ORIGINAL PAPER

Online network monitoring

Anna Malinovskaya[1] · Philipp Otto[1]

© The Author(s) 2021

Abstract

An important problem in network analysis is the online detection of anomalous behaviour. In this paper, we introduce a network surveillance method bringing together network modelling and statistical process control. Our approach is to apply multivariate control charts based on exponential smoothing and cumulative sums in order to monitor networks generated by temporal exponential random graph models (TERGM). The latter allows us to account for temporal dependence while simultaneously reducing the number of parameters to be monitored. The performance of the considered charts is evaluated by calculating the average run length and the conditional expected delay for both simulated and real data. To justify the decision of using the TERGM to describe network data, some measures of goodness of fit are inspected. We demonstrate the effectiveness of the proposed approach by an empirical application, monitoring daily flights in the United States to detect anomalous patterns.

Keywords MCUSUM · MEWMA · Multivariate Control Charts · Network Modelling · Network Monitoring · Statistical Process Control · TERGM

1 Introduction

The digital information revolution offers a rich opportunity for scientific progress; however, the amount and variety of data available requires new analysis techniques for data mining, interpretation and application of results to deal with the growing complexity. As a consequence, these requirements have influenced the development of networks, bringing their analysis beyond the traditional sociological scope into

✉ Anna Malinovskaya
anna.malinovskaya@ikg.uni-hannover.de

Philipp Otto
philipp.otto@ikg.uni-hannover.de

1   Leibniz University Hannover, Hannover, Germany

many other disciplines, as varied as are physics, biology and statistics (cf. Amaral et al. 2000; Simpson et al. 2013; Chen et al. 2019).

An important problem in network study is the detection of anomalous behaviour. There are two types of network monitoring, differing in the treatment of nodes and links: fixed and random network surveillance (cf. Leitch et al. 2019). In this work, we concentrate on the modelling and monitoring of networks with randomly generated edges across time, describing a surveillance method of the second type. When talking about anomalies in temporal networks, the major interest is to find the point of time when a significant change happened and, if appropriate, to identify the vertices, edges or subgraphs which considerably contributed to the change (cf. Akoglu et al. 2014). Further differentiation depends on at least two factors: characteristics of the network data and available time granularity. Hence, given a particular network to monitor it is worth first defining what is classified as "anomalous".

To analyse the network data effectively, it is important to account for its complex structure and the possibly high computational costs. Our approach to mitigate these issues and simultaneously reflect the stochastic and dynamic nature of networks is to model them applying a temporal random graph model. We consider a general class of Exponential Random Graph Models (ERGM) (cf. Frank and Strauss 1986; Robins et al. 2007; Schweinberger et al. 2020), which was originally designed for modelling cross-sectional networks. This class includes many prominent random network configurations such as dyadic independence models and Markov random graphs, enabling the ERGM to be generally applicable to many types of complex networks. Furthermore, if a network has many covariates (also called attributes), i.e., variables that provide additional information about the graph's edges and nodes, it can be both computationally and theoretically challenging to recognise meaningful patterns. In this case, it is beneficial to apply ERGM that facilitate data reduction by summarising the network in the form of sufficient statistics, i.e., network statistics that capture the relevant features of the network. Knowing the observed values of sufficient statistics, we can derive the model parameters from the network data and make inferences.

Hanneke et al. (2010) developed a dynamic extension based on ERGM which is known as the Temporal Exponential Random Graph Model (TERGM). These models contain the overall functionality of the ERGM, while additionally enabling time-dependent covariates. Thus, our monitoring procedure for this class of models allows for many applications in different disciplines which are interested in analysing networks of medium sizes, such as sociology, political science, engineering, economics and psychology (cf. Carrington et al. 2005; Ward et al. 2011; Das et al. 2013; Jackson 2015; Fonseca-Pedrero 2018).

In the field of change detection, according to Basseville and Nikiforov (1993), there are three classes of problems: online detection of a change, off-line hypotheses testing and off-line estimation of the change time. Our method refers to the first class, meaning that the change point should be detected as soon as possible after it occurred. In this case, real-time monitoring of complex structures becomes necessary: for instance, if the network is observed every minute, the monitoring procedure should be faster than one minute. To perform online surveillance for real-

time detection, the efficient way is to use tools from the field of Statistical Process Control (SPC). SPC corresponds to an ensemble of analytical tools originally developed for industrial purposes, which is applied for the achievement of process stability and variability reduction (cf. Montgomery 2009).

One of the leading SPC tools is a control chart, which exists in various forms in terms of the number of variables, data type and statistics being of interest. For example, the monitoring of network topology statistics applying the Cumulative Sum (CUSUM) chart and illustrating its effectiveness on the analysis of military networks was presented by McCulloh and Carley (2011). Wilson et al. (2019) used the dynamic Degree-Corrected Stochastic Block Model (DCSBM) to generate the networks and then performed surveillance over the Maximum Likelihood (ML) estimates using the Shewhart and Exponentially Weighted Moving Average (EWMA) charts. One of the possibilities to bring together the ERGM in form of a Markov Graph and EWMA and Hotelling's T^2 charts was proposed by Sadinejad et al. (2020). Farahani et al. (2017) evaluate the combination of Multivariate EWMA (MEWMA) and Multivariate CUSUM (MCUSUM) applied with the Poisson regression model for monitoring social networks. Hosseini and Noorossana (2018) apply EWMA and CUSUM to degree measures for detecting outbreaks on a weighted undirected network. The distribution-free MCUSUM is introduced by Liu et al. (2019) to analyse longitudinal networks. Salmasnia et al. (2019) present a comparative study of univariate and multivariate EWMA for social network monitoring. An overview of further studies is provided by Noorossana et al. (2018).

In this paper, we present an online monitoring procedure based on SPC, which enables one to detect significant changes in the network structure in real time. The foundations of this approach, together with the description of the selected network model and multivariate control charts, are discussed in Sect. 2. Section 3 outlines the simulation study and includes performance evaluation of the designed control charts. In Sect. 4 we monitor daily flights in the United States and explain the detected anomalies. We conclude with a discussion of outcomes and present several directions for future research.

2 Network monitoring

Network monitoring is a form of online surveillance procedure to detect deviations from a so-called in-control state, i.e., the state when no unaccountable variation of the process is present. This is done by sequential hypothesis testing over time, which has a strong connection to control charts. In other words, the purpose of control charting is to identify occurrences of unusual deviation of the observed process from a prespecified target (or in-control) process, distinguishing common from special causes of variation (cf. Johnson and Wichern 2007). To be precise, the aim is to test the null hypothesis

$H_{0,t}$: The network observed at time point t is in its target state

against the alternative

$H_{1,t}$: The network observed at time point t deviates from its target state.

In this work, we concentrate on the monitoring of networks, which are modelled by the TERGM that is briefly described below.

2.1 Network modelling

A network (also interchangeably called "graph") can be presented by its adjacency matrix $\boldsymbol{Y} := (Y_{ij})_{i,j=1,\ldots,N}$, where N represents the total number of nodes. Two vertices (or nodes) i, j are adjacent if they are connected by an edge (also called a tie or link). In this case, $Y_{ij} = 1$, otherwise, $Y_{ij} = 0$. If the network is undirected, the adjacency matrix $\boldsymbol{Y}$ is symmetric. The connections of a node with itself are mostly not relevant to the majority of the networks, therefore, we assume that $Y_{ii} = 0$ for all $i = 1, \ldots, N$. Formally, we define a network model as a collection $\{\mathbb{P}_\theta(\boldsymbol{Y}), \boldsymbol{Y} \in \mathscr{Y} : \boldsymbol{\theta} \in \Theta\}$, where $\mathscr{Y}$ denotes the ensemble of possible networks, $\mathbb{P}_\theta$ is a probability distribution on $\mathscr{Y}$ and $\boldsymbol{\theta}$ is a vector of parameters, ranging over possible values in a subset of p-dimensional Euclidean space $\Theta \subseteq \mathbb{R}^p$ with $p \in \mathbb{N}$ (Kolaczyk 2009). In case of a directed graph, where the edges have a direction assigned to them, this stochastic mechanism determines which of the $N(N-1)$ edges are present, i.e., it assigns probabilities to each of the $2^{N(N-1)}$ graphs (cf. Cannings and Penman 2003).

The ERGM functional representation is given by

$$\mathrm{P}_{\boldsymbol{\theta}}(\boldsymbol{Y}) = \frac{\exp[\boldsymbol{\theta}'\boldsymbol{s}(\boldsymbol{Y})]}{c(\boldsymbol{\theta})}, \tag{1}$$

where $\boldsymbol{Y}$ is the adjacency matrix of an observed graph with $\boldsymbol{s} : \mathscr{Y} \to \mathbb{R}^p$ being a p-dimensional statistic describing the essential properties of a network based on $\boldsymbol{Y}$ (cf. Frank 1991; Wasserman and Pattison 1996). The network terms $\boldsymbol{s}(\boldsymbol{Y})$ are sufficient statistics which summarise the network data, so that the inference about the model parameters $\boldsymbol{\theta}$ depends on the graph data $\boldsymbol{Y}$ only through the values $\boldsymbol{s}(\boldsymbol{Y})$. There are several types of network terms, including dyadic dependent terms, for example, a statistic capturing transitivity, and dyadic independent terms, for instance, a term describing graph density (Morris et al. 2008). It is also possible to include nodal and edge attributes into the statistics, whose variety depends on the type of network. Although the overall concept presented in this work is valid for both graph types, we explicitly consider directed graphs from now on.

The model parameters $\boldsymbol{\theta}$ can be defined as respective coefficients of $\boldsymbol{s}(\boldsymbol{Y})$ which are of considerable interest in understanding the structural properties of a network. They reflect, on the network level, the tendency of a graph to exhibit certain sub-structures relative to what would be expected from a model by chance, or, on the tie level, the probability to observe a specific edge, given the rest of the graph (Block et al. 2018). The last interpretation follows from the representation of the problem as a log-odds ratio. The normalising constant in the denominator ensures that the sum of probabilities is equal to one, meaning it includes all possible network configurations $c(\boldsymbol{\theta}) = \sum_{\boldsymbol{Y} \in \mathscr{Y}}[\exp \boldsymbol{\theta}'\boldsymbol{s}(\boldsymbol{Y})]$ in the ensemble $\mathscr{Y}$.

In dynamic network modelling, a random sequence of Y_t for $t = 1, 2, \ldots$ with $Y_t \in \mathcal{Y}$ defines a stochastic process for all t. Unlike the relational event models, where the edges have no duration (cf. Butts 2008), in this work, we contemplate edges with duration. To conduct surveillance over Y_t, we propose to consider only the dynamically estimated characteristics of a graph in order to reduce computational complexity and to allow for real-time monitoring. In most of the cases, the dynamic network models serve as an extension of well-known static models. Similarly, the discrete temporal expansion of the ERGM is known as TERGM (cf. Hanneke et al. 2010) and can be seen as further advancement of a family of network models proposed by Robins and Pattison (2001).

