

RESEARCH

Open Access



A novel control chart scheme for online social network monitoring using multivariate nonparametric profile techniques

Arya Karami^{1*} and Seyed Taghi Akhavan Niaki¹

*Correspondence:
karami_ariya@ie.sharif.edu

¹ Department of Industrial Engineering, Sharif University of Technology, Tehran, Iran

Abstract

Online social networks have become indispensable in modern life, facilitating knowledge sharing, social communication, and business marketing. To gain a deeper understanding of individuals' behavior within social networks, researchers have undertaken essential analytical tasks such as change point detection. Recently, nonparametric change point detection methods have attracted researchers' attention due to their generality and flexibility. However, existing methods exhibit limitations, including overlooking network structure, reliance on case-based network attributes, and neglecting the dynamic nature of data, which may have correlations in evolving social networks. In this study, we propose a novel multivariate mixed-effects nonparametric profile control (MENPC) algorithm to address these limitations. The advantage of MENPC relies on its unique point of view in approaching network data, where it incorporates the dynamic nature of data into the monitoring process without assuming internal independence of networks over time. Additionally, it takes into account the network structure by considering both nodal and network-level attributes. Furthermore, by introducing an updating trick formula, the proposed algorithm simplifies computations, effectively balancing memory and speed for online monitoring. We evaluate the effectiveness of MENPC through comprehensive numerical experiments using the degree correlated stochastic block model to simulate interactions in evolving online social networks. The results demonstrate MENPC's superior performance in terms of expected detection delay, showcasing its accuracy and efficiency in comparison to competing approaches including Wilson, and eigenvalue methods. Applying MENPC to the Enron email network dataset further confirms its significant progress in social network monitoring, expanding its potential for various applications.

Keywords: Social network monitoring, Non-parametric profiles, Multivariate mixed effects nonparametric profile control, Network measures

Introduction

In today's digital age, extensive use of information technology enables diverse entities and individuals to communicate, resulting in the formation of intricate networks. Various types of networks have emerged, each serving distinct purposes across numerous applications in the field of applied network science, such as the Internet of Things

(IoT), protein–protein interaction networks, transportation networks, and social networks. Social networks play a fundamental role in knowledge dissemination, social communications, and business marketing. However, the massive volume and complexity of data generated on these platforms present challenges for social networks analysis (Sterchi et al. 2021; He and Wang 2023). Given the increasing trend of using social network applications across various aspects of human life, researchers have increasingly focused on social network analysis, particularly social network monitoring. Social network monitoring involves observing and analyzing activities, interactions, and trends within a social network platform. It encompasses tracking and gathering data related to user behavior, content sharing, engagement, and other relevant metrics. The primary goal of social network monitoring is to gain insights into the network's dynamics, user preferences, sentiment, and overall performance (Stevens et al. 2021a). Notably, social network change point detection finds applications in user behavioral analysis on online platforms, spammer detection, and community analysis. The ability to detect change points in social networks holds significant potential for understanding and improving user experiences, identifying anomalous behaviors, and enhancing network security. As social networks continue to evolve and influence various aspects of our lives, effective monitoring and analysis become imperative for making informed decisions and ensuring positive outcomes for users and businesses alike (Kumar et al. 2023; Dey et al. 2023).

Furthermore, there have been significant advancements in network systems, particularly in social science, shifting away from traditional approaches that treated networks as either static or dynamic entities. Instead, emerging models now allow networks to exhibit various changes over time. Efforts to model time-varying processes in social networks, such as information diffusion, reciprocity, and leadership, have led to a diverse range of modeling techniques. Consequently, there is a growing need to develop approaches for change point detection tasks in social networks (Camacho et al. 2020). Numerous researchers have emphasized the importance of change-point detection in social networks (Wilson et al. 2019; Salmasnia et al. 2020). Identifying and analyzing change points in social networks hold immense significance, as they can reveal crucial transitions and shifts in network dynamics, shedding light on evolving patterns of interactions and behaviors. By uncovering such changes, researchers can better understand the underlying mechanisms that drive network evolution and improve predictive modeling and decision-making processes. The exploration of change point detection methods in social networks continues to captivate researchers in the fields of statistical process control, network monitoring, and social science, as they strive to unravel the complexities and nuances of these dynamic and interconnected systems.

Traditionally, change detection and monitoring have been prevalent in manufacturing processes, aiming to recognize deviations from normal operating conditions. Two significant kinds of change points can occur in in-control processes: changes in the statistical distribution and changes in the distribution's parameters (Reis and Gins 2017). For example, changes may occur in the distribution of a variable of interest in a water treatment process. In addition, monitoring parameters like mean or variance can detect various changes, such as single-step shifts, multiple-step changes, sporadic changes, monotonic changes, and linear or nonlinear trend changes in the process distribution. The problem of change-point detection in social networks has garnered significant

attention from researchers (P Peixoto and Gauvin 2018). They have recognized the common categorization of change point detection methods in social networks, leading to two distinct methodologies: model-based approaches focusing on changes in the network model, and free-model approaches targeting changes in the network structure (Stevens et al. 2021a).

