moderate	1.22	1.20	1.20	68.86	1.23	86.83	1.22	16.89	19.82
	heavy	1.00	1.00	1.00	45.02	1.00	102.13	1.00	4.71	9.82
LLC	small	9.19	7.70	8.40	25.22	9.19	27.44	9.53	60.99	36.00
	moderate	14.57	9.97	12.14	3.14	14.57	3.45	14.57	6.04	4.43
	heavy	22.52	11.60	17.67	1.41	32.52	1.46	21.72	2.27	1.81
GNC	small	144.02	2.19	1.41	45.33	2.17	78.00	1.34	8.16	11.98
	moderate	142.55	1.01	1.00	9.82	1.01	73.06	1.00	2.02	3.03
	heavy	135.90	1.00	1.00	1.47	1.00	61.49	1.00	1.34	1.06
LNC	small	93.22	39.53	1.06	1.16	9.01	1.07	1.80	1.07	1.00
	moderate	90.57	95.01	4.45	1.02	38.66	1.00	4.46	1.07	1.00
	heavy	88.24	93.16	23.84	1.01	62.04	1.01	13.16	1.02	1.01

(b) Multivariate setup performances								
Type	Intensity	SB_dE	SD_mC_d	B_dD_dS	C_mD_mS	$B_dC_dD_d$	FSB_d	SAL
GLC	small	20.56	15.63	19.00	8.90	25.31	23.52	41.34
	moderate	2.22	1.82	2.25	1.66	6.28	2.39	10.61
	heavy	1.00	1.00	1.00	1.00	1.04	1.00	1.06
LLC	small	27.72	18.47	25.47	68.36	22.07	29.05	23.56
	moderate	4.66	3.49	4.07	13.81	2.92	4.76	2.94
	heavy	1.86	1.60	1.76	17.27	1.52	2.04	1.51
GNC	small	7.49	12.38	6.87	6.06	8.88	23.94	7.09
	moderate	1.48	1.50	1.34	1.17	1.61	5.04	1.44
	heavy	1.00	1.00	1.00	1.00	1.00	1.00	1.00
LNC	small	1.02	1.03	1.03	1.00	1.02	1.50	1.11
	moderate	1.02	1.05	1.04	1.11	1.13	1.81	1.10
	heavy	1.01	1.01	1.01	2.00	1.01	1.01	1.01

overview of the results for the univariate procedure and the representatives of the multivariate monitoring strategies. The results for all 84 triple combinations are given in Tables 8, 9, 10, 11, 12, 13 and 14 in the Appendix.

Overall, the results meet our expectations. The univariate metrics might perform reasonably in special scenarios, but show 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 flexibility compared to the univariate approach. On a further note, the results are handily interpretable and are aligned with the statements of our derived categorization of Sect. 3.1. We can, e.g., take a closer look at the performance of a representative of Category I, i.e., the balanced set SB_dE (consisting of S, B_d , and E). Regarding the involved univariate performances, S is able to handle global changes, but has problems with LLCs and especially LNCs. For B_d , 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 E, 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 of Categories IIa, IIb, IIIa, IIIb is similar as they provide more flexibility overall compared to the involved univariate candidates. Particularly, the ARL_1 for the “non-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 $B_dC_dD_d$ performance for GNCs or the C_mD_mS 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.

Table 6 SBM parameter 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	$K = 3, p_1 = 0.7, p_2 = 0.6,$ $p_3 = 0.3, p_{ij} = 0.1,$ $n_1 = 30, n_2 = 20,$ $n_3 = 50$	$K = 4, p_3 = p_4 = 0.49$ $n_3 = 30, n_4 = 20$

Lastly, the results for the performance of SAL underline our previous interpretations as this metric set achieves reasonable results for local changes, but struggles more in global change cases compared to other multivariate sets. This behavior underlines our classification into Category IIIb of Sect. 3.1. Furthermore, the additional consideration of the network density as a fourth metric compared to the set $B_dC_dD_d$ seems to not contribute relevant added value as the performances are quite similar.

4.4 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 practically more relevant application examples. For this purpose, we consider mixed-type changes (MTC) in the popular scenario of community changes in stochastic blockmodels (SBM) (Holland et al. 1983). The explicit setups can be found in Table 6, where K represents the number of communities, p_i the intra-group link probability of group $i \in \{1, \dots, K\}$, 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. 4 are different to those before. Whereas the nonparametric version clearly outperformed the parametric control charts in Sect. 4.2, the performances are more equal now. The nonparametric