The TERGM defines the probability of a network at the discrete time point t both as a function of subgraph counts in t and by including the network terms based on the previous graph observations until the particular time point $t - v$. That is

$$\mathrm{P}_{\boldsymbol{\theta}}(\boldsymbol{Y}_t | \boldsymbol{Y}_{t-1}, \ldots, \boldsymbol{Y}_{t-v}, \boldsymbol{\theta}) = \frac{\exp[\boldsymbol{\theta}' \boldsymbol{s}(\boldsymbol{Y}_t, \boldsymbol{Y}_{t-1}, \ldots, \boldsymbol{Y}_{t-v})]}{c(\boldsymbol{\theta}, \boldsymbol{Y}_{t-1}, \ldots, \boldsymbol{Y}_{t-v})}, \tag{2}$$

where v represents the maximum temporal lag, capturing the networks which are incorporated into the $\boldsymbol{\theta}$ estimation at t, hence, defining the complete temporal dependence of $\boldsymbol{Y}_t$ that corresponds to the Markov structure of order $v \in \mathbb{N}$ (Hanneke et al. 2010). In Sects. 3 and 4, we assume $v = 1$, leading to $(\boldsymbol{Y}_t \perp\!\!\!\perp \{\boldsymbol{Y}_1, \ldots, \boldsymbol{Y}_{t-2}\} | \boldsymbol{Y}_{t-1})$, where $\perp\!\!\!\perp$ defines conditional independence.

To model the joint probability of z networks between the time stamps $v + 1$ and $v + z$, we define $\mathrm{P}_{\boldsymbol{\theta}}$ based on the conditional independence assumption as

$$\mathrm{P}_{\boldsymbol{\theta}}(\boldsymbol{Y}_{v+1}, \ldots, \boldsymbol{Y}_{v+z} | \boldsymbol{Y}_1, \ldots, \boldsymbol{Y}_v, \boldsymbol{\theta}) = \prod_{t=v+1}^{v+z} \mathrm{P}_{\boldsymbol{\theta}}(\boldsymbol{Y}_t | \boldsymbol{Y}_{t-1}, \ldots, \boldsymbol{Y}_{t-v}, \boldsymbol{\theta}). \tag{3}$$

Regarding the network statistics in the TERGM, $\boldsymbol{s}(\cdot)$ includes "memory terms" such as dyadic stability or reciprocity (Leifeld et al. 2018). To distinguish the processes leading to the dissolution and formation of links, Krivitsky and Handcock (2014) presented Seperable TERGM (STERGM). To be precise, the STERGM is a subclass of the TERGM class, which can reproduce any transition process captured by the parameters $\boldsymbol{\theta} = (\boldsymbol{\theta}^+, \boldsymbol{\theta}^-)$ and the network terms $\boldsymbol{s} = (\boldsymbol{s}^+, \boldsymbol{s}^-)$, where $\boldsymbol{\theta}^+$ and $\boldsymbol{s}^+$ belong to the formation model, $\boldsymbol{\theta}^-$ and $\boldsymbol{s}^-$ to the dissolution model.

The careful selection of the network statistics is relevant from several points of view. First of all, under the ML estimation, the expected value of the network statistics is equal to the observed value, i.e., $\mathbb{E}_{\boldsymbol{\theta}}(\boldsymbol{s}(\boldsymbol{Y})) = \boldsymbol{s}(\boldsymbol{Y}_{obs})$ (cf. van Duijn et al. 2009). To be precise, on average, we reproduce the observed network $\boldsymbol{Y}_{obs}$ in terms of the sufficient statistics $\boldsymbol{s}(\boldsymbol{Y})$. Second, the selected network statistics determine our understanding of the network formation, combining our knowledge about the important terms to recover the graph structure with the interest of including additional statistics for monitoring. The dimension of the sufficient statistics can differ over time, however, we assume that in each time stamp t we have the same configuration $\boldsymbol{s}(\cdot)$. In general, the selection of terms extensively depends on the field and context, although the statistical modelling standards such as avoidance of linear dependencies among the terms should be also considered

(Morris et al. 2008). It is also helpful to perform goodness of fit tests, which enable one to find a compromise between the model's complexity and its explanatory power.

An improper selection of the network terms can often lead to a near-degenerate model, resulting in algorithmic issues and lack of fit (cf. Handcock 2003; Schweinberger 2011). In this case, as well as fine-tuning the configuration of statistics, one can modify some settings which design the estimation procedure of the model parameters. Considering the Markov Chain Monte Carlo (MCMC) ML estimation, for example, the run time, the sample size or the step length could be adjusted (Morris et al. 2008). Another possible improvement would be to add some stable statistics such as Geometrically-Weighted Edgewise Shared Partnerships (GWESP) (Snijders et al. 2006). However, the TERGM is less prone to degeneracy issues compared to the ERGM as ascertained by Hanneke et al. (2010) and Leifeld and Cranmer (2019). Overall, we assume that most of the network surveillance studies can reliably estimate beforehand the type of anomalies which are possible to occur. This assumption guides the choice of terms in the models throughout the work.

2.2 Monitoring process

Although the monitoring procedure can be constructed by supervising $\boldsymbol{Y}_t$ directly, this approach is likely to become computationally intricate as it depends on the order of a graph, leading to the curse of dimensionality. In the case of TERGM, we believe there are two reasonable choices of network monitoring, namely it can be performed either in terms of the (normalised) network statistics or the model parameters whose dimension remains independent from the network evolvement.

To obtain a time series of the corresponding estimates, we propose to apply a moving window approach with the window size z. More precisely, we take into account the past z observations of the network $\{\boldsymbol{Y}_{t-z+1}, \ldots, \boldsymbol{Y}_t\}$ to estimate the respective quantities at time point t.

Let $\boldsymbol{\theta}$ be the true model parameters and $\hat{\boldsymbol{\theta}}_t$ their estimates at time point t based on the last z network states. Similarly, the expected value of the network statistics $\mathbb{E}_{\boldsymbol{\theta}}(\boldsymbol{s}(\boldsymbol{Y}))$ can be estimated as

$$\hat{\boldsymbol{s}}_t = \frac{1}{z} \sum_{n=0}^{z-1} \boldsymbol{s}(\boldsymbol{Y}_{t-n}). \tag{4}$$

Note that for the first time point, we cannot compute the memory terms because v previous network observations are not present. The same holds for the window size z, i.e., at least $v + z$ past network states must be observed before the monitoring.

Concerning the choice of monitoring the network statistics or the model parameters, it is worth noting that there is a one-to-one relationship between $\boldsymbol{\theta}$ and $\mathbb{E}_{\boldsymbol{\theta}}(\boldsymbol{s}(\boldsymbol{Y}))$. That is, for every $\boldsymbol{\theta}$, there is only one expectation of $\boldsymbol{s}(\boldsymbol{Y})$. Hence, one can monitor the network based on the estimates of either $\boldsymbol{\theta}$ or $\mathbb{E}_{\boldsymbol{\theta}}(\boldsymbol{s}(\boldsymbol{Y}))$. Since the monitoring procedure is identical for $\hat{\boldsymbol{\theta}}_t$ and $\hat{\boldsymbol{s}}_t$, we introduce a new notation $\hat{\boldsymbol{c}}_t$ for

the estimates of the network characteristics. Consequently, we refer to $\boldsymbol{c}$ meaning either $\boldsymbol{\theta}$ or $\mathbb{E}_{\boldsymbol{\theta}}(\boldsymbol{s}(\boldsymbol{Y}))$.

Let p be the number of network terms, which describe the in-control state and can reflect the deviations in the case of an out-of-control state. Thus, at time point t there is a p-dimensional vector $\hat{\boldsymbol{c}}_t = (\hat{c}_{1t}, \ldots, \hat{c}_{pt})'$ that estimates the network characteristics $\boldsymbol{c}$. Moreover, let $F_{\boldsymbol{c}_0, \boldsymbol{\Sigma}}$ be the target distribution of these estimates with $\boldsymbol{c}_0 = \mathbb{E}_0(\hat{c}_1, \ldots, \hat{c}_p)'$ being the expected value and $\boldsymbol{\Sigma}$ the respective $p \times p$ variance-covariance matrix of the network characteristics (Montgomery 2009). Thus,

$$\hat{\boldsymbol{c}}_t \quad \sim \quad \begin{cases} F_{\boldsymbol{c}_0, \boldsymbol{\Sigma}} & \text{if } t < \tau \\ F_{\boldsymbol{c}_\tau, \boldsymbol{\Sigma}} & \text{if } t \geq \tau \end{cases}, \tag{5}$$

where τ denotes a change point to be detected and $\boldsymbol{c}_\tau \neq \boldsymbol{c}_0$. If $\tau = \infty$ or $t < \tau$ the network is said to be in-control, whereas it is out of control in the case of $\tau \leq t < \infty$. Furthermore, we assume that the estimation precision of the parameters does not change across t, i.e., $\boldsymbol{\Sigma}$ is constant for the in-control and out-of-control state. Hence, the monitoring procedure is based on the expected values of $\hat{\boldsymbol{c}}_t$. In fact, we can specify the above mentioned hypothesis as follows

$$H_{0,t}: \ \mathbb{E}(\hat{\boldsymbol{c}}_t) = \boldsymbol{c}_0 \qquad \text{against} \qquad H_{1,t}: \ \mathbb{E}(\hat{\boldsymbol{c}}_t) \neq \boldsymbol{c}_0.$$

Typically, a multivariate control chart consists of the control statistic depending on one or more characteristic quantities, plotted in time order, and a horizontal line, called the upper control limit (UCL) that indicates the amount of acceptable variation. A hypothesis H_0 is rejected if the control statistic is equal to or exceeds the value of the UCL.

Subsequently, we discuss several control statistics together with presenting a method to determine the respective in-control parameters and UCLs.

2.3 Multivariate cumulative sum and exponentially weighted moving average control charts

The strength of the multivariate control chart over the univariate control chart is the ability to monitor several interrelated process variables. It implies that the corresponding test statistic should take into account the correlations of the data, be dimensionless and scale-invariant, as the process variables can considerably differ from each other. The squared Mahalanobis distance, which represents the general form of the control statistic, fulfils these criteria and is defined as

$$D_t^{(1)} = (\hat{\boldsymbol{c}}_t - \boldsymbol{c}_0)' \boldsymbol{\Sigma}^{-1} (\hat{\boldsymbol{c}}_t - \boldsymbol{c}_0), \tag{6}$$

being the part of the respective "data depth" expression—Mahalanobis depth that measures a deviation from an in-control distribution (cf. Liu 1995). Hence, $D_t^{(1)}$ maps the p-dimensional characteristic quantity $\hat{\boldsymbol{c}}_t$ to a one-dimensional measure. It is important to note that the characteristic quantity at time point t is usually the mean of several samples at t, but in our case, we only observe one network at each

instant of time. Thus, the characteristic quantity $\hat{c}_t$ is the value of the obtained estimates and not the average of several samples.