In model-based approaches there is the assumption that the static or streaming network data is generated from a specified network model. Thus, the network model is a prespecified information in the network monitoring task. Therefore, the data should be tested for its randomness to judge whether it's generated from the prespecified network model or a change point is recognized. Additionally, the network's structure is of interest in the free-model approaches, in which there are time series of network-related variables, i.e., nodes and edges, which might be used for change point detection. Researchers have adapted industrial process monitoring methods, including the Shewhart control chart, cumulative sum chart (CUSUM), exponentially weighted moving average (EWMA), and the Bayesian approaches in social network monitoring (Noorossana et al. 2011). While there are various methods for change point detection in industrial processes, a commonly accepted procedure for social network change point detection tasks has not yet been established. In this regard, researchers continue to propose various methods for change point detection tasks in social networks, leveraging network characteristics and conducting case studies. The ongoing efforts in this field indicate its growth and potential for further advancements.

Reviewing the related works in social network change point detection reveals a significant research gap characterized by the following issues: (1) neglecting the network structure, where the state of the network is summarized into a vector based on node states, leading to the loss of some information (Wang et al. 2023). (2) overreliance on case-based attributes, wherein certain network-level measures such as centrality measures are considered according to the case study, while features of nodes are neglected (Salmasnia et al. 2020). Furthermore, most research often disregards the dynamic nature of online social networks, resulting in: (3) the unrealistic assumption of independence among network snapshots over time, despite real social networks exhibiting correlation in their evolution (Wilson et al. 2019). These limitations extend to (4) high computational costs and poor scalability with large networks, resulting in low sensitivity for detecting fast change points (Hazrati-Marangaloo and Noorossana 2021). Addressing the four mentioned challenges is crucial for advancing the state-of-the-art in social network change point detection and fostering more accurate and efficient monitoring capabilities. To tackle these challenges, this study proposes a novel multivariate mixed-effects nonparametric profile control (MENPC) algorithm for change point detection in online social networks. MENPC tackles these limitations through the following approaches: (1) It considers both a function for capturing the effects of network structure and another function for node effects within the multivariate mixed-effects model. (2) As a multivariate approach, it utilizes both nodal and network topological measures over time, including node degrees, vertex count, edge count, network centrality degree, network density, triangle count, and network betweenness, which can also be supplemented by other user-defined network measures. (3) MENPC seamlessly integrates the dynamic nature of data into constructing the multivariate Mixed-Effects model, without

assuming internal network independence across time frames. (4) Additionally, MENPC benefits from recursive equations to simplify computations and improve efficiency.

The main contributions of this study are as follows: (1) Monitoring by using both nodal and network attributes to consider the features of social networks, allowing for broad applicability in varying applications. (2) Scalability of the proposed algorithm for large social networks achieved through a novel updating trick formula, ensuring both memory and time efficiency. (3) Accuracy in fast detection of change points, as evidenced by the ARL and EDD criteria. These contributions pave the way for more effective and comprehensive social network change point detection and monitoring.

The rest of the paper is organized as follows: section "[Literature review](#)" reviews the related works in statistical change point detection. Section "[Problem definition](#)" specifies the scope of the change points detection problem and introduces the assumptions that are used for profile monitoring. Section "[Methodology](#)" provides the methodology proposed in this paper for change point detection in social networks. Additionally, experimental results on both synthetic and real-world datasets are presented in section "[Experiments](#)". Finally, the conclusion and suggestions for further research avenues are outlined in section "[Practical implications and insights](#)".

Literature review

In the realm of social network analysis (SNA), human activities can be modeled by nodes and edges where the nodes represent individuals and the edges present communications among them. In this regard, networks are potent means for representing human behaviors and making it possible for researchers to study, analyze, and predict varying events, which might indicate occurrences in the real world. Moreover, in other domains such as information technology, biology, sensors network, transportation, and supply chain network, it has shown its capabilities, which made network analysis a growing interdisciplinary field of science for the last decades (Khalilzadeh et al. 2020; Hazrati-Marangaloo and Noorossana 2021). The scope of the problem and contextual constraints inspires the network type and its characteristics, for which there are different types: directed or undirected, weighted or unweighted, static or dynamic, and deterministic or random.

Recently, the problem of social network change point detection has attracted the attention of many researchers in both network science and statistical process control areas (Stevens et al. 2021a). There are primarily two approaches to addressing this problem: model-based and free-model approaches.