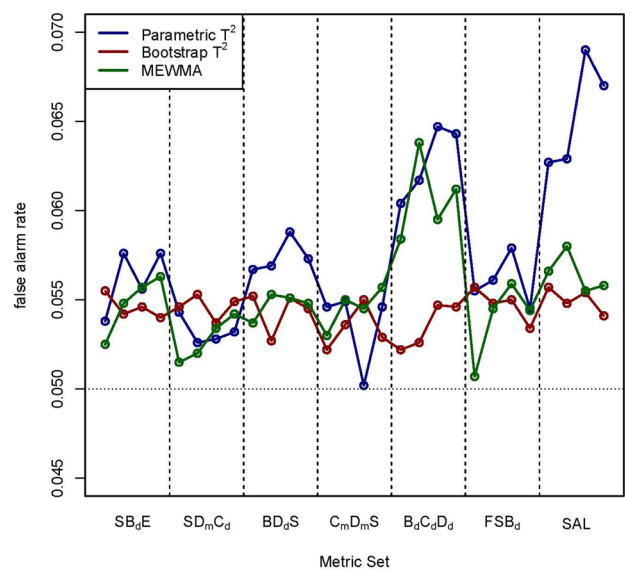


Fig. 4 Average empirical false alarm rates for all simulated situations

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 reasonably 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 nonparametric candidate. An explanation is the higher value of α , for which the charts are designed, as explained in Sect. 3.2. However, the sensitivity to the applied metric set still holds as the performance gets quite unstable for unbalanced sets.

The ARL_1 results in Table 7 can be interpreted similarly as before. Situation 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 reasonably, 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

Table 7 ARL_1 for the community changes. Best ARL_1 for each scenario across all setups is printed in bold

a. Univariate Setup Performances									
Case	F	S	C_m	C_d	D_m	D_d	B_m	B_d	E
1	1.00	1.00	1.00	19.32	1.00	10.63	1.00	1.49	1.06
2	20.34	15.71	17.90	18.59	18.34	22.77	19.42	12.37	1.62
3	1.01	3.02	1.00	2.55	1.00	6.13	1.00	5.89	1.28
4	25.60	17.33	10.61	16.46	25.60	17.93	10.44	18.50	20.03
b. Multivariate Setup Performances									
Case	SB_dE	SD_mC_d	B_dD_dS	C_mD_mS	$B_dC_dD_d$	FSB_d	SAL		
1	1.00	1.00	1.00	1.00	1.00	1.00	1.00		
2	1.87	5.00	9.87	5.12	14.58	4.77	13.75		
3	1.39	1.00	2.21	1.00	1.43	1.00	1.03		
4	8.94	17.18	14.14	5.12	10.74	15.25	10.92		

as before which results in an overall decreased link amount due to the decreased link probability of those nodes, which were in one group before and are in different ones afterward. Hence, it is a relatively easy GLC situation and all applied metrics achieve a satisfactory performance. We designed the last scenario such that the overall link probability 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.

5 Empirical data examples

To further underline the applicability and the handling of the approach, we analyze two real-world dynamic networks from economics and social sciences.

5.1 International trade data

This publicly available data set from the World Trade Organization (WTO) contains all reported international import–export relationships that are responsible for 90% of the overall worldwide trade volume. The data is collected quarterly and contains the time span of Q1 2010 - Q1 2022 which results in a dynamic network of $T = 49$ time points. For the selection of the in-control phase, we use the first 22 time points, i.e., the time span Q1 2010–Q2 2015. We are not aware of major trading conflicts in this time period and therefore expect stable behavior. As this in-control phase is rather short, we use the bootstrap Hotellings T^2 chart, because it promises to yield the most reliable results in these circumstances compared to the other charts as we saw in

the simulation study. We do not expect a particular change type, since changes triggered by nodes and links both in a global and in a local fashion are conceivable. Therefore, we use the balanced metric set SB_dE in order to increase the probability for the detection of arbitrary change types. The result is shown in the upper part of Fig. 5.

The procedure detects a change in Q2 2020 that is probably triggered by the corona pandemic. Many countries were forced to reduce their trading activities and mainly restrict them to neighboring states and the most important partner nations. This behavior pushed the trade network to a more centralized layout with leading trade nations

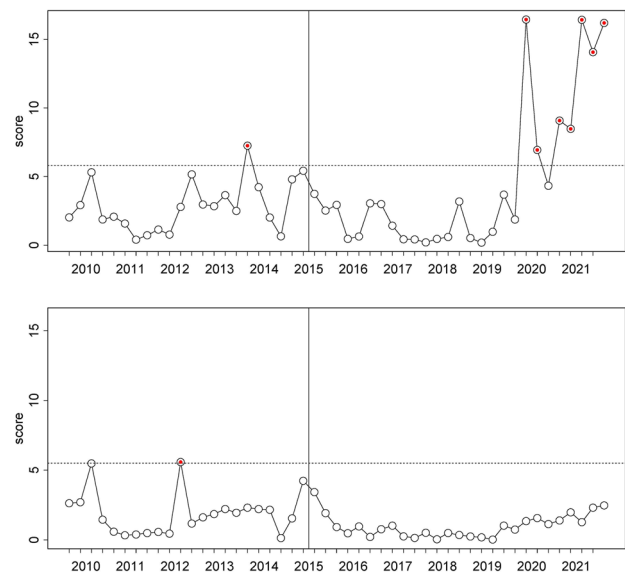


Fig. 5 Control charts using a balanced metric set (upper chart) and an unbalanced metric set (lower chart). In-control phase ends in Q2 2015. Red dots signalize alarms