In this work, we apply two control chart types and compare their performance in network monitoring. Firstly, we discuss Multivariate CUSUM (MCUSUM) charts (cf. Woodall and Ncube 1985; Joseph et al. 1990; Ngai and Zhang 2001). One of the widely used versions was proposed by Crosier (1988) and is defined as follows

$$C_t = \left[(\boldsymbol{r}_{t-1} + \hat{\boldsymbol{c}}_t - \boldsymbol{c}_0)'\boldsymbol{\Sigma}^{-1}(\boldsymbol{r}_{t-1} + \hat{\boldsymbol{c}}_t - \boldsymbol{c}_0)\right]^{1/2}, \tag{7}$$

where

$$\boldsymbol{r}_t = \begin{cases} \boldsymbol{0} & \text{if } C_t \leq k, \\ (\boldsymbol{r}_{t-1} + \hat{\boldsymbol{c}}_t - \boldsymbol{c}_0)(1 - k/C_t) & \text{if } C_t > k, \end{cases}$$

given that $\boldsymbol{r}_0 = \boldsymbol{0}$ and $k > 0$. The respective chart statistic is

$$D_t^{(2)} = \boldsymbol{r}_t'\boldsymbol{\Sigma}^{-1}\boldsymbol{r}_t, \tag{8}$$

and it signals if $\sqrt{D_t^{(2)}}$ is greater than or equals the UCL. Certainly, the values k and UCL considerably influence the performance of the chart. The parameter k, also known as reference value or allowance, reflects the variation tolerance, taking into consideration δ—the deviation from the mean measured in the standard deviation units we aim to detect. According to Page (1954) and Crosier (1988), the chart is approximately optimal if $k = \delta/2$.

Secondly, we consider multivariate charts based on exponential smoothing (EWMA). Lowry et al. (1992) proposed a multivariate extension of the EWMA control chart (MEWMA), which is defined as follows

$$\boldsymbol{l}_t = \lambda(\hat{\boldsymbol{c}}_t - \boldsymbol{c}_0) + (1 - \lambda)\boldsymbol{l}_{t-1} \tag{9}$$

with $0 < \lambda \leq 1$ and $\boldsymbol{l}_0 = \boldsymbol{0}$ (cf. Montgomery 2009). The corresponding chart statistic is

$$D_t^{(3)} = \boldsymbol{l}_t'\boldsymbol{\Sigma}_{\boldsymbol{l}_t}^{-1}\boldsymbol{l}_t, \tag{10}$$

where the covariance matrix is defined as

$$\boldsymbol{\Sigma}_{\boldsymbol{l}_t} = \frac{\lambda}{2 - \lambda}\left[1 - (1 - \lambda)^{2t}\right]\boldsymbol{\Sigma}. \tag{11}$$

Together with the MCUSUM, the MEWMA is an advisable approach for detecting relatively small but persistent changes. However, the detection of large shifts is also possible by setting the reference parameter k or the smoothing parameter λ high. For instance, in case of the MEWMA with $\lambda = 1$, the chart statistic coincides with $D_t^{(1)}$. Thus, it is equivalent to Hotelling's T^2 control procedure, which is suitable for the detection of substantial deviations. It is worth mentioning that the discussed methods are directionally invariant, therefore, the investigation of the data at the signal time point is necessary if the change direction is of particular interest.

2.4 Estimation of the in-control parameters

In practice, the in-control parameters c_0 and Σ are usually unknown and therefore have to be estimated. Thus, one subdivides the sequence of network observations into Phase I and Phase II. In Phase I, the process must coincide with the in-control state so that the true in-control parameters c_0 and Σ can be estimated by the sample mean vector $\bar{c}$ and the sample covariance matrix S from $\hat{c}_t$.

It is important that Phase I replicates the natural behaviour of a network, so that possible dynamics related to its growth or changes in its topological structure are considered. Similarly, if the network is prone to remain constant, this fact should be captured in Phase I for reliable estimation and later network surveillance. After the estimates of c_0, Σ and the UCL are obtained, the calibrated control chart can be applied to the actual data in Phase II.

2.5 Computation of control limits

If $D_t^{(2)}$ or $D_t^{(3)}$ is equal to or exceeds the UCL, it means that the charts signal a change. To determine the UCLs, one typically assumes that the chart has a predefined (low) probability of false alarms, i.e., signals when the process is in control, or a prescribed in-control Average Run Length ARL_0 that represents the number of expected time steps until the first signal. To compute the UCLs corresponding to ARL_0 theoretically, a prevalent number of multivariate control charts require a normally distributed target process (cf. Johnson and Wichern 2007; Porzio and Ragozini 2008; Montgomery 2009). In our case, this assumption would need to be valid for the estimates of the model parameters/the network statistics. However, while there are some studies on the distributions of particular network statistics $s(Y)$ (cf. Yan and Xu 2013; Yan et al. 2016; Sambale and Sinulis 2018), only a few results are obtained about the parameter estimates of θ. Primarily, the difficulties to determine the distribution is that the assumption of independent and identically distributed data is violated in the ERGM case. In addition, the parameters depend on the choice of the model terms and network size (He and Zheng 2015). Kolaczyk and Krivitsky (2015) proved asymptotic normality for the ML estimators in a simplified context of the ERGM, pointing out the necessity to establish a deeper understanding of the distributional properties of parameter estimators.

In case that the normality assumption is violated to a slight or moderate degree, the control charts still will remain robust (Montgomery 2009). The most crucial assumption that needs to be satisfied is the independence of the observations at different time points (Qiu 2013). If the data is autocorrelated, the theoretically derived UCLs become invalid, so that their implementation would lead to inaccurate results. Here, we consider networks which are dependent over time. Moreover, the networks used for estimation of the characteristics $\hat{c}_t$ are overlapping due to the application of the moving window approach. As shown in Sect. 3.2, the characteristics that are based on the averaged network statistics $\hat{s}_t$ can violate this assumption substantially. Regarding the estimates $\hat{\theta}_t$, if their computation does not

involve overlapping of the networks by the sliding window approach of size z, i.e., each graph is involved only once in the estimation of $\boldsymbol{\theta}$, and the size of z is enough for recovering the temporal dependence completely, then the estimates become uncorrelated. However, as we design an online monitoring procedure, we support the idea of computing $\hat{\boldsymbol{c}}_t$ immediately as soon as a new data point is available. In this case, we account for the correlation between the estimated characteristics $\hat{\boldsymbol{c}}_t$.

There are several works which apply control charts in the presence of autocorrelation, advising either using the residuals of the time series models as observations, calculating theoretical control limits under autocorrelation or designing a simulation study to determine the control limits corresponding to the desired ARL_0 (cf. Montgomery and Mastrangelo 1991; Alwan 1992; Runger and Willemain 1995; Schmid and Schöne 1997; Zhang 1997; Lu and Reynolds Jr 1999, 2001; Sheu and Lu 2009). It is worth noting that the residual charts have different properties from the traditional charts, which we consider in this work. Hence, we determine the UCLs via Monte Carlo simulations described in Sect. 3.2.

3 Simulation study

To verify the applicability and effectiveness of the discussed approach, we design a simulation study followed by the surveillance of real-world data with the goal of obtaining some insights into its temporal development.

3.1 Generation of network time series

To compute $\bar{c}$ and $\boldsymbol{S}$ we need a certain number of in-control networks. For this purpose, we generate 2500 temporal graphs of desired length $T < \tau$, where each graph consists of $N = 100$ nodes. The simulation of synthetic networks is based on the Markov chain principle: the network observation in time point $\boldsymbol{Y}_t$ is simulated from its previous state $\boldsymbol{Y}_{t-1}$ by selecting randomly a fraction ϕ of elements of the adjacency matrix and setting them to either 1 or 0, according to a specified transition matrix $\boldsymbol{M}$. This setting allows us to include the memory term during the estimation of the TERGM that reflects the stability of both edges and non-edges between the previous and the current network observation. The in-control values are $\phi_0 = 0.01$ and

$$\boldsymbol{M}_0 = \begin{pmatrix} m_{00,0} & m_{01,0} \\ m_{10,0} & m_{11,0} \end{pmatrix} = \begin{pmatrix} 0.9 & 0.1 \\ 0.4 & 0.6 \end{pmatrix},$$

where $m_{ij,0}$ denotes the probability of a transition from i to j in the in-control state.

At the beginning of each sequence, a directed network which is called the "base network" is simulated by applying an ERGM with predefined network terms and corresponding coefficients so that it is possible to control the "network creation" indirectly. This procedure helps to guarantee that the temporal networks have a stochastic but analogous initialisation. In our case, we select three network statistics, namely an edge term, a triangle term and a parameter that defines asymmetric dyads. These terms are used later for estimating network characteristics.

Subsequently, a new graph is produced by applying the in-control fraction ϕ and the transition matrix $\boldsymbol{M}$.

Next, we need to confirm that the generated samples of networks behave according to the requirements of Phase I, i.e., capturing only the usual variation of the target process. For this purpose, we can exploit Markov chain properties and calculate its steady-state equilibrium vector $\boldsymbol{\pi}$, as it follows that the expected number of non-edges and edges is given by $\boldsymbol{\pi}$. Using eigenvector decomposition, we find the steady-state to be $\boldsymbol{\pi} = (0.8, 0.2)'$. Consequently, the expected number of edges in the graph in its steady-state is 1980. However, the network density is only one of the aspects to define the in-control process, as the temporal development and the topology are also involved in the network creation. Hence, we identify the suitable start of the considered network sequence by computing the network statistics $s(\boldsymbol{Y}_t)$ over multiple network time series. By plotting the behaviour, we determined that all four terms become stable by $t = 1000$. Thus, we simulate 1000 network observations in a burn-in period so that the in-control sequence of network states starts at $t = 1001$.

3.2 Calibration of the charts in Phase I

After the generation of temporal networks, we compute $\hat{\boldsymbol{\theta}}_t$ by fitting the TERGM and $\hat{s}_t$ by applying Eq. (4) with a certain window size z using the four network terms, namely edge term, a triangle term, a term that defines asymmetric dyads and a memory term which describes the stability of both edges and non-edges over time with the temporal lag $v = 1$. Currently, there are two widely used approaches to estimate the TERGM: Maximum Pseudolikelihood Estimation (MPLE) with bootstrapped confidence intervals and MCMC ML estimation (Leifeld et al. 2018). The chosen estimation method to derive $\hat{\boldsymbol{\theta}}_t$ is the bootstrap MPLE which is appropriate to handle a relatively large number of nodes and time points (Leifeld et al. 2018). Next, we calculate the in-control parameters $\bar{c}$ and $\boldsymbol{S}$ for both monitoring cases. Finally, we calibrate different control charts by obtaining the UCLs with respect to the predefined ARL_0 via the bisection method. For two window sizes $z = \{7, 14\}$, Tables 1 and 3 summarise the obtained results for surveillance of $\boldsymbol{\theta}$, and Tables 2 and 4 for surveillance of $s(\boldsymbol{Y})$ with the MEWMA and MCUSUM charts respectively. If the reader wishes to apply the TERGM with the same network terms and similar window size as we did in this work, the presented UCLs can be used directly. Otherwise, it is necessary to conduct different Monte Carlo simulations that address the specific settings of the TERGM.

As both network characteristics describe the same process, we would expect the UCL results to be similar. However, in Fig. 1 the analysis of the autocorrelation functions applied to the estimates of one of the generated network time series shows that the dependence structures of $\hat{\boldsymbol{\theta}}_t$ and $\hat{s}_t$ considerably differ. While the elimination of the overlap in the calculation procedure removes correlation in the case of the parameter estimates $\hat{\boldsymbol{\theta}}_t$, there is only a slight improvement regarding the averaged network statistics $\hat{s}_t$. Thus, the UCLs are different for both cases.