The model-based approach treats the network model as a fixed entity for network change point detection. In this approach, the edges between nodes represent an unobserved conditional relationship between some network variables. Testing the obtained data can show whether there is a change point and if so, detect changes based on the model's deviation or the network data distribution. Such a method is similar to the traditional ways used for change point detection in time series such as the likelihood ratio test. In this regard, (Rajabi et al. 2020) developed a method for monitoring changes in social networks through the application of the exponential random graph model (ERGM) and statistical process control techniques. This method addresses the inherent complexity of social network data by utilizing the flexible framework provided by ERGM to capture the intricate dependencies among tie variables. However, the reliance

on the ERGM model poses a potential limitation, as its assumptions may not fully encapsulate the complexity of real-world social networks. Moreover, while the results indicate the superiority of exponentially weighted moving average (EWMA) charts over Hotelling's T^2 , they are suitable for univariate models. The incorporation of multivariate vectors necessitates theoretical development. Additionally, the lack of utilization of nodal attributes in this method limits its generality for application in various social networks. Furthermore, the proposed method is suitable for static cases, and for dynamic social networks, it is preferable to use temporal ERGM (TERGM). Focusing on dynamic networks, (Ghoshal et al. 2022) developed a parallel algorithm for computing a newly defined anomaly score to detect change points in weighted graphs, adapted from the mixed membership stochastic block model (MSBM), where the weights represent the level of interactions among individuals. They integrated the overlapping community structure into their change point detection algorithm and demonstrated that focusing solely on nodes with multiple community memberships is sufficient for effective change detection. Experimental results on both synthetic data and real-world network data (specifically, the Enron Email dataset) indicate both the computational cost and precision superiority of their proposed method over existing methods like SpotLight (Eswaran et al. 2018) and AnomRank (Yoon et al. 2019). However, the method may face challenges in scenarios where the network structure is highly dynamic or noisy, potentially leading to false positives or missed anomalies. There is a need to extend the method to consider the correlated nature of social networks. Additionally, while the parallel implementation enhances computational efficiency, it may require substantial computational resources, limiting its applicability in resource-constrained environments. Moreover, not utilizing nodal features in the change point analysis constrains the application of the proposed method. Framing the changepoint detection problem as a composite hypothesis testing task, (Sharpnack et al. 2013) proposed a spectral scan statistic (SSS) that utilizes the combinatorial Laplacian of the graph to detect anomalies efficiently. They have demonstrated that the SSS outperforms existing methods such as χ^2 testing and naive testing based on edge thresholding, particularly when the signal-to-noise ratio is low. However, scalability could be a challenge for the SSS when applied to extremely large graphs. Moreover, the evaluation of the SSS primarily relies on simulated data from balanced binary trees, the two-dimensional lattice and the Kronecker graphs, which may not fully capture the complexities of real-world scenarios. Real-world graph-structured data often exhibit additional noise, heterogeneity, correlation and dynamic behavior that may not be adequately addressed by theoretical assumptions or simulated experiments. Adapting techniques from industrial process monitoring, (Salmasnia et al. 2020) proposed a multivariate exponentially weighted moving average (MEWMA) chart for social network monitoring. This method considers four correlated network attributes: density, degree centrality, betweenness, and closeness. In addition, their experimental results demonstrate the superiority of their proposed method against univariate control charts. However, limitation of their work lies in the assumption of a static network, where the number of members (nodes) remains constant throughout the monitoring cycle. This assumption may not hold in dynamic social networks, where nodes may join or leave over time, thereby altering the network structure. Moreover, the research assumes an undirected and unweighted network, which may oversimplify the complexity of

real-world social interactions where relationships can be directional and carry varying degrees of importance. Additionally, the evaluation of the proposed method relies solely on simulated data, potentially limiting its ability to capture the nuanced dynamics of real social networks.

Alongside with the first stream of model-based approaches, the second stream of free-model approaches attracted the researchers' attention due to their generalization capabilities. In the free-model approach, there is no assumption for the network model, and the changes in the network structure are of interest. For example, the monitoring may occur by observing time series of variables such as the number of nodes, edges, or some network-related attributes, e.g., centrality measures for the network. Farahani et al. (2017) developed both MEWMA and multivariate cumulative sum (MCUSUM) charts to monitor interactions in networks following Poisson distribution. They considered the correlations among network measures, the effect of node attributes for establishing connections, and the number of relationships between nodes in their proposed method. The experimental evaluation illustrated that the MEWMA chart is more efficient than the MCUSUM chart, for their proposed monitoring statistic. However, the assumption of a Poisson distribution for the number of interactions in social networks limits the application of the proposed method, as many social networks exhibit reciprocity behavior among their members. Additionally, although the researchers attempted to account for the correlation of network measures, the model they used for generation independently generates network interactions.

Additionally, considering the applications of attributed social networks, Fotuhi et al. (2023) investigated attributed social networks based on categorical data analysis using the generalized estimating equations (GEE) approach. They incorporated the autocorrelation of social networks into their proposed method, and after applying the GEE approach, the obtained model parameters were monitored through the MEWMA method and Hotelling's T² control chart. Although the conducted experiments showed that the MEWMA method has better performance, the proposed method relies heavily on simulated studies, which may not fully capture the complexity and variability present in real-world social networks. Additionally, the proposed method focuses on monitoring social networks based on categorical attributes and contingency tables, which may not fully capture the richness of interactions present in real social networks. This limitation suggests that the method may not be suitable for all types of social networks.