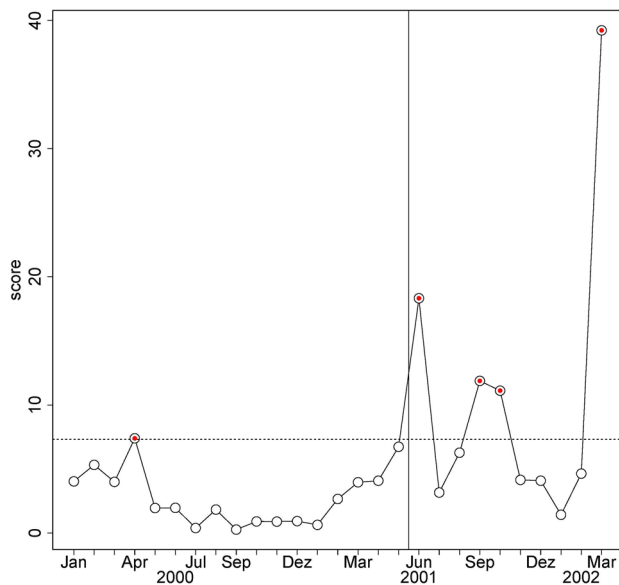


Fig. 6 Control Chart with SB_dE for the Enron data. In-control phase ends in May 2001. Red dots signalize alarms

as the hubs. Consequently, we can classify the observed change as a local change. Interestingly, we cannot detect global changes for this data as more unbalanced metric sets, that are focussed on global change types, do not trigger an alarm. This is visualized in the lower control chart of Fig. 5, for which we used the metric set $C_m D_m S$. Hence, the pandemic apparently did not have a huge influence on the amount of relationships that are responsible for 90% of the overall worldwide trade volume, but it did change the network on a structural level as explained. This example again underlines the importance of a careful execution of the procedure and the flexibility advantage of a balanced metric set.

5.2 Enron email data

This data contains the email communication network between employees of the former US company Enron that has been made public by the US Department of Justice and was rigorously analyzed in many publications (Priebe et al. 2005; Park et al. 2009; Chapanond et al. 2005). It comprises a time span from January 2000 until April 2002 on a monthly level. This results in $T = 28$ time points, from which we use the first 17 for the in-control phase and we aim to detect the effects of the accounting fraud scandal in the late 2001 s and early 2002 s. We again use the unbalanced metric set SB_dE in a bootstrap Hotellings T^2 chart in order to be flexible regarding the existence of different types of change. The output of the resulting control chart is depicted in Fig. 6.

The procedure triggers alarms that are in accordance with other social network analyses of this data set (Kendrick et al. 2018). In June 2001, the CEO Jeffrey Skilling tried to fire the chairman and former CEO Kenneth Lay, in September/October 2001 Enron reported a huge loss and it was announced that a SEC (Security and Exchange Commission) inquiry has become a formal investigation, and in February/March 2002 more and more details of the fraud were made public.

6 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 nonparametric 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 particularly recommend to use a balanced metric set like SB_dE together with a nonparametric control chart like the bootstrap Hotelling T^2 chart in order to achieve reliable results in flexible change situations. We further validated our statements with the help of a simulation study and some real-world examples in which a thoroughly designed multivariate approach outperforms the univariate procedure by offering a more flexible solution to the problem of change detection in social network analysis. As a main part of this paper was studying and evaluating different multivariate network metric sets and their capability of capturing relevant network information, the results can be beneficial for further statistical analysis procedures. In our view, this especially yields for statistical testing procedures (e.g., goodness-of-fit, two-sample tests) for network data where test statistics (i.e., metrics) need to be derived that characterize the network structure as comprehensively as possible. Furthermore, our results can be beneficial also for researchers that aim to derive network monitoring procedures for a certain application scenario, e.g., with pre-known network structures (like in model-based approaches). In this context, our results can be used as a benchmark method as it is particularly designed for detecting changes of more general form.

Appendix

On the next pages: Tables for the performance of all potential metric sets for Sect. 4.3.