Table 1 Upper control limits for the MEWMA chart based on the estimates $\hat{\theta}_t$ and $ARL_0 \in \{50, 75, 100\}$ for two different windows sizes $z = 7$ and $z = 14$

z	ARL_0/λ	0.1	0.2	0.3	0.4	0.5	0.6	0.7	0.8	0.9	1.0
7	50	39.32	35.76	31.04	26.52	22.58	19.22	16.46	14.10	12.14	10.46
	75	45.76	41.03	35.20	29.81	25.27	21.61	18.53	15.86	13.66	11.72
	100	50.52	45.30	38.58	32.56	27.62	23.48	20.00	17.13	14.71	12.66
14	50	55.14	43.44	33.80	26.96	21.98	18.16	15.16	12.76	10.76	9.15
	75	65.63	50.94	39.29	31.23	25.26	20.68	17.24	14.50	12.20	10.32
	100	73.85	56.20	42.97	33.97	27.44	22.52	18.72	15.69	13.21	11.15

Table 2 Upper control limits for the MEWMA chart based on the estimates $\hat{s}_t$ and $ARL_0 \in \{50, 75, 100\}$ for two different windows sizes $z = 7$ and $z = 14$

z	ARL_0/λ	0.1	0.2	0.3	0.4	0.5	0.6	0.7	0.8	0.9	1.0
7	50	65.03	43.79	32.23	25.29	20.36	16.72	13.91	11.66	9.85	8.31
	75	82.40	52.41	38.13	29.52	23.59	19.23	15.88	13.23	11.13	9.42
	100	96.23	59.39	42.92	32.96	26.19	21.24	17.58	14.69	12.32	10.35
14	50	71.09	45.47	32.66	24.81	19.51	15.80	12.95	10.74	8.96	7.53
	75	89.03	55.26	38.71	29.06	22.85	18.40	15.07	12.46	10.35	8.66
	100	103.00	62.73	43.73	32.56	25.40	20.40	16.65	13.73	11.43	9.57

Table 3 Upper control limits for the MCUSUM chart based on the estimates $\hat{\theta}_t$ and $ARL_0 \in \{50, 75, 100\}$ for two different windows sizes $z = 7$ and $z = 14$

z	ARL_0/k	0.5	0.6	0.7	0.8	0.9	1.0	1.1	1.2	1.3	1.4	1.5
7	50	21.01	19.36	17.75	16.21	14.83	13.52	12.25	11.10	9.99	9.00	8.03
	75	25.19	22.90	20.82	18.95	17.33	15.91	14.47	13.19	11.94	10.74	9.61
	100	28.25	25.59	23.31	21.18	19.38	17.67	16.12	14.67	13.27	11.96	10.73
14	50	30.10	27.84	25.64	23.67	21.69	19.83	17.92	16.17	14.44	12.75	11.06
	75	37.35	34.60	31.91	29.25	26.86	24.53	22.37	20.20	18.14	16.19	14.32
	100	43.06	39.52	36.20	33.15	30.45	27.84	25.43	23.16	20.97	18.81	16.77

## 3.3 Design of the anomalous behaviour

To test how well the proposed control charts can detect the changes in the network's development, it is necessary to compose different anomalous cases and generate samples from Phase II. Since our focus is on the detection of shifts in the process mean, an anomalous change can occur either in the proportion of the asymmetric

Table 4 Upper control limits for the MCUSUM chart based on the estimates $\hat{s}_t$ and $ARL_0 \in \{50, 75, 100\}$ for two different windows sizes $z = 7$ and $z = 14$

z	ARL_0/k	0.5	0.6	0.7	0.8	0.9	1.0	1.1	1.2	1.3	1.4	1.5
7	50	51.85	47.41	42.96	38.27	33.87	29.71	25.94	22.51	19.01	15.79	13.06
	75	75.93	69.26	62.97	56.83	50.79	44.82	39.02	33.73	29.05	24.95	20.95
	100	97.58	89.46	81.68	73.51	65.78	59.01	52.43	45.96	39.96	34.40	29.23
14	50	55.72	51.27	46.63	41.90	37.54	33.29	29.18	25.46	21.69	18.21	15.36
	75	80.28	73.70	67.32	61.13	54.75	48.86	43.15	37.76	32.81	28.26	24.20
	100	102.34	94.25	85.88	78.05	70.90	63.96	57.03	50.65	44.31	38.71	33.39
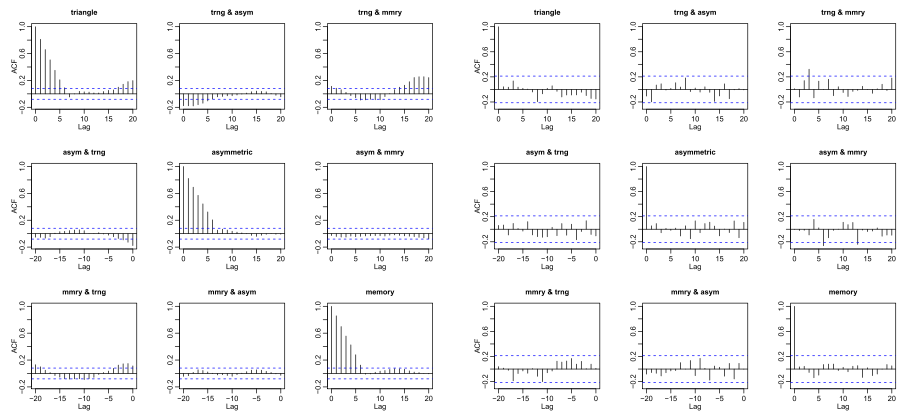
(a) ACF of the estimates $\hat{\theta}_t$ on the example of triangle, asymmetric and memory network terms.
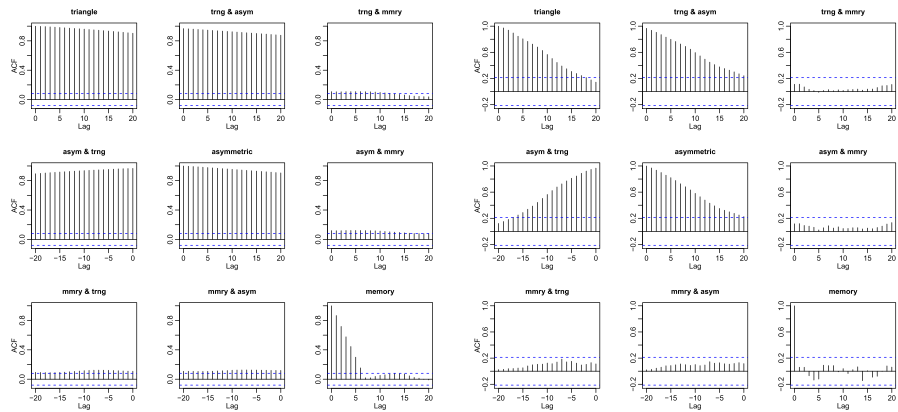
(b) ACF of the estimates $\hat{s}_t$ on the example of triangle, asymmetric and memory network terms.

Fig. 1 Comparison of the autocorrelation function (ACF) values when the network characteristics are estimated with a sliding window approach of size $z = 7$ containing (left) and not containing (right) overlapping network states
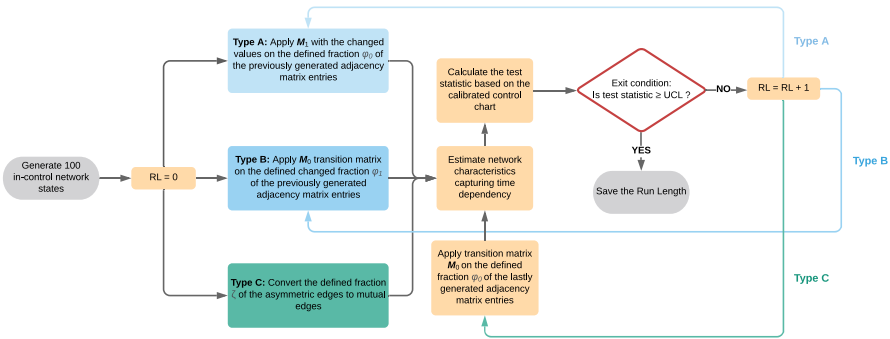
Fig. 2 Anomaly types for the generation of observations from Phase II and calculation of the associated run length

edges, in the fraction of the randomly selected adjacency matrix entries ϕ or the transition matrix $\boldsymbol{M}$. Thus, we subdivide these scenarios into three different anomaly types which are briefly described in the flow chart presented in Fig. 2.

We define a Type A anomaly as a change in the values of $\boldsymbol{M}$. That is, there is a transition matrix $\boldsymbol{M}_1 \neq \boldsymbol{M}_0$ when $t \geq \tau$. Similar to Type A, we consider anomalies of Type B by introducing a new fraction value ϕ_1 in the generation process when $t \geq \tau$. Both types are instances of a persistent change (also known as simply a "change"), where the abnormal development continues for all $t \geq \tau$ (Ranshous et al. 2015). Anomalies of Type C differ from the previous two types as they represent a "point change" (also referred to as an "event")—the abnormal behaviour occurs only at a single point of time τ but its outcome may also affect subsequent network states in our case due to the Markov property. We recreate this type of anomaly by converting a fraction ζ of asymmetric edges into mutual links. This process happens at time point τ only. Afterwards, the new network states are created similar to Phase I by applying $\boldsymbol{M}_0$ and ϕ_0 up until the anomaly is detected. The considered cases are summarised in Table 5.

3.4 Performance of the charts in Phase II

In the next step, we analyse the performance of the proposed charts in terms of their detection speed. As a performance measure, we compute the conditional expected delay (CED) of detection, conditional on a false signal not having been occurred before the (unknown) time of change τ (Kenett and Pollak 2012). For our simulation, we set $\tau = 101$. Using 250 simulations, we estimate the CED based on the UCLs with $ARL_0 = 50$ for each setting. That means we would expect CED = 50 if no change happened and it should be considerably smaller in the case of an anomaly. Figures 3, 4 and 5 present the results of the simulation for anomalies of Type A, B and C, respectively.

There are several aspects to assess fully the obtained results. First of all, the comparison of performance between the MCUSUM and the MEWMA control charts. In most of the cases, the CED of the MEWMA chart is smaller compared to

the corresponding MCUSUM chart. However, for the best choice of the reference parameter k or the smoothing parameter λ, both charts are competitive. The respective values are indicated by the large dots indicating the minimum on the CED curve. For instance, the weakest change of Type A.1 (Fig. 3a) is detected quicker by the MCUSUM chart with the low parameters k. In contrast, the MEWMA charts perform better for bigger changes such as in Cases 2 and 3.

Generally speaking, we see that the CED is decreasing if the shift size or the intensity of the change is increasing. Moreover, if the reference parameter k or the smoothing parameter λ is smaller, less intense anomalies can be detected. If in practical implementation the detection of larger changes is required, these parameters should also be higher. It is worth reminding that the MEWMA chart coincides with Hotelling's T^2 chart if $\lambda = 1$, i.e., the control statistic depends only on the current value.