Moreover, Anastasiou et al. (2022) presented the cross-covariate isolate detection (CCID) method for change point detection in functional connectivity (FC) networks. It can detect change points in the second-order structure of multivariate time series. The CCID algorithm uses the scaled CUSUM statistic in an aggregated form considering only those CUSUM statistics over multivariate time series which pass a prespecified threshold. The experiments demonstrated the superiority and novelty of the proposed algorithm over methods such as sparsified binary segmentation (SBS) (Cho and Fryzlewicz 2015) and Factor with common components (Factor com) (Barigozzi et al. 2018), typically used in detecting change points in biological data, in terms of computational costs and sensitivity to small signals. However, while simulated data may offer insights into the performance of the CCID method, it may not fully capture the complexities and variability present in real-world fMRI or other neuroimaging data. Additionally, the

comparison with existing methods may not encompass all possible change scenarios, potentially limiting the generalizability of the findings. Moreover, it requires significant computational resources, particularly with high-dimensional data or numerous regions of interest (ROIs), which could present challenges in real-world scenarios with extensive datasets or limited computational resources.

Industrial process monitoring using profiles has garnered significant attention due to its various applications (Noorossana et al. 2011). One of the early efforts for monitoring social networks through profiles dates back to (Wasserman and Pattison 1996), which proposed logistic regression for social network modeling. Continuing in this way, (Azarnoush et al. 2016) developed a new approach based on logistic regression and node attributes for social network monitoring purposes. It should be noted that detecting changes in logistic regression models falls into the profile monitoring methods. Additionally, profile monitoring consists of two phases: Phase I involves analyzing in-control data to estimate the profile parameters and design the control chart, while Phase II aims to rapidly detect changes in the process using the designed control chart. While the proposed approach offers flexibility in detecting anomalies arising from different network edge-formation mechanisms, it also assumes that the relationships between entities can be fully captured by vertex attributes. However, in real-world scenarios, the formation of network edges may be influenced by various other factors, including the autocorrelation of networks. Hazrati-Marangaloo and Noorossana (2021) proposed a nonparametric change detection approach based on eigenvalues of adjacency matrices to monitor network streams, offering advantages over parametric methods by avoiding strict distributional assumptions. By utilizing a sliding window of reference networks and comparing distributions of eigenvalues, the method can detect structural changes in network connectivity, including changes in communication rates and community structure, without relying on predefined models or assumptions about network dynamics. However, while the proposed method offers flexibility and robustness, it may come with certain limitations. One potential disadvantage is computational complexity, especially when dealing with large-scale networks or high-dimensional data for calculating eigenvalues. Additionally, the effectiveness of nonparametric methods may depend on the choice of hyperparameters, such as the window size or divergence threshold, which could introduce subjectivity and require careful tuning for optimal performance. In addition, the lack of utilization of nodal attributes in this method limits its generality for application in various social networks.

Wilson et al. (2019) proposed an approach to detect structural changes in dynamic networks. By utilizing degree corrected stochastic block model (DCSBM), which accounts for heterogeneous connectivity and community structure, the proposed monitoring method demonstrates flexibility in handling complex network dynamics. The proposed set of statistics, Shewhart control charts and exponentially weighted moving average (EWMA) control charts used for change point detection. The study demonstrates the utility of this framework through its application to real-world datasets. Moreover, experiments show that the proposed method effectively detects local and global structural changes in dynamic networks, providing valuable insights into the behavior of system. While the proposed methodology presents advantages such as its flexibility and ability to detect significant changes accurately, it has limitations such as computational

complexity associated with calculation of monitoring statistics in large-scale network data. Furthermore, the effectiveness of the surveillance strategy may depend on the choice of monitoring statistic, requiring careful tuning for optimal performance.

Recently, researchers have proposed nonparametric models for change point analysis which have the capability to capture interdependency of evolving networks over time while considering common features of social networks (Qiu et al. 2010; Klyushin and Martynenko 2021; Yeganeh et al. 2022; Wang et al. 2022). While traditional parametric profile approaches are widely used for monitoring, they may be inadequate when the relationship is too complex to be accurately described by a parametric model. In such cases, nonparametric profile monitoring offers a promising alternative. Qiu et al. (2010) developed a novel approach to nonparametric profile monitoring by incorporating mixed-effects modeling and local linear kernel smoothing into the exponentially weighted moving average (EWMA) control scheme. The proposed method allows for the detection of structural changes or shifts in nonparametric profiles while accounting for within-profile correlations. Through numerical studies and application examples, the proposed control chart demonstrates its effectiveness in detecting step shifts and certain profile drifts in various scenarios. However, the proposed method is developed for univariate cases, and the nodal attributes are not considered in the model, limiting the generalization of the method.