Table 8 ARL₁ multivariate performances

Type	Intensity	FSC _m	FSC _d	FSB _m	FSB _d	FSD _m	FSD _d	FSE	FC _m C _d	FC _m B _m	FC _m B _d	FC _m D _m	FC _m D _d	FC _m E
GLC	Small	5.30	16.76	12.53	23.53	4.39	14.06	8.73	5.36	8.66	5.74	4.46	5.13	4.96
	Moderate	1.42	2.07	1.83	2.39	1.26	1.87	1.69	1.41	1.74	1.43	1.27	1.39	1.36
	Heavy	1.00	1.00	1.00	1.00	1.00	1.00	1.00	1.00	1.00	1.00	1.00	1.00	1.00
LLC	Small	34.63	18.69	74.03	29.05	13.74	17.94	16.28	28.43	19.79	44.36	17.51	29.25	34.47
	Moderate	12.51	3.41	44.51	4.76	20.30	3.79	5.28	4.42	15.99	5.32	16.79	4.76	6.56
	Heavy	15.05	1.60	25.55	2.04	31.21	1.74	1.99	1.84	8.44	2.38	45.68	2.00	3.13
GNC	small	1.48	3.74	1.81	23.94	2.60	2.99	2.47	13.70	1.36	7.62	1.36	6.64	4.17
	Moderate	1.00	1.00	1.00	1.04	1.00	1.00	1.00	1.12	1.00	1.13	1.00	1.06	1.04
	Heavy	1.00	1.00	1.00	1.00	1.00	1.00	1.00	1.00	1.00	1.00	1.00	1.00	1.00
LNC	Small	1.01	1.00	1.00	1.00	1.50	1.00	1.00	1.00	1.01	1.00	1.09	1.00	1.00
	Moderate	1.44	1.01	1.08	1.01	1.81	1.00	1.01	1.00	2.48	1.01	3.53	1.00	1.01
	Heavy	3.12	1.01	1.91	1.01	1.86	1.01	1.01	1.01	12.40	1.01	14.85	1.01	1.01

Table 9 ARL₁ multivariate performances

Type	Intensity	FC _d B _m	FC _d B _d	FC _d D _m	FC _d D _d	FC _d E	FB _m B _d	FB _m D _m	FB _m D _d	FB _m E	FB _d D _m	FB _d D _d	FB _d E	FD _m D _d
GLC	Small	16.04	30.75	4.38	35.88	17.99	15.71	5.33	14.72	14.45	4.39	28.35	25.61	4.23
	Moderate	1.91	2.51	1.26	10.72	2.02	2.10	1.39	1.84	1.83	1.26	2.18	2.21	1.26
	Heavy	1.00	1.00	1.00	1.11	1.00	1.00	1.00	1.00	1.00	1.00	1.00	1.00	1.00
LLC	Small	45.83	26.73	13.54	23.00	20.67	105.27	18.46	71.36	107.86	15.00	28.36	29.95	13.66
	Moderate	7.43	3.60	9.59	3.05	3.20	12.15	36.01	9.89	42.68	12.64	4.23	4.53	12.18
	Heavy	2.11	1.60	2.72	1.52	1.54	2.66	52.74	2.39	4.55	3.35	1.77	1.93	3.30
GNC	Small	2.82	6.72	3.96	11.79	4.37	3.14	1.38	2.49	2.17	3.83	5.47	7.65	2.84
	Moderate	1.00	1.15	1.00	1.59	1.03	1.00	1.00	1.00	1.00	1.03	1.15	1.78	1.00
	Heavy	1.00	1.00	1.00	1.00	1.00	1.00	1.00	1.00	1.00	1.00	1.00	1.00	1.00
LNC	Small	1.00	1.02	1.00	1.00	1.04	1.00	1.05	1.00	1.00	1.00	1.02	1.02	1.00
	Moderate	1.00	1.04	1.00	1.06	1.04	1.00	5.77	1.00	1.01	1.00	1.03	1.02	1.00
	Heavy	1.01	1.01	1.01	1.01	1.01	1.01	33.69	1.01	1.01	1.01	1.01	1.01	1.01