The disadvantage of both approaches is that small and persistent changes are not detected quickly when the parameters k or λ are not optimally chosen. For example, considering Case A.1 in Fig. 3b, we can notice that at the high values of the parameter λ the CED slightly exceeds the ARL_0 reflecting the poor performance. However, a careful selection of the parameters can overcome this problem. Also, the choice of the window size plays a significant role in detecting the anomalies reliably, being a trade-off between a precise description of the process and the ability to reflect the sudden changes in its behaviour.

Regarding the differences in results with respect to the quantities $\hat{\boldsymbol{\theta}}_t$ and $\hat{\boldsymbol{s}}_t$, we notice a similar performance in Anomaly Types A and B. It is interesting that in most of the cases the MEWMA control charts work better for $\hat{\boldsymbol{s}}_t$ and the CUSUM control charts for $\hat{\boldsymbol{\theta}}_t$. However, looking at the detection of anomaly Type C.2, we observe a considerable advantage of applying $\hat{\boldsymbol{\theta}}_t$ rather than $\hat{\boldsymbol{s}}_t$. To confirm that this behaviour is supported by another example, we created an additional test case with

Table 5 Anomaly cases

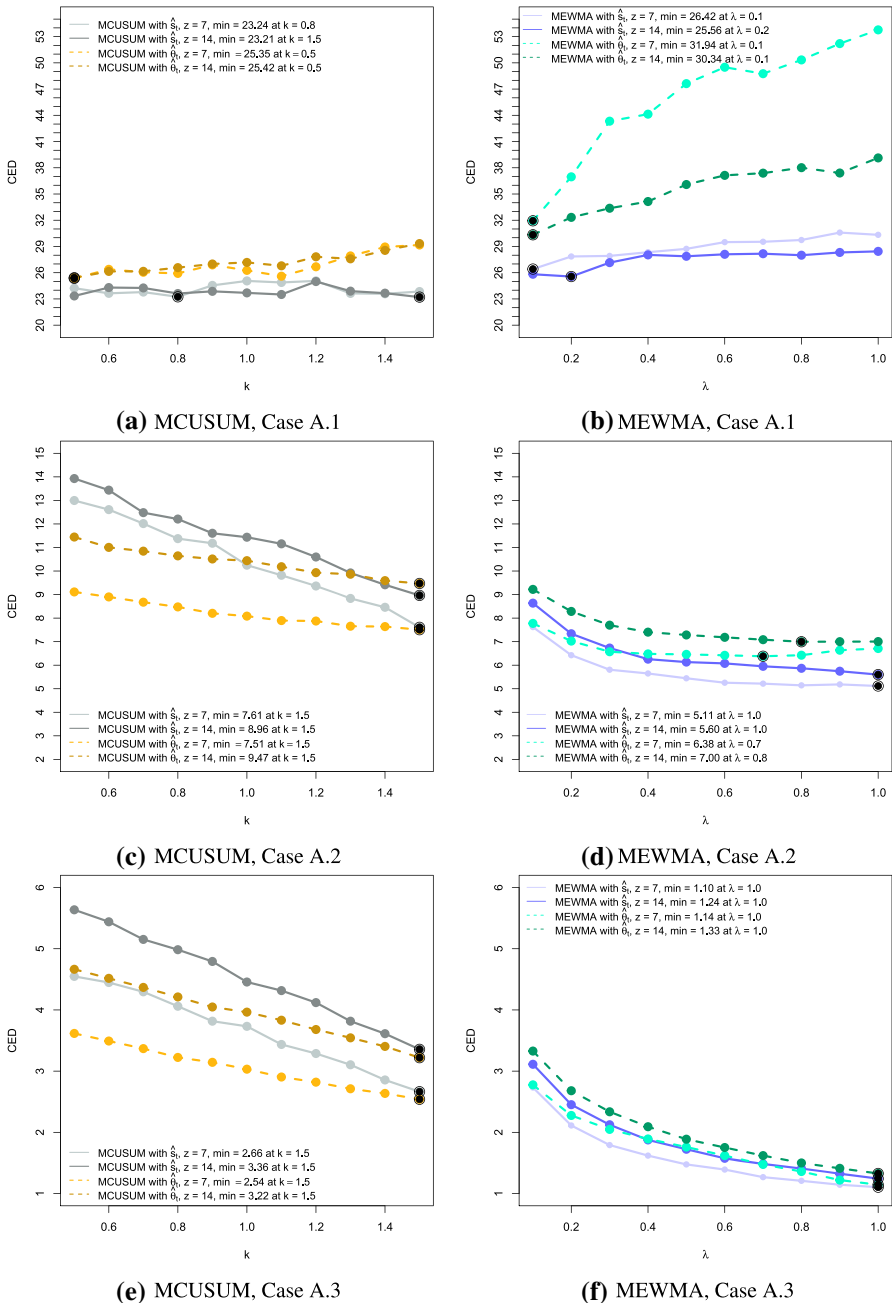| Anomaly Type	Description	Case	
Type A	Change in the transition matrix M	A.1	$m_{00,1} = 0.89$ $(m_{00,0} = 0.9)$
			$m_{01,1} = 0.11$ $(m_{01,0} = 0.1)$
		A.2	$m_{10,1} = 0.6$ $(m_{10,0} = 0.4)$
			$m_{11,1} = 0.4$ $(m_{11,0} = 0.6)$
		A.3	$m_{00,1} = 0.5$ $(m_{00,0} = 0.9)$
			$m_{11,1} = 0.5$ $(m_{11,0} = 0.6)$
Type B	Change of the fraction $\phi_0 = 0.01$	B.1	$\phi_1 = 0.009$
		B.2	$\phi_1 = 0.015$
		B.3	$\phi_1 = 0.02$
Type C	Increase of the proportion of mutual edges by ζ	C.1	$\zeta = 0.005$
		C.2	$\zeta = 0.01$
		C.3	$\zeta = 0.05$

Fig. 3 Conditional expected delays for anomalies of Type A for MCUSUM (left) and MEWMA (right) together with the different choices of the reference parameter k and the smoothing parameter λ, the window sizes $z = 7$ and $z = 14$, and the network estimates $\hat{s}_t$ (solid lines) and $\hat{\theta}_t$ (dashed lines). Black points indicate the minimum CED for each setting
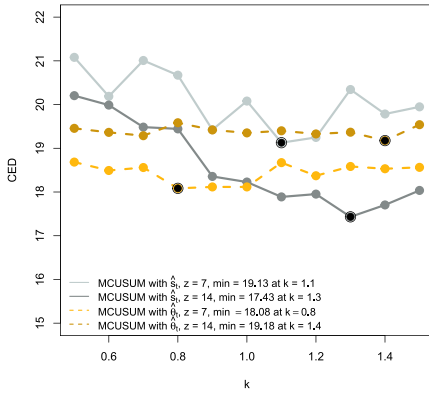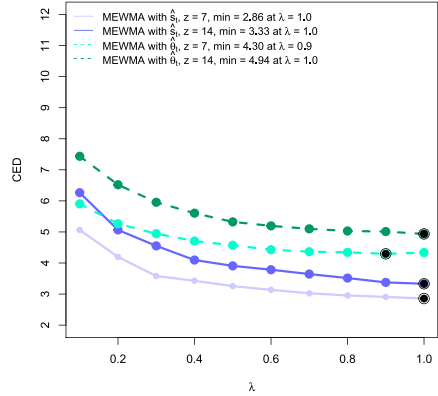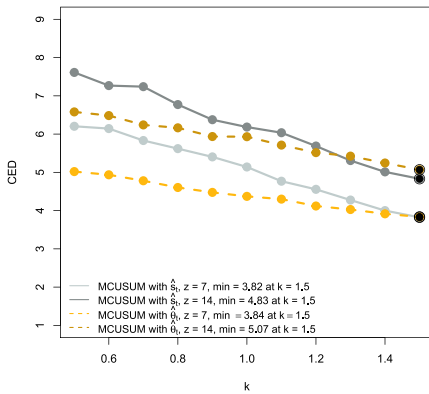
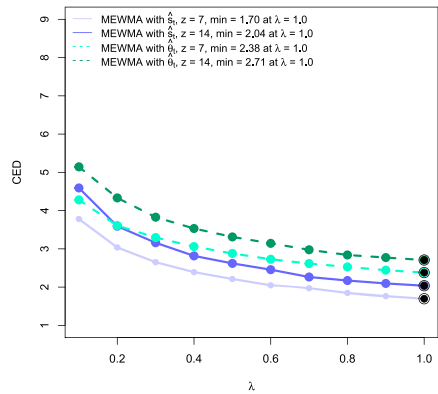**(a)** MCUSUM, Case B.1
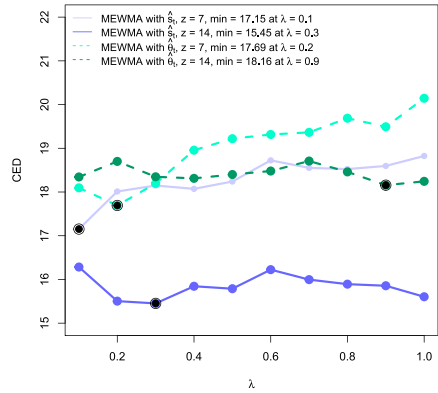
(b) MEWMA, Case B.1
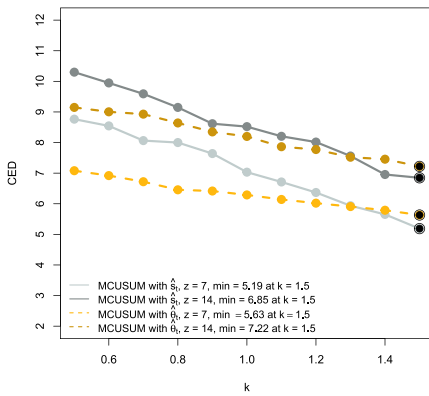
(c) MCUSUM, Case B.2