The two streams of model-based and free-model approaches for change point analysis in social networks, have experienced fluctuations. However, tools for network monitoring, change point detection, and data analysis have consistently prevailed and found application across a wide range of domains (Schweinberger et al. 2021). Upon reviewing the existing literature on social network change point detection, a significant research gap emerges, marked by several key issues. Firstly, there is a tendency to overlook the network structure, wherein the network's state is condensed into a vector based on node states, consequently losing valuable information. Secondly, there is an overreliance on case-based attributes, where certain network-level metrics such as centrality measures are selectively considered, leaving out crucial node features. Additionally, many studies tend to ignore the dynamic nature of online social networks, making the unrealistic assumption of independence among network snapshots over time, despite real networks showing correlated evolution. These limitations are further compounded by the high computational costs and poor scalability observed with large networks, resulting in a diminished sensitivity for detecting rapid change points. Addressing these challenges is imperative for advancing the state-of-the-art in social network change point detection and enhancing monitoring capabilities.

Table 1 provides a comparison and summary of the literature methods, outlining their respective capabilities. As evident from the table, there exists a research gap for a multivariate nonparametric mixed-effects (NME) approach while also incorporating nodal attributes for social network monitoring.

To address the aforementioned issues and research gap, this study proposes a novel multivariate mixed-effects nonparametric profile control (MENPC) algorithm tailored for change point detection in online social networks. In other words, we contribute to the second stream, which is the model-free approach of change point detection. To this aim, this study proposes nonparametric profiles (NP) by considering common features

Table 1 Comparison of literature methods for social network monitoring

Method	Free of network model	Nonparametric profile approach	Not assuming i.i.d. conditions	Considering correlation in dynamic network	Multivariate approach	Node features	Network attributes
ERGM based	×	×	×	×	×	×	✓
Community Based Approach for MSBM	×	×	×	×	×	×	✓
SSS	×	×	×	×	×	×	✓
Salmasnia Method	×	×	×	×	✓	×	✓
Farahani method for Poisson Networks	×	×	×	✓	✓	✓	✓
GEE	×	×	×	✓	✓	✓	✓
CCID	✓	×	✓	✓	×	×	✓
Azarnoush method	✓	×	×	×	✓	✓	×
Eigenvalue	✓	×	✓	×	×	×	×
Wilson	✓	×	×	×	✓	×	✓
NME	✓	✓	✓	✓	×	×	✓

of social networks, including sparsity, heterogeneity, and reciprocity. In addition, in Phase I, we utilize NP profiles for modeling the data and the least squares (LS) method (Craven and Islam 2011) for parameter estimation. Subsequently, the profiles are utilized for monitoring purposes in Phase II.

In short, as there is no attempt accomplished for online monitoring of social networks through multivariate NP profiles, which are naturally dynamic and should be appropriately incorporated into profiles, this paper contributes to both theoretical and practical aspects of social network monitoring. The novelty comes from the theory for which the profiles are constructed and its capability to consider social network features. In addition, the proposed tool can be used in industries and governments. Some potential applications of our proposed algorithm include monitoring communities in social networks for marketing and business insights, as well as tracking criminal activities, disease spread, etc. (Stevens et al. 2021b). The next section presents the problem of change point detection in dynamic social networks and provides its assumptions.

Problem definition

Statistical modeling and mathematical analysis of networks reveal many insights regarding complex systems. In this article, the data changes over the network are the subject of interest. There are different frameworks for presenting graphs, including exchangeable random measures on a plane, point processes, and adjacency lists (Todeschini et al. 2020). In this study, a discrete structure like an adjacency matrix, commonly used for network analysis, is employed. Table 2 provides the notation and parameter definition.

The following equations characterize the dynamic social network:

Table 2 Social network change point detection problem notation

Notation	Definition
$G(t)$	The network state at time t
$V(t)$	The set of network individuals at time t
$E(t)$	The set of network edges at time t
T	The number of periods
p	The number of nodes
e_{ijt}	The number of interactions between node i and node j at time t