Table 10 ARL₁ multivariate performances

Type	Intensity	FD _m E	FD _d E	SC _m C _d	SC _m B _m	SC _m B _d	SC _m D _m	SC _m D _d	SC _m E	SC _d B _m	SC _d B _d	SC _d D _m	SC _d D _d	SC _d E
GLC	Small	4.30	13.64	13.05	4.37	16.80	8.90	11.86	8.82	17.46	22.59	15.63	32.38	14.90
	Moderate	1.26	1.85	1.74	1.33	1.83	1.66	1.69	1.54	1.87	2.31	1.82	8.48	2.11
	Heavy	1.00	1.00	1.00	1.11	1.00	1.00	1.00	1.00	1.00	1.00	1.00	1.02	1.00
LLC	Small	14.40	18.96	24.54	11.48	45.51	68.36	27.85	29.39	23.02	24.58	18.47	22.99	17.41
	Moderate	19.61	3.32	3.50	8.22	4.59	13.81	3.96	6.19	3.49	3.52	3.49	3.00	3.16
	Heavy	14.55	1.67	1.68	6.38	2.12	17.27	1.82	2.30	1.76	1.62	1.60	1.49	1.52
GNC	Small	2.51	2.91	16.80	1.36	8.75	6.06	8.08	5.68	2.81	7.76	12.38	14.03	5.84
	Moderate	1.00	1.00	1.48	1.00	1.37	1.17	1.24	1.14	1.00	1.52	1.50	2.00	1.18
	Heavy	1.00	1.00	1.00	1.00	1.00	1.00	1.00	1.00	1.00	1.00	1.00	1.00	1.00
LNC	Small	1.00	1.04	1.00	1.41	1.00	1.00	1.00	1.00	1.00	1.03	1.00	1.00	1.06
	Moderate	1.01	1.02	1.00	12.48	1.00	1.11	1.00	1.02	1.00	1.05	1.03	1.06	1.06
	Heavy	1.01	1.01	1.01	49.96	1.01	2.00	1.01	1.01	1.01	1.01	1.01	1.01	1.01

Table 11 ARL₁ multivariate performances

Type	Intensity	SB _m B _d	SB _m D _m	SB _m D _d	SB _m E	SB _d D _m	SB _d D _d	SB _d E	SD _m D _d	SD _m E	SD _d E	C _m C _d B _m	C _m C _d B _d	C _m C _d D _m
GLC	Small	28.69	99.70	15.27	10.08	22.25	19.00	20.56	13.13	9.47	12.23	5.86	20.42	6.75
	Moderate	2.29	12.37	1.84	1.64	2.15	2.25	2.22	1.79	1.60	1.85	1.59	2.08	1.66
	Heavy	1.00	1.01	1.00	1.11	1.00	1.00	1.00	1.00	1.00	1.00	1.00	1.00	1.00
LLC	Small	43.46	100.17	27.04	27.51	31.81	25.47	27.72	17.70	16.64	16.29	13.68	23.80	36.03
	Moderate	4.76	135.18	3.98	5.93	4.69	4.07	4.66	3.69	5.20	3.27	3.66	3.52	3.61
	Heavy	2.13	32.76	1.82	2.27	2.04	1.76	1.86	1.72	1.97	1.65	1.79	1.62	1.83
GNC	Small	2.84	2.26	2.40	2.00	8.39	6.87	7.49	9.93	8.40	3.43	1.53	8.85	16.30
	Moderate	1.00	1.00	1.00	1.00	1.57	1.34	1.48	1.42	1.34	1.04	1.00	1.62	1.47
	Heavy	1.00	1.00	1.00	1.00	1.00	1.00	1.00	1.00	1.00	1.00	1.00	1.00	1.00
LNC	Small	1.00	1.00	1.00	1.00	1.01	1.03	1.02	1.00	1.00	1.06	1.00	1.00	1.00
	Moderate	1.00	1.08	1.00	1.01	1.02	1.04	1.02	1.01	1.01	1.02	1.00	1.01	1.00
	Heavy	1.01	1.81	1.01	1.01	1.01	1.01	1.01	1.01	1.01	1.01	1.01	1.01	1.01

Table 12 ARL₁ multivariate performances

Type	Intensity	C _m C _d D _d	C _m C _d E	C _m B _m B _d	C _m B _m D _m	C _m B _m D _d	C _m B _m E	C _m B _d D _m	C _m B _d D _d	C _m B _d E	C _m D _m D _d	C _m D _m E
GLC	Small	33.35	12.82	7.21	25.38	5.74	5.52	8.75	17.13	15.08	6.26	5.99
	Moderate	8.33	1.88	1.63	2.76	1.47	1.44	1.74	1.90	1.81	1.56	1.52
	Heavy	1.02	1.01	1.00	1.11	1.00	1.00	1.00	1.00	1.00	1.00	1.00
LLC	Small	22.94	16.79	18.21	23.52	14.35	12.59	81.27	26.80	26.55	39.48	63.41
	Moderate	3.03	3.13	4.86	24.81	4.08	5.23	4.95	4.07	4.45	4.13	5.05
	Heavy	1.52	1.54	2.10	8.61	1.87	2.49	2.53	1.77	1.94	2.01	3.17
GNC	Small	12.60	5.76	1.60	1.36	1.52	1.48	8.90	7.79	8.61	8.29	5.84
	Moderate	2.85	1.26	1.00	1.00	1.00	1.00	1.39	1.38	1.49	1.24	1.14
	Heavy	1.00	1.00	1.00	1.00	1.00	1.00	1.00	1.00	1.00	1.00	1.00
LNC	Small	1.16	1.00	1.00	1.02	1.01	1.03	1.02	1.00	1.00	1.06	1.00
	Moderate	1.08	1.00	1.00	2.66	1.00	1.01	1.01	1.01	1.07	1.00	1.02
	Heavy	1.01	1.01	1.01	12.31	1.01	1.01	1.01	1.01	1.01	1.01	1.01