(d) MEWMA, Case B.2

(e) MCUSUM, Case B.3

(f) MEWMA, Case B.3

Fig. 4 Conditional expected delays for anomalies of Type B for MCUSUM (left) and MEWMA (right) together with the different choices of the reference parameter k and the smoothing parameter λ, the window sizes $z = 7$ and $z = 14$, and the network estimates $\hat{s}_t$ (solid lines) and $\hat{\theta}_t$ (dashed lines). Black points indicate the minimum CED for each setting
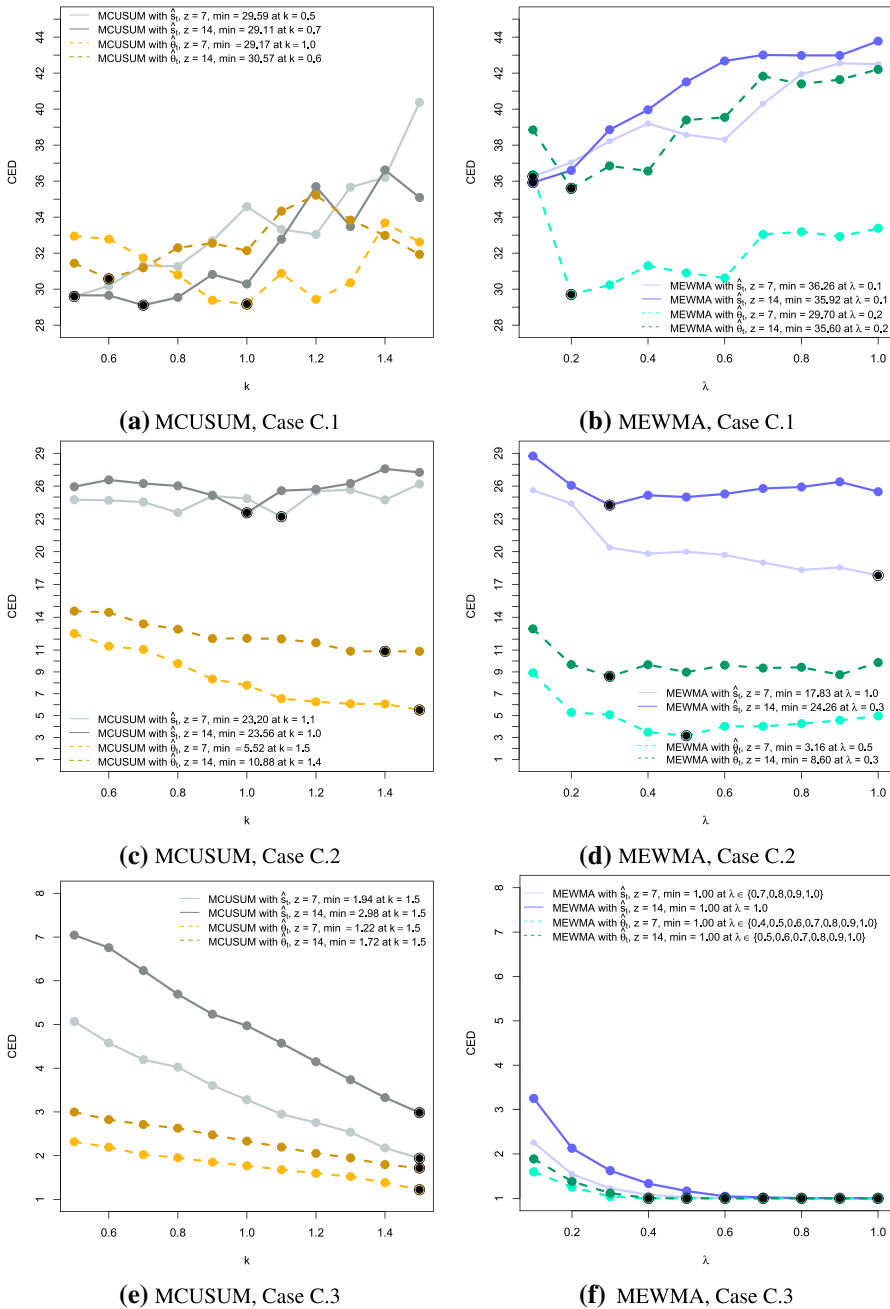
Fig. 5 Conditional expected delays for anomalies of Type C for MCUSUM (left) and MEWMA (right) together with the different choices of the reference parameter k and the smoothing parameter λ, the window sizes $z = 7$ and $z = 14$, and the network estimates $\hat{s}_t$ (solid lines) and $\hat{\theta}_t$ (dashed lines). Black points indicate the minimum CED for each setting

Table 6 Summary of the CED results to detect anomalies of Type C with the additional test case $\zeta = 0.02$

Case Parameter ζ	CED	$\hat{s}_t$ C.1 0.005	C.2 0.01	 0.02	C.3 0.05	$\hat{\theta}_t$ C.1 0.005	C.2 0.01	 0.02	C.3 0.05
MEWMA with $z = 7$	Min.	36.26	17.83	1.85	<u>1.00</u>	29.70	3.16	1.00	<u>1.00</u>
	λ_{min}	0.1	1.0	1.0	0.7	0.2	0.5	0.9	0.4
	Max.	42.55	25.62	7.16	2.26	36.36	8.92	2.56	1.60
	λ_{max}	0.9	0.1	0.1	0.1	0.1	0.1	0.1	0.1
MEWMA with $z = 14$	Min.	35.92	24.26	3.57	<u>1.00</u>	35.60	8.60	<u>1.00</u>	<u>1.00</u>
	λ_{min}	0.1	0.3	0.9	1.0	0.2	0.3	1.0	0.5
	Max.	43.78	28.75	9.90	3.25	42.21	12.94	3.49	1.89
	λ_{max}	1.0	0.1	0.1	0.1	1.0	0.1	0.1	0.1
MCUSUM with $z = 7$	Min.	29.59	23.20	6.66	1.94	29.17	<u>5.52</u>	<u>2.10</u>	<u>1.22</u>
	k_{min}	0.5	1.1	1.5	1.5	1.0	1.5	1.5	1.5
	Max.	40.38	26.19	18.70	5.07	33.68	12.51	3.91	2.32
	k_{max}	1.5	1.5	0.5	0.5	1.4	0.5	0.5	0.5
MCUSUM with $z = 14$	Min.	<u>29.11</u>	23.56	9.45	2.98	30.57	10.88	3.14	1.72
	k_{min}	0.7	1.0	1.5	1.5	0.6	1.4	1.5	1.5
	Max.	36.62	27.58	18.86	7.04	35.22	14.56	6.19	3.00
	k_{max}	1.4	1.4	0.5	0.5	1.2	0.5	0.5	0.5

The corresponding smoothing and reference parameters λ and k are provided under the respective CED. The minimum CED for each case and the control chart group are underlined. The maximum CED represents the "worst-case" scenario. In case several values of the parameter λ correspond to the CED result, only the smallest value is reported

$\zeta = 0.02$. These results as well as the others from Type C anomalies are summarised in Table 6. As we can observe, if the change is too small, then both groups of control charts created on the basis of $\hat{\theta}_t$ and $\hat{s}_t$ need relatively long to detect it. In case when $\zeta = 0.05$, representing a substantial anomaly, the change is identified quickly by both options. However, when the change is of a moderate degree, for example, $\zeta = 0.02$, then the control charts based on $\hat{\theta}_t$ signal the anomalous behaviour considerably quicker. Whether the main reason for such difference is the particular type of anomaly, namely it is an example of a point change, cannot be said generally as additional variations of such anomalies should be examined. However, from the evidence in this work, the authors hypothesise that the estimates $\hat{\theta}_t$ might be more suitable for general network monitoring when it is assumed that a point, as well as a persistent change can occur, though the comparison between the performance of $\hat{\theta}_t$ and $\hat{s}_t$ is worth further investigation.

To summarise, the effectiveness of the presented charts to detect structural changes depends significantly on the accurate estimation of the anomaly size one

aims to detect. Thus, to ensure that no anomalies were missed, it can be effective to apply paired charts and benefit from the strengths of each of them to detect varying types and sizes of anomalies, if the information on the possible change is not available or not reliable.

4 Empirical illustration

To demonstrate the applicability of the described method, we monitor the daily flight data of the United States (US) published by the US Bureau of Transportation Statistics using the parameter estimates $\hat{\theta}_t$. Each day can be represented as a directed network, where nodes are airports and directed edges define flights between airports. In Fig. 6, examples of flight network data in 2018, 2019 and 2020 (until the end of April) are presented.

Flexibility in choosing the network terms, according to the type of anomalies one would like to detect, enables different perspectives on the same network data. In our case, we aim to identify considerable changes in network development. The intuition of how the flight network usually operates guides the choice of its terms. At the time of writing, due to the current Coronavirus disease (COVID-19) pandemic in the year 2020, some regions have paused the operation of transport systems with the aim to reduce the number of new infections. However, the providers enable mobility by establishing connections through territories which allow travelling. That means, instead of having a direct journey from one geographical point to another, currently the route passes through several locations, which can be interpreted as nodes. Thus, the topology of the graph has changed: instead of directed mutual links, the number of intransitive triads and asymmetric links starts to increase significantly. We can incorporate both terms, together with the edge term and a memory term ($v = 1$), and expect the estimates of the respective coefficients belonging to the first two statistics to be close to zero or strongly negative in the in-control case.

Initially, we need to decide which data are suitable to define observations coming from Phase I, i.e., the in-control state. There were no considerable events which
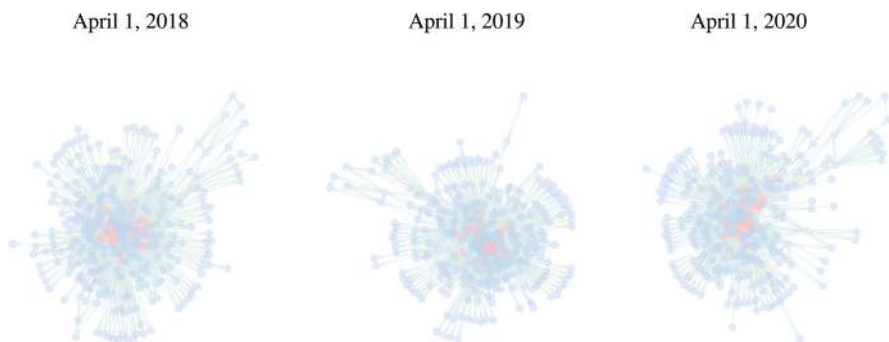

April 1, 2018   April 1, 2019   April 1, 2020

Fig. 6 Illustration of the flight network on April 1 of each year excluding isolated vertices. It can be seen that the topology of the network has changed. The red coloured nodes represent the 30 busiest airports

would seriously affect the US flight network known to the authors in the year 2018, therefore, we chose this year to characterise the in-control state. Consequently, the years 2019 and 2020 represent Phase II. To capture the weekly patterns, a time window of size $z = 7$ was chosen, so that the first instant of time when the monitoring begins represents January 8, 2018. In this case, Phase I consists of 358 observations and Phase II of 486 observations. To guarantee that only factual flight data are considered, we remove cases when a flight was cancelled. Additionally, we eliminate multiple edges. The main descriptive statistics for Phase I and II are reported in Table 7. There are no obvious changes when considering the descriptive statistics. Hence, control charts, which are only based on such characteristics, could fail to detect the possible changes in 2019 and 2020. When considering the estimates $\hat{\boldsymbol{\theta}}_t$ of the TERGM described by a series of boxplots in Fig. 7, we can observe substantial changes in the values.

Before proceeding with the analysis, it is important to evaluate whether a TERGM fits the data well (Hunter et al. 2008). For each of the years, we randomly selected one period of the length z and simulated 500 networks based on the parameter estimates from each of the corresponding networks. Figure 8 depicts the results for the time frame April 3–9 2019, where the grey boxplots of each of the statistics represent the simulations, and the solid black lines connect the median values of the observed networks. Despite the relatively simple definition of the model, some typical network characteristics such as the distributions of edge-wise shared partners, the vertex degrees, various triadic configurations (triad census) and geodesic distances (the value of infinity replicates the existence of isolated nodes) match the observed distributions of the same statistics satisfactory.