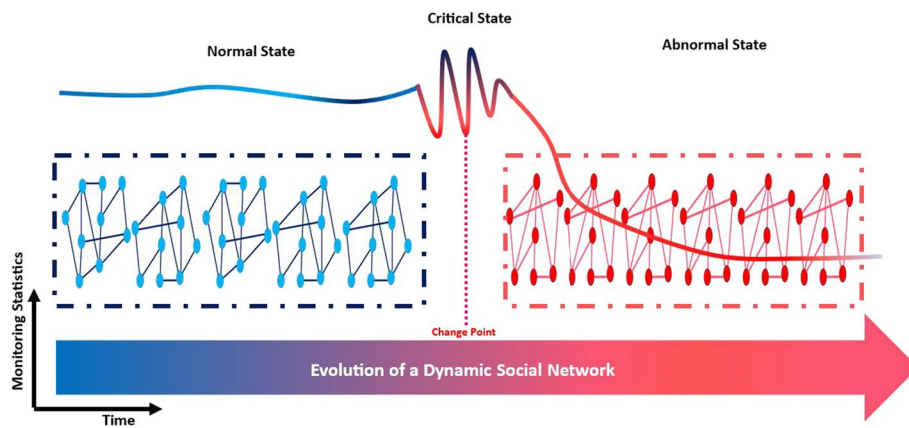


Fig. 1 Change point in a dynamic social network

$$G(t) = (V(t), E(t)) \quad t = 1, 2, \dots, T \tag{1}$$

$$V(t) = (v_1, v_2, \dots, v_p) \tag{2}$$

$$E(t) = (e_{12t}, \dots, e_{ijt}, \dots, e_{p(p-1)t}) \tag{3}$$

where $V(t)$ and $E(t)$ represent nodes and edges in the graph at period t , respectively. In addition, the relationship among nodes can be in the form of any communication facilities such as emails, messages, comments, reshare, or likes available on the platform of which they are a member.

Figure 1 illustrates the evolution of a dynamic social network, denoted as $G(t)$, over time. It depicts a sample change point in time, where the network state undergoes a transition. The occurrence of changes in social networks and individual interactions can be influenced by a wide range of factors, both external and internal. Among external factors, technological advancements, demographic shifts, and social trends can contribute to altering the dynamics of social networks (Stevens et al. 2021a). For example, technological advancements, like the emergence of new competing social platforms, can influence user preferences and behaviors, such as transitioning to a new platform or reducing daily usage. These changes often reflect in the network’s topological measures over time. Similarly, demographic shifts, such as alterations

in the age demographics or geographic distribution of users, can impact the structure and dynamics of social communities within the network. For instance, individuals with distinct local dialects may be less inclined to form new communities within the social network. Moreover, social trends like viral challenges or trending topics that garner widespread attention and participation within a social network can swiftly shape the type of content shared and the conversations taking place. These trends often spark a surge in user-generated content related to the trend, leading to increased engagement and alterations in interactions within the social network.

Moreover, internal factors, such as changes in platform algorithms, community management features, privacy concerns, and the influence of key users like influencers and celebrities play roles in driving changes within social networks (Jeske et al. 2018). Social platforms frequently update their algorithms to enhance user experience and personalize content. These algorithmic changes have a significant impact on the content users encounter, how it is prioritized, and its dissemination within the network. Consequently, these changes manifest in various aspects of the network, including nodes, edges, and network-level measures. Changes in community management features on social platforms, such as group functionalities, can also shape user behavior and interaction patterns within social networks. These alterations may influence the nature of discussions, practices regarding content sharing, and the overall dynamics of the community. Similarly, modifications to privacy settings can influence user trust and engagement on social platforms. These adjustments might lead to shifts in user behavior, resulting in changes in interaction patterns and content sharing practices. Furthermore, influencers and celebrities wield considerable influence over social network dynamics through their content, endorsements, and interactions. Their actions and endorsements have the power to set trends, amplify messages, and drive user engagement, thus shaping the overall tone and direction of conversations within the social network.

Adapting and responding to changes is crucial for social platforms to provide a positive and engaging user experience. Understanding the complexity and interconnectedness of these factors is essential for ensuring the continued growth and success of social networks.

This research endeavors to introduce a novel charting statistic for tackling the problem of detecting change points in dynamic social networks, considering the following challenges:

1. Incorporating both nodal and network-level attributes of the social network for change point detection.
2. The necessity of considering the interdependency among consecutive network states over time.
3. Ensuring scalability and accuracy in rapidly detecting change points over time.

Methodology

In this section, we begin by providing some contextual background. Then, a novel charting statistic for online monitoring of social networks is proposed.

Preliminaries in profile monitoring

Historically, quality engineers proposed profile monitoring techniques to monitor industrial processes in which one or more response variables have functional relation to some explanatory variables known as quality variables (Noorossana et al. 2011). The following equation represents the general form of profiles

$$y_i = f(x_i) + \varepsilon_i \quad i = 1, 2, \dots, n. \quad (4)$$

where y_i is the vector of measured values, $f(\cdot)$ presents the functional relationship, x_i is the vector of explanatory variables, ε_i is the vector of error term usually assumed white noise. Additionally, profiles can take on different functional forms, either parametric or nonparametric, including simple linear, multiple linear, polynomial profile, binary response, parametric nonlinear, and nonparametric nonlinear. Note that Eq. 4 can be reduced to a parametric form by considering $\tilde{\beta}$ as the vector of parameters. The following equation presents the reduced form.

$$y_i = f(x_i, \tilde{\beta}) + \varepsilon_i \quad i = 1, 2, \dots, n. \quad (5)$$

In parametric profiles, based on the type of profile, there are varying methods such as maximum likelihood estimator (MLE) (Meeker and Escobar 1994), ordinary least squares (OLS) (Craven and Islam 2011), and weighted least square (WLS) (Cohen and Migliorati 2017) methods to estimate the parameters. Furthermore, the literature on profile monitoring commonly addresses issues related to the dependency of profiles over time or the dependency between explanatory variables (Noorossana et al. 2011). The profiles exhibit dependency in two forms: either through between-profile correlation (BPC), indicating dependence between the profiles, or through within-profile autocorrelation (WPA), indicating that the measurements within a profile are not independent of each other. Note that BPC and WPA refer to the data and error term, not the profile parameters (Zhang et al. 2015). As a result, researchers often assume independence among profiles or variables in many cases (Noorossana et al. 2011; Salmasnia et al. 2020). In the context of online social network monitoring, the issue of dependency is crucial, as individuals' interactions naturally depend on previous connections. The section "[The proposed method for online monitoring of social networks](#)" proposes our method for tackling this issue.