Table 13 ARL₁ multivariate performances

Type	Intensity	C _m D _d E	C _d B _m B _d	C _d B _m D _m	C _d B _m D _d	C _d B _m E	C _d B _d D _m	C _d B _d D _d	C _d B _d E	C _d D _m D _d	C _d D _m E	C _d D _d E
GLC	Small	8.94	23.41	104.23	37.26	14.70	23.39	25.31	43.36	36.64	14.09	19.58
	Moderate	1.72	2.20	13.19	9.82	2.00	2.09	6.28	9.33	9.71	2.00	4.58
	Heavy	1.00	1.00	1.01	1.07	1.00	1.00	1.03	1.13	1.05	1.00	1.01
LLC	Small	17.02	27.48	47.80	22.92	20.63	24.81	22.07	27.40	22.76	18.66	17.89
	Moderate	3.27	3.59	10.05	3.03	3.19	3.59	2.92	3.29	3.03	3.19	2.99
	Heavy	1.62	1.61	2.06	1.52	1.57	1.59	1.52	1.60	1.52	1.54	1.50
GNC	Small	4.12	4.46	2.73	4.06	4.45	7.73	8.88	5.81	13.90	5.48	4.09
	Moderate	1.07	1.05	1.00	1.08	1.04	1.56	1.61	1.16	2.00	1.18	1.12
	Heavy	1.00	1.00	1.00	1.00	1.00	1.00	1.00	1.00	1.00	1.00	1.00
LNC	Small	1.00	1.00	1.00	1.00	1.00	1.04	1.02	1.07	1.00	1.07	1.00
	Moderate	1.00	1.00	1.00	1.05	1.03	1.05	1.13	1.07	1.06	1.06	1.08
	Heavy	1.01	1.01	1.01	1.01	1.01	1.01	1.01	1.01	1.01	1.01	1.01

Table 14 ARL₁ multivariate performances

Type	Intensity	B _m B _d D _m	B _m B _d D _d	B _m B _d E	B _m D _m D _d	B _m D _m E	B _m D _d E	B _d D _m D _d	B _d D _m E	B _d D _d E	D _m D _d E
GLC	Small	109.54	23.02	20.11	118.34	147.58	11.09	21.38	19.62	103.25	10.65
	Moderate	17.90	2.03	1.94	13.49	13.14	1.77	2.03	1.94	23.57	1.77
	Heavy	1.01	1.00	1.00	1.01	1.01	1.00	1.00	1.00	1.33	1.00
LLC	Small	102.06	29.13	30.14	66.03	113.64	19.25	26.42	26.96	35.14	17.40
	Moderate	17.35	4.21	4.55	17.86	106.32	3.29	4.16	4.50	4.22	3.27
	Heavy	2.67	1.81	1.94	2.35	4.70	1.69	1.77	1.93	1.82	1.67
GNC	Small	2.88	4.28	3.68	2.34	2.01	3.35	6.80	8.04	3.96	3.36
	Moderate	1.00	1.02	1.02	1.00	1.00	1.01	1.33	1.44	1.03	1.044
	Heavy	1.00	1.00	1.00	1.00	1.00	1.00	1.00	1.00	1.00	1.00
LNC	Small	1.00	1.00	1.00	1.00	1.00	1.00	1.03	1.03	1.06	1.04
	Moderate	1.00	1.00	1.00	1.00	1.01	1.01	1.04	1.02	1.06	1.02
	Heavy	1.01	1.01	1.01	1.01	1.01	1.01	1.01	1.01	1.01	1.01

Acknowledgements This research was financially supported by the Mercator Research Center Ruhr (MERCUR) with project number PR-2019-0019.

Author contributions Conceptualization: J.F, R.F., C.J. Methodology: J.F., C.J. Writing: J.F. Simulation Study: J.F. Reviewing of Manuscript: J.F., R.F., C.J.

Funding Open Access funding enabled and organized by Projekt DEAL.

Declarations

Conflict of interest The authors declare that there is no conflict of interest.

Open Access This article is licensed under a Creative Commons Attribution 4.0 International License, which permits use, sharing, adaptation, distribution and reproduction in any medium or format, as long as you give appropriate credit to the original author(s) and the source, provide a link to the Creative Commons licence, and indicate if changes were made. The images or other third party material in this article are included in the article's Creative Commons licence, unless indicated otherwise in a credit line to the material. If material is not included in the article's Creative Commons licence and your intended use is not permitted by statutory regulation or exceeds the permitted use, you will need to obtain permission directly from the copyright holder. To view a copy of this licence, visit <http://creativecommons.org/licenses/by/4.0/>.