To select appropriate control charts, we need to take into consideration the specifications of the flight network data. Firstly, it is common to have 3–4 travel peaks per year around holidays, which are not explicitly modelled, so that we can detect these changes as verifiable anomalous patterns. It is worth noting that one

Table 7 Descriptive statistics of the US flight network data		2018	2019	2020
Phase		I	II	II
Number of nodes		358	360	354
Density	Min.	0.031	0.033	0.022
	Median	0.037	0.038	0.038
	Max.	0.039	0.040	0.041
	Min.	0.97	0.96	0.89
Reciprocity	Median	0.99	0.99	0.99
	Max.	1.00	1.00	1.00
	Min.	0.315	0.322	0.263
Transitivity	Median	0.339	0.339	0.326
	Max.	0.357	0.354	0.345

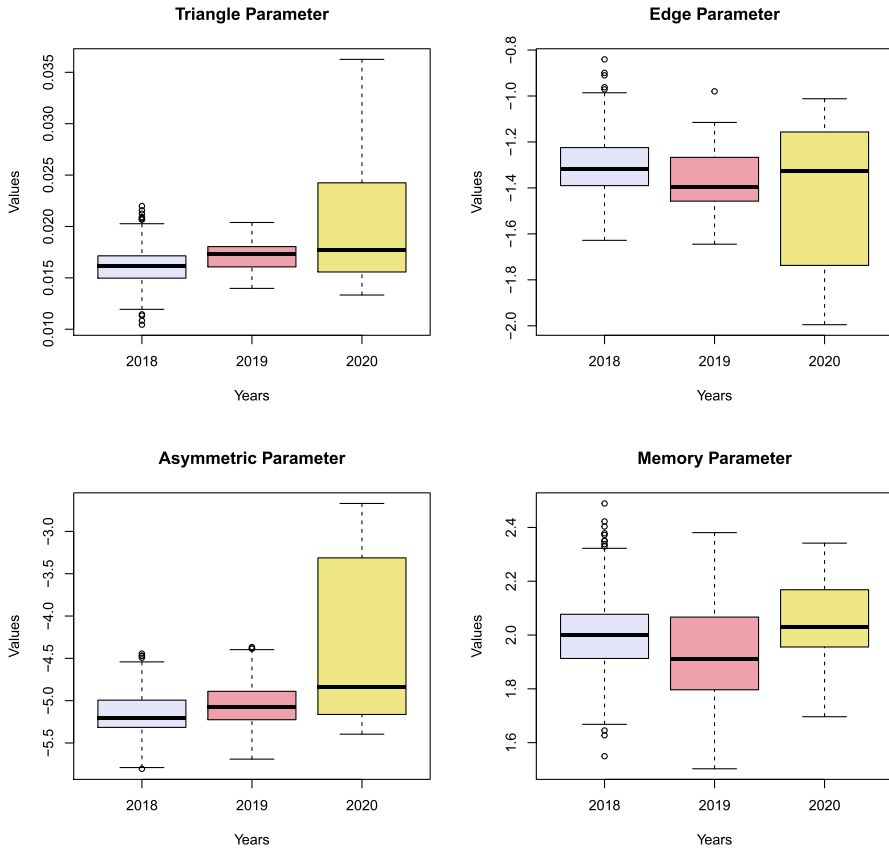Density is calculated on networks without multiple edges

Fig. 7 Distribution of the estimated coefficients $\hat{\theta}_t$ in 2018, 2019 and 2020

could account for such seasonality by including nodal or edge covariates. Secondly, as we aim to detect considerable deviations from the in-control state, we are more interested in sequences of signals. Thus, we have chosen $k = 1.5$ for MCUSUM and $\lambda = 0.9$ for the MEWMA chart. The target ARL_0 is set to 100 days, therefore, we can expect roughly 3.65 in-control signals per year by the construction of the charts.

Figure 9 depicts the results of both charts for monitoring the US flight network data. In Phase I there are slightly more in-control signals than expected, which we leave without investigation as they occur as single instances. Considering Phase II, there are several anomalous behaviours which were detected. The first series of signals in summer 2019 is due to a particularly increased demand for flights during the holidays. The second sequence of signals corresponds to the development of the COVID-19 pandemic. On March 19, the State Department issued a Level 4 "do not travel" advisory, recommending that United States citizens avoid any global travel. Although this security measure emphasises international flights, it also influences domestic aerial connections. The continuous sequence of the signals in case of the MEWMA begins on March 21, 2020. In case of the MCUSUM, the start is on
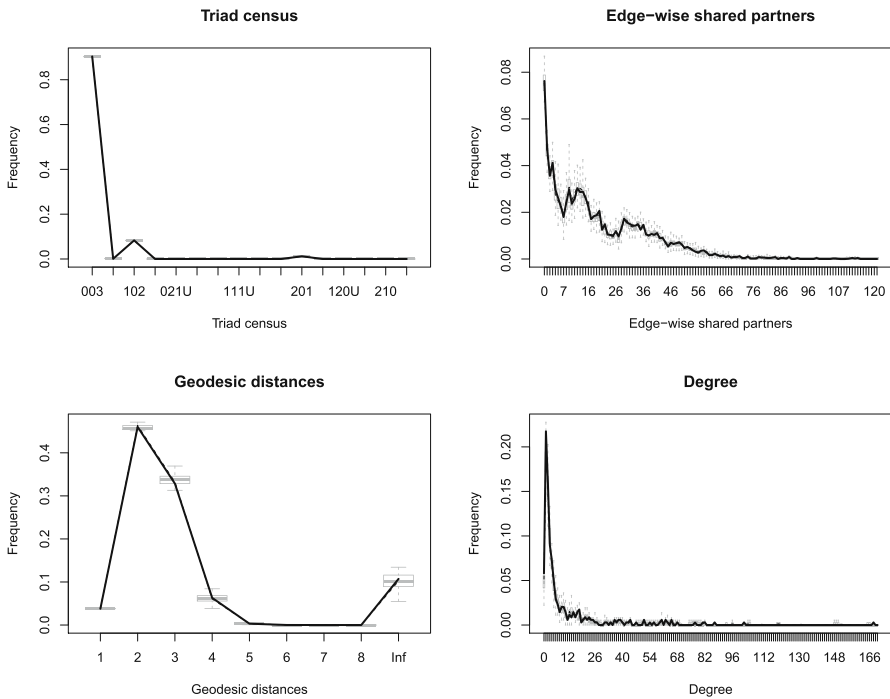
Fig. 8 Illustration of the goodness of fit assessment for the TERGM. The considered networks belong to the period April 3–9 2019

March 24. Although in both cases the control statistic resets to zero after each signalling, the repeated violation of the upper control limit is a clear indicator of this shift in network behaviour.

To identify smaller and more specific changes in the daily flight data of the US, one could also integrate nodal and edge covariates, which would refer to further aspects of the network. Additionally, control charts with smaller k and λ can be applied.

5 Conclusion

Statistical methods can be remarkably powerful for the surveillance of networks. However, due to the complex structure and potentially large size of the adjacency matrix, traditional tools for multivariate process control cannot directly be applied, as the network's complexity must be reduced first. For instance, this can be done by statistical modelling of the network. The choice of the model is crucial as it decides constraints and simplifications of the network which later influence the types of changes we are able to detect. In this paper, we show how multivariate control charts can be used to detect changes in dynamic networks generated by the TERGM. The proposed methods can be applied in real time. This general approach is applicable for various types of networks in terms of the edge direction and
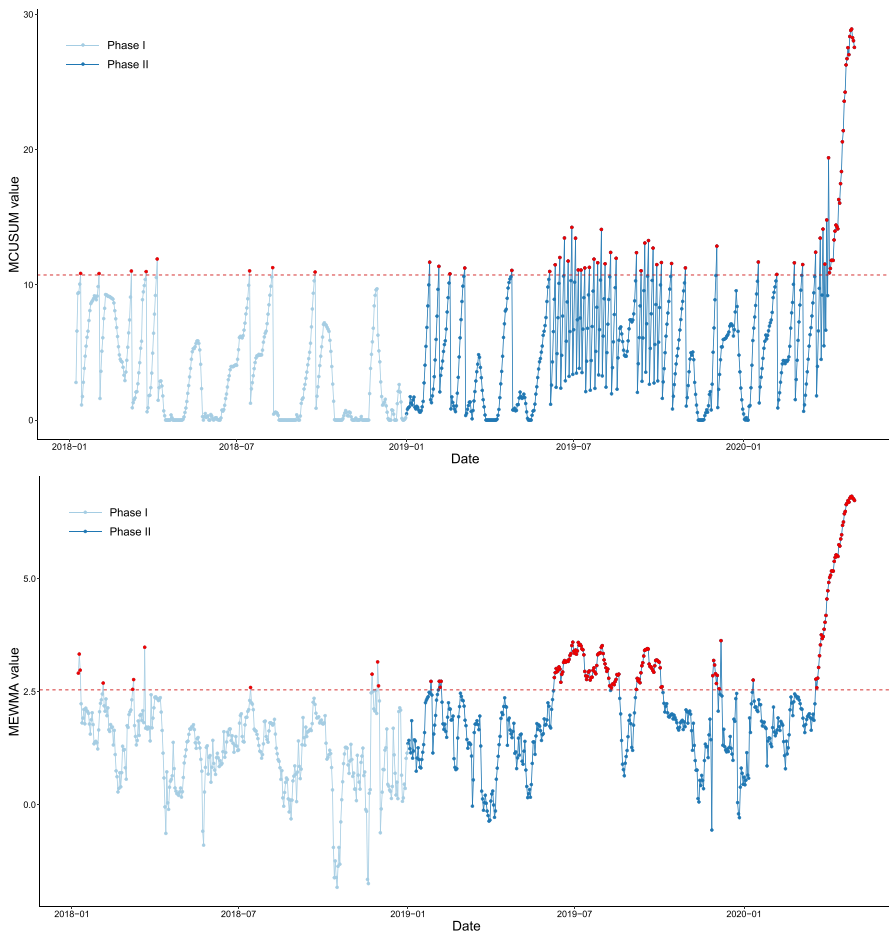
Fig. 9 The MCUSUM control chart (above) and the logarithmic MEWMA control chart (below). The horizontal red line corresponds to the upper control limit and the red points to the occurred signals

topology. Moreover, it enables the integration of nodal and edge covariates and considers temporal dependence.

The performance of our procedure is evaluated for different anomalous scenarios by comparing the CED of the calibrated control charts. According to the classification and explanation of anomalies provided by Ranshous et al. (2015), the surveillance method presented in this work is applicable for event and change detection in temporal networks.

Finally, we illustrated the applicability of our approach by monitoring daily flights in the United States. Both control charts were able to detect the beginning of the lock-down period due to the COVID-19 pandemic. The MEWMA chart signalled a change just two days after a Level 4 "no travel" warning was issued.

Despite the benefits of the TERGM, such as the incorporation of the temporal dimension and representation of the network in terms of its sufficient statistics, there

are several considerable drawbacks. Other than the difficulty to determine a suitable combination of the network terms, the model is not applicable for networks of large size (Block et al. 2018). Furthermore, the temporal dependency statistics in the TERGM depend on the selected temporal lag and the size of the time window over which the data is modelled (Leifeld and Cranmer 2019). Thus, the accurate modelling of the network strongly relies on the analyst's knowledge about its nature. A helpful extension of the approach would be the implementation of the STERGM. In this case, it could be possible to subdivide the network monitoring into two distinct streams, so that the interpretation of changes in the network would become clearer.

Considering a network where the number of vertices differs over time, the current TERGM framework would model it by either removing or incorporating particular nodes as structural zeroes. However, the development of alternative solutions to address this issue (for instance, Krivitsky et al. (2011) introduce an offset term for the ERGM), as well as the expansion of the presented approach for monitoring the node composition, is subject to future research.