Online profile monitoring

Generally, process monitoring is divided into two phases. In Phase I, the goal is to achieve stability and estimate model parameters. Once the parameters and baselines are established, Phase II involves monitoring online data to detect deviations from these baselines. To ensure desirable efficiency, it is essential to minimize the probability of type I and II errors for unstable points in Phase I. Additionally, quick detection of anomalies in Phase II is another critical factor for effective monitoring (Noorossana et al. 2011). Although it is true to say that Phase II conducts online monitoring due to the dynamic nature of processes, most research efforts do not address the dynamic nature of data in Phase I. In fact, one could argue that a method monitors

online processes while the technique itself possesses a dynamic nature. Therefore, it is crucial to update baselines periodically, a consideration incorporated in the proposed method outlined in this article.

The proposed method for online monitoring of social networks

In the literature on profile monitoring, it is common to assume i.i.d random errors in the model, although this assumption is not always accurate. Similarly, in networks such as social networks, there are often evident correlations among data, where the connections between nodes depend on the network's previous state or structure, as well as on the attribute values of certain nodes. This correlation within profile data poses a challenging issue in profile setup. To address this, researchers have proposed various methods (Noorossana et al. 2011; Zhang et al. 2015). One notable method is the nonparametric mixed effects (NME) model (Qiu et al. 2010), which is well-suited for longitudinal datasets. Importantly, the proposed NME model enables monitoring of social networks without imposing parameters, thereby increasing its applicability.

Assuming $x_{tij} \in [0,1]$, Eq. 6 presents a multivariate NME model (Qiu et al. 2010)

$$y_{tij} = g(x_{tij}) + f_t(x_{tij}) + \varepsilon_{tij} \quad (6)$$

where y_{tij} is the weight of the connection between nodes i and j at time point t , $g(\cdot)$ is the profile function for connections in the network, $f_t(\cdot)$ is the random effect term to characterize the variation of the profile from $g(\cdot)$, and ε_{tij} is the white noise. In addition, the assumption behind the proposed model is the independency of $f_t(\cdot)$ and ε_{tij} , and the $f_t(\cdot)$ is a realization of a zero-mean process with the following covariance function.

$$\gamma(x_1, x_2) = \mathbb{E}(f_t(x_1) * f_t(x_2)) \quad (7)$$

Because of NME model flexibility, it requires a relatively large amount of training data in model estimation. As mostly there is a large amount of in-control data available in social networks as well as industrial processes, the proposed model can effectively be used in these applications. Moreover, to construct a Phase-II control chart statistic, it is essential to have $g(\cdot)$, $f_t(\cdot)$ or $\gamma(\cdot)$, and σ^2 based on in-control data. In this regard, according to Wu and Zhang (2002) the combination of Linear Mixed Effects modeling and the Local Linear Kernel Smoothing method provides an efficient estimation of $g(\cdot)$, and by the same approach, leading to consistent estimates of $f_t(\cdot)$ and σ^2 . These estimations facilitate the proposal of the charting statistic.

Linear Mixed Effects modeling is a statistical technique used in nonparametric approaches for analyzing data that has both fixed and random effects. It is an extension of linear regression that is particularly useful when dealing with data that exhibit correlation or grouping structures (West et al. 2022). In addition, Local Linear Kernel smoothing is a non-parametric statistical technique used for estimating the underlying relationship between variables, when it is not assumed to be strictly linear and when there might be local variations in data. Implementing local linear kernel Smoothing involves selecting an appropriate kernel function and bandwidth, and then computing the weighted linear regression estimates for each target point (Zara et al. 2022).

With the inspiration of Wu and Zhang method (Wu and Zhang 2002) and extending it for the multivariate case, for each $\zeta \in [0,1]$, assuming $\widehat{g}(\zeta) = e_1^T \widehat{\beta}(\zeta)$, $\widehat{f}_t(\zeta) = e_1^T \widehat{\alpha}_t(\zeta)$ the Local Linear Mixed Effects model for $g(\zeta)$, and $f_t(\zeta)$ can be obtained by minimizing Eq. 8.

$$\sum_{t=1}^T \frac{1}{\sigma^2} \sum_{i=1}^v \sum_{j=1}^v \left[y_{tij} - z_{tij}^T(\beta + \alpha_t) \right]^2 K_h(x_{tij} - s) + \alpha_t^T D^{-1} \alpha_t + \ln|D| + n_t \ln(\sigma^2) \tag{8}$$

where x_{tij} is a p dimensional vector of explanatory variables, K is a symmetric kernel function, h is the bandwidth, $e_1 = (1,0)^T$ and $K_h(\cdot) = K(\cdot/h)/h$. Additionally, $\alpha_t \sim (0, D)$ is a random variable with mean 0 and variance D , and β is a two-dimensional vector of coefficients for obtaining $g(\cdot)$. Algorithm 1 proposes an iterative procedure for obtaining $g(\cdot)$ and $f_t(\cdot)$ of Eq. 6.

Algorithm 1. An iterative procedure to obtain $g(\cdot)$ and $f_t(\cdot)$

1	Set the initial values of D and σ^2 from the following formulae, and consider an arbitrary p dimensional s and $D_{(0)} = I_2$ $\sigma_{(0)}^2 = \frac{1}{T} \sum_{i=1}^T \frac{1}{v^2} \sum_{i=1}^v \sum_{j=1}^v [y_{tij} - \widehat{g}^{(P)}(x_{tij})]^2$ Where $\widehat{g}^{(P)}(\cdot)$ is standard LLKE constructed from the data
2	Estimate β and α_t at iteration $k \geq 0$ by the following formulae: $\widehat{\beta}^{(k)} = \left\{ \sum_{t=1}^T Z_t^T \Sigma_t Z_t \right\}^{-1} \left\{ \sum_{t=1}^T Z_t^T \Sigma_t y_t \right\}$ $\widehat{\alpha}_t^{(k)} = \left\{ Z_t^T K_t Z_t + \sigma_{(k)}^2 [D_{(k)}]^{-1} \right\}^{-1} \left\{ Z_t^T K_t (y_t - Z_t \widehat{\beta}^{(k)}) \right\}$ Where $Z_t = (z_{t11}, \dots, z_{tww})^T$, $y_t = (y_{t11}, \dots, y_{tww})^T$, $\Sigma_t = (Z_t D_{(k)} Z_t^T + \sigma_{(k)}^2 K_t^{-1})^{-1}$, and $K_t = \text{diag}\{K_h(x_{t11} - s), \dots, K_h(x_{tww} - s)\}$.
3	Update D and σ^2 based on $\widehat{\beta}^{(k)}$ and $\widehat{\alpha}_t^{(k)}$ $D_{(k+1)} = \frac{1}{T} \sum_{t=1}^T \widehat{\alpha}_t^{(k)} [\widehat{\alpha}_t^{(k)}]^T$ $\sigma_{(k+1)}^2 = \frac{1}{T} \sum_{t=1}^T \frac{1}{v^2} \left[y_t - Z_t (\widehat{\beta}^{(k)} + \widehat{\alpha}_t^{(k)}) \right]^T K_t [y_t - Z_t (\widehat{\beta}^{(k)} + \widehat{\alpha}_t^{(k)})]$
4	Repeat Steps 2 and 3 until the following condition is satisfied $\frac{\ D_{(0)} - D_{(l-1)}\ _1}{\ D_{(l-1)}\ _1} \leq \epsilon$ Where ϵ is a small positive value, i.e., $\epsilon = 10^{-4}$
5	After obtaining the estimates of β and α_t through Steps 1 to 4, the following equations can be defined for $g(\cdot)$, $f_t(\cdot)$ and $\gamma(\cdot)$ $\widehat{g}(\zeta) = e_1^T \widehat{\beta}(\zeta)$ $\widehat{f}_t(\zeta) = e_1^T \widehat{\alpha}_t(\zeta)$ $\widehat{\gamma}(\zeta_1, \zeta_2) = \frac{1}{T} \sum_{t=1}^T \widehat{f}_t(\zeta_1) \widehat{f}_t(\zeta_2), \zeta_1, \zeta_2 \in [0,1]$

After implementing the proposed procedure for multivariate data on in-control dataset, we can obtain an estimation of $g(\cdot)$ and $f_t(\cdot)$ in the NME model (Eq. 6) for the desired social network interaction data. The remaining task is to pave the way for Phase II monitoring through proposing the charting statistic.

Phase II monitoring using the charting statistic based on the proposed NME model

In Phase II, we assume that variance function $v^2(\cdot)$, $g(\cdot)$, and $f(\cdot)$ for the in-control dataset are known, and obtained from Algorithm 1. In this phase, online monitoring is a challenging task mainly for two reasons. The first is that correlations between profiles and within profiles increase computational time since the model considers historical data to tackle the correlation issue, in this regard we are proposing a computational trick

for the recursive computation of charting statistic. The second is related to the applicability of the GLR test in which the variation of design points is necessary, thus instead of that here according to Qiu et al. (2010) and extending it for the multivariate case the local weighted negative log likelihood function minimization is used to overcome the mentioned issue.

$$WL(a, b; s, \lambda, \tau) = \sum_{t=1}^{\tau} \sum_{i=1}^v \sum_{j=1}^v \frac{[y_{tij} - a - b(