References

Barnett I, Onnela JP (2016) Change point detection in correlation networks. *Sci Rep* 6(1):1–11
 Bassett DS, Sporns O (2017) Network neuroscience. *Nature Neurosci* 20(3):353–364
 Basseville M, Nikiforov IV (1993) *Detection of abrupt changes: theory and application*, Volume 104. Prentice Hall Englewood Cliffs
 Belkin M, Niyogi P (2001) Laplacian eigenmaps and spectral techniques for embedding and clustering. *Adv Neural Inform Process Syst* 14:102

Bunke H, Dickinson PJ, Kraetzl M, Wallis WD (2007) A graph-theoretic approach to enterprise network dynamics. Springer, London, p 24
 Carrington PJ, Scott J, Wasserman S (2005) Models and methods in social network analysis. Cambridge University Press, Cambridge, p 28
 Chapanond A, Krishnamoorthy MS, Yener B (2005) Graph theoretic and spectral analysis of enron email data. *Comput Math Organizat Theory* 11(3):265–281
 Chen L, Zhou J, Lin L (2021) Hypothesis testing for populations of networks. *Commun Statist Theory Methods* 7:1–24
 Cui P, Wang X, Pei J, Zhu W (2018) A survey on network embedding. *IEEE Trans Knowl Data Eng* 31(5):833–852
 Duan, D, Tong L, Li Y, Lu J, Shi L, Zhang C (2020) Aane: Anomaly aware network embedding for anomalous link detection. In: 2020 IEEE International Conference on Data Mining (ICDM), pp. 1002–1007. IEEE
 Durante D, Dunson DB (2014) Bayesian dynamic financial networks with time-varying predictors. *Statist Probab Lett* 93:19–26
 Durante D, Dunson DB, Vogelstein JT (2017) Nonparametric bayes modeling of populations of networks. *J Am Statist Associat* 112(520):1516–1530
 Farahani EM, Baradaran Kazemzadeh R, Noorossana R, Rahimian G (2017) A statistical approach to social network monitoring. *Commun Statist Theory Methods* 46(22):11272–11288
 Flossdorf J, Jentsch C (2021) Change detection in dynamic networks using network characteristics. *IEEE Trans Signal Inform Process Over Netw* 7:451–464
 Grattarola D, Zambon D, Livi L, Alippi C (2019) Change detection in graph streams by learning graph embeddings on constant-curvature manifolds. *IEEE Trans Neural Netw Learn Syst* 31(6):1856–1869
 Grover A, Leskovec J (2016) node2vec: Scalable feature learning for networks. In *Proceedings of the 22nd ACM SIGKDD international conference on Knowledge discovery and data mining*, pp. 855–864
 Gürsoy F, Haddad M, Bothorel C (2021) Alignment and stability of embeddings: measurement and inference improvement. arXiv preprint [arXiv:2101.07251](https://arxiv.org/abs/2101.07251)
 Haddad M, Bothorel C, Lenca P, Bedart D (2020) Temporalnode2vec: Temporal node embedding in temporal networks. In *Complex Networks and Their Applications VIII: Volume 1 Proceedings of the Eighth International Conference on Complex Networks and Their Applications COMPLEX NETWORKS 2019* 8, pp. 891–902. Springer

- Hazrati-Marangaloo H, Noorossana R (2021) A nonparametric change detection approach in social networks. *Qual Reliabil Eng Int* 37(6):2916–2935
- Hewapathirana IU (2019) Change detection in dynamic attributed networks. *Int Rev Data Min Knowl Disc* 9(3):1286–1306
- Hewapathirana IU, Lee D, Moltchanova E, McLeod J (2020) Change detection in noisy dynamic networks: a spectral embedding approach. *Soc Netw Anal Min* 10:1–22
- Holland PW, Laskey KB, Leinhardt S (1983) Stochastic blockmodels: first steps. *Soc Netw* 5(2):109–137
- Kendrick L, Musial K, Gabrys B (2018) Change point detection in social networks: critical review with experiments. *Comput Sci Rev* 29:1–13
- Knoth S (2017) Arl numerics for mewma charts. *J Qual Technol* 49(1):78–89
- Koutra D, Shah N, Vogelstein JT, Gallagher B, Faloutsos C (2016) Deltacon: Principled massive-graph similarity function with attribution. *ACM Trans Knowl Disc Data (TKDD)* 10(3):1–43
- Lee DH, Dong M (2009) Dynamic network design for reverse logistics operations under uncertainty. *Trans Res Part E logist Trans Rev* 45(1):61–71
- Li Y, Lu A, Wu X, Yuan S (2019) Dynamic anomaly detection using vector autoregressive model. In *Advances in Knowledge Discovery and Data Mining: 23rd Pacific-Asia Conference, PAKDD 2019, Macau, China, April 14–17, 2019, Proceedings, Part I 23*, pp. 600–611. Springer
- Lin Ch, Xu L, Yamanishi K (2022) Network change detection based on random walk in latent space. *IEEE Trans Knowl Data Eng.* <https://doi.org/10.1109/TKDE.2022.3167062>
- Lowry CA, Woodall WH, Champ CW, Rigdon SE (1992) A multivariate exponentially weighted moving average control chart. *Technometrics* 34(1):46–53
- Malinovskaya A, Otto P (2021) Online network monitoring. *Statist Methods Appl* 30(5):1337–1364
- McCulloh I, Carley KM (2011) Detecting change in longitudinal social networks. *Military Acad West Point NY Network Sci Cent (NSC)* 12:1–37
- Montgomery DC (2012) *Statistical quality control*. Wiley Global Education
- Motalebi N, Owlia MS, Amiri A, Fallahnezhad MS (2021) Monitoring social networks based on zero-inflated poisson regression model. *Commun Statist Theory Methods* 12:1–17
- Neil J, Hash C, Brugh A, Fisk M, Storlie CB (2013) Scan statistics for the online detection of locally anomalous subgraphs. *Technometrics* 55(4):403–414
- Ofori-Boateng D, Gel YR, Cribben I (2021) Nonparametric anomaly detection on time series of graphs. *J Computat Graph Stat* 30(3):756–767
- Ou M, Cui P, Pei J, Zhang Z, Zhu W (2016) Asymmetric transitivity preserving graph embedding. In: *Proceedings of the 22nd ACM SIGKDD international conference on Knowledge discovery and data mining*, pp. 1105–1114
- Park Y, Priebe C, Marchette D, Youssef A (2009) Anomaly detection using scan statistics on time series hypergraphs. In: *Link Analysis, Counterterrorism and Security (LACTS) Conference*, pp. 9. SIAM Pennsylvania
- Perozzi B, Al-Rfou R, Skiena S (2014) Deepwalk: Online learning of social representations. In: *Proceedings of the 20th ACM SIGKDD international conference on Knowledge discovery and data mining*, pp. 701–710
- Phaladiganon P, Kim SB, Chen VC, Baek JG, Park SK (2011) Bootstrap-based t2 multivariate control charts. *Commun Stat Simulat Comput* 40(5):645–662
- Pincombe B (2005) Anomaly detection in time series of graphs using arma processes. *Asor Bull* 24(4):2
- Prabhu SS, Runger GC (1997) Designing a multivariate ewma control chart. *J Qual Technol* 29(1):8–15
- Priebe CE, Conroy JM, Marchette DJ, Park Y (2005) Scan statistics on enron graphs. *Computat Math Organiz Theory* 11(3):229–247
- Prill RJ, Iglesias PA, Levchenko A (2005) Dynamic properties of network motifs contribute to biological network organization. *PLoS Biol* 3(11):e343
- Ranshous S, Shen S, Koutra D, Harenberg S, Faloutsos C, Samatova NF (2015) Anomaly detection in dynamic networks: a survey. *Int Rev Computat Statist* 7(3):223–247
- Salmasnia A, Mohabbati M, Namdar M (2020) Change point detection in social networks using a multivariate exponentially weighted moving average chart. *J Inform Sci* 46(6):790–809
- Sarkar P, Moore AW (2005) Dynamic social network analysis using latent space models. *ACM Sigkdd Explorat Newsl* 7(2):31–40
- Stoumbos ZG, Sullivan JH (2002) Robustness to non-normality of the multivariate EWMA control chart. *J Qual Technol* 34(3):260–276
- Sun J, Tao D, Faloutsos C (2006) Beyond streams and graphs: dynamic tensor analysis. In: *Proceedings of the 12th ACM SIGKDD international conference on Knowledge discovery and data mining*, pp. 374–383
- Sun T, Liu Y (2018) A dynamic network change detection method using network embedding. In: *Cloud Computing and Security: 4th International Conference, ICCCS 2018, Haikou, China, June 8–10, 2018, Revised Selected Papers, Part I 4*, pp. 63–74. Springer
- Wilson JD, Stevens NT, Woodall WH (2019) Modeling and detecting change in temporal networks via the degree corrected stochastic block model. *Qual Reliabil Eng Int* 35(5):1363–1378
- Xie Y, Wang W, Shao M, Li T, Yu Y (2023) Multi-view change point detection in dynamic networks. *Inform Sci* 629:344–357
- Zhou L, Yang Y, Ren X, Wu F, Zhuang Y (2018) Dynamic network embedding by modeling triadic closure process. In: *Proceedings of the AAAI conference on artificial intelligence, Volume 32*
- Zou N, Li J (2017) Modeling and change detection of dynamic network data by a network state space model. *IISE Trans* 49(1):45–57

Publisher's Note Springer Nature remains neutral with regard to jurisdictional claims in published maps and institutional affiliations.