Another topic that demands additional research is the determination of cases when it is reliable to use the averaged network statistics $\hat{s}_t$ to construct the monitoring procedure and not the parameter estimates $\hat{\theta}_t$, as their calculation is more complex than of $\hat{s}_t$. Also, it would be beneficial to consider other estimators to compute $\hat{s}_t$ and compare their effectiveness to detect anomalies. In addition, the construction of a generator that simulates the network time series directly from a TERGM with the desired configuration would enhance the further analysis of the considered methods and support the development of novel approaches.

Concerning the multivariate control charts, there are also some aspects to regard. Referring to Montgomery (2009), the multivariate control charts perform well if the number of process variables is not too large, usually up to 10. Also, a possible extension of the procedure is to design a monitoring process when the values for Σ can vary between the in-control and out-of-control states. Whether this factor would beneficially enrich the surveillance remains open for future research.

In this paper, we customise the application using simulation methods to calibrate the charts. Hence, further development of adaptive control charts is interesting as they could remarkably improve the performance of the anomaly detection (cf. Sparks and Wilson 2019).

Availability of data and material For the empirical study, we used the daily flight data published by the Bureau of Transportation Statistics of the United States. The simulation study was performed on the cluster system at the Leibniz University of Hannover and the simulated data are available upon request.

Code availability All analyses were done using open-source software. Code is available upon request.

Declarations

Conflict of interest The authors declare that they have no conflict of interest.

Ethics approval There are no specific ethical aspects to the studies.

Consent to participate Not applicable.

Consent for publication Yes.

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

Akoglu L, Tong H, Koutra D (2014) Graph-based anomaly detection and description: a survey. Data Mining Knowl Disc 29(3):626–688

Alwan LC (1992) Effects of autocorrelation on control chart performance. Commun Stat Theory Methods 21(4):1025–1049

Amaral LAN, Scala A, Barthelemy M, Stanley HE (2000) Classes of small-world networks. Proc Natl Acad Sci 97(21):11149–11152

Basseville M, Nikiforov IV (1993) Detection of abrupt changes: theory and application, vol 104. Prentice Hall Englewood Cliffs

Block P, Koskinen J, Hollway J, Steglich C, Stadtfeld C (2018) Change we can believe in: comparing longitudinal network models on consistency, interpretability and predictive power. Social Netw 52:180–191

Butts CT (2008) A relational event framework for social action. Sociol Methodol 38(1):155–200

Cannings C, Penman D (2003) Models of random graphs and their applications. Stoch Process Modell Simul 21:51–91

Carrington PJ, Scott J, Wasserman S (2005) Models and methods in social network analysis, vol 28. Cambridge University Press, Cambridge

Chen CYH, Härdle WK, Okhrin Y (2019) Tail event driven networks of SIFIs. J Economr 208(1):282–298

Crosier RB (1988) Multivariate generalizations of cumulative sum quality-control schemes. Technometrics 30(3):291–303

Das H, Mishra SK, Roy DS (2013) The topological structure of the Odisha power grid: a complex network analysis. IJMCA 1(1):012–016

Farahani EM, Baradaran Kazemzadeh R, Noorossana R, Rahimian G (2017) A statistical approach to social network monitoring. Commun Stat Theory Methods 46(22):11272–11288

Fonseca-Pedrero E (2018) Network analysis in psychology. Papeles del Psicólogo 39(1):1–12

Frank O (1991) Statistical analysis of change in networks. Statistica Neerlandica 45(3):283–293

Frank O, Strauss D (1986) Markov graphs. J Am Stat Assoc 81(395):832–842

Handcock MS (2003) Assessing degeneracy in statistical models of social networks. Working Paper No. 39, Center for Statistics and the Social Sciences, University of Washington, Seattle

Hanneke S, Fu W, Xing EP (2010) Discrete temporal models of social networks. Electron J Stat 4:585–605

He R, Zheng T (2015) GLMLE: graph-limit enabled fast computation for fitting exponential random graph models to large social networks. Social Netw Anal Mining 5(1):8

Hosseini SS, Noorossana R (2018) Performance evaluation of EWMA and CUSUM control charts to detect anomalies in social networks using average and standard deviation of degree measures. Qual Reliab Eng Int 34(4):477–500

Hunter DR, Goodreau SM, Handcock MS (2008) Goodness of fit of social network models. J Am Stat Assoc 103(481):248–258

Jackson M (2016) The past and future of network analysis in economics. In: The Oxford handbook of the economics of networks

Johnson RA, Wichern DW (2007) Applied multivariate statistical analysis. Pearson Prentice Hall, Upper Saddle River, New Jersey

Joseph J, Pignatiello J, Runger GC (1990) Comparisons of multivariate CUSUM charts. J Qual Technol 22(3):173–186

Kenett RS, Pollak M (2012) On assessing the performance of sequential procedures for detecting a change. Qual Reliab Eng Int 28(5):500–507

Kolaczyk ED (2009) Statistical analysis of network data. Springer Series in Statistics

Kolaczyk ED, Krivitsky PN (2015) On the question of effective sample size in network modeling: an asymptotic inquiry. Stat Sci Rev J Inst Math Stat 30(2):184

Krivitsky PN, Handcock MS, Morris M (2011) Adjusting for network size and composition effects in exponential-family random graph models. Stat Methodol 8(4):319–339

Krivitsky PN, Handcock MS (2014) A separable model for dynamic networks. J Roy Stat Soc Ser B (Stat Methodol) 76(1):29–46

Leifeld P, Cranmer SJ (2019) A theoretical and empirical comparison of the temporal exponential random graph model and the stochastic actor-oriented model. Netw Sci 7(1):20–51

Leifeld P, Cranmer SJ, Desmarais BA (2018) Temporal exponential random graph models with btergm: estimation and bootstrap confidence intervals. J Stat Softw 83(6)

Leitch J, Alexander KA, Sengupta S (2019) Toward epidemic thresholds on temporal networks: a review and open questions. Appl Netw Sci 4(1)

Liu RY (1995) Control charts for multivariate processes. J Am Stat Assoc 90(432):1380–1387

Liu Y, Liu L, Yan Y, Feng H, Ding S (2019) Analyzing dynamic change in social network based on distribution-free multivariate process control method. Comput Mater Continua 60(3):1123–1139

Lowry CA, Woodall WH, Champ CW, Rigdon SE (1992) A multivariate exponentially weighted moving average control chart. Technometrics 34(1):46–53

Lu CW, Reynolds MR Jr (1999) Control charts for monitoring the mean and variance of autocorrelated processes. J Qual Technol 31(3):259–274

Lu CW, Reynolds MR Jr (2001) Cusum charts for monitoring an autocorrelated process. J Qual Technol 33(3):316–334

McCulloh I, Carley KM (2011) Detecting change in longitudinal social networks. Tech. rep, Military Academy West Point NY Network Science Center (NSC)

Montgomery DC (2009) Introduction to statistical quality control. John Wiley & Sons Inc

Montgomery DC, Mastrangelo CM (1991) Some statistical process control methods for autocorrelated data. J Qual Technol 23(3):179–193

Morris M, Handcock MS, Hunter DR (2008) Specification of exponential-family random graph models: terms and computational aspects. J Stat Softw 24(4):1548

Ngai HM, Zhang J (2001) Multivariate cumulative sum control charts based on projection pursuit. Stat Sinica 11:747–766

Noorossana R, Hosseini SS, Heydarzade A (2018) An overview of dynamic anomaly detection in social networks via control charts. Qual Reliab Eng Int 34(4):641–648

Page ES (1954) Continuous inspection schemes. Biometrika 41(1/2):100–115

Porzio GC, Ragozini G (2008) Multivariate control charts from a data mining perspective. Recent Adva Data Mining Enterp Data Algo Appl 6:413–462

Qiu P (2013) Introduction to statistical process control. CRC Press

Ranshous S, Shen S, Koutra D, Harenberg S, Faloutsos C, Samatova NF (2015) Anomaly detection in dynamic networks: a survey. Wiley Interdisc Rev Comput Stat 7(3):223–247

Robins G, Pattison P (2001) Random graph models for temporal processes in social networks. J Math Sociol 25(1):5–41

Robins G, Pattison P, Kalish Y, Lusher D (2007) An introduction to exponential random graph (p*) models for social networks. Social Netw 29(2):173–191

Runger GC, Willemain TR (1995) Model-based and model-free control of autocorrelated processes. J Qual Technol 27(4):283–292

Sadinejad S, Saghaei A, Rajabi F (2020) Monitoring of social network and change detection by applying statistical process: ERGM. J Optim Indus Eng 13(1):131–143

Salmasnia A, Mohabbati M, Namdar M (2019) Change point detection in social networks using a multivariate exponentially weighted moving average chart. J Inform Sci

Sambale H, Sinulis A (2018) Logarithmic Sobolev inequalities for finite spin systems and applications. arXiv preprint arXiv:1807.07765

Schmid W, Schöne A (1997) Some properties of the ewma control chart in the presence of autocorrelation. Ann Stat 25(3):1277–1283

Schweinberger M (2011) Instability, sensitivity, and degeneracy of discrete exponential families. J Am Stat Assoc 106(496):1361–1370

Schweinberger M, Krivitsky PN, Butts CT, Stewart J (2020) Exponential-family models of random graphs: Inference in finite-, super-, and infinite population scenarios. Stat Sci

Sheu SH, Lu SL (2009) Monitoring the mean of autocorrelated observations with one generally weighted moving average control chart. J Stat Comput Simul 79(12):1393–1406

Simpson SL, Bowman FD, Laurienti PJ (2013) Analyzing complex functional brain networks: fusing statistics and network science to understand the brain. Stat Surv 7:1–36

Snijders TAB, Pattison PE, Robins GL, Handcock MS (2006) New specifications for exponential random graph models. Sociol Methodol 36(1):99–153

Sparks R, Wilson JD (2019) Monitoring communication outbreaks among an unknown team of actors in dynamic networks. J Qual Technol 51(4):353–374

van Duijn MA, Gile K, Handcock MS (2009) Comparison of maximum pseudo likelihood and maximum likelihood estimation of exponential family random graph models. Social Netw 31(1):52–62

Ward MD, Stovel K, Sacks A (2011) Network analysis and political science. Ann Rev Polit Sci 14:245–264

Wasserman S, Pattison P (1996) Logit models and logistic regressions for social networks: I. An introduction to Markov graphs and p*. Psychometrika 61(3):401–425. https://doi.org/10.1007/BF02294547

Wilson JD, Stevens NT, Woodall WH (2019) Modeling and detecting change in temporal networks via the degree corrected stochastic block model. Qual Reliab Eng Int 35(5):1363–1378

Woodall WH, Ncube MM (1985) Multivariate cusum quality-control procedures. Technometrics 27(3):285–292

Yan T, Xu J (2013) A central limit theorem in the β-model for undirected random graphs with a diverging number of vertices. Biometrika 100(2):519–524

Yan T, Leng C, Zhu J (2016) Asymptotics in directed exponential random graph models with an increasing bi-degree sequence. Ann Stat 44(1):31–57

Zhang NF (1997) Detection capability of residual control chart for stationary process data. J Appl Stat 24(4):475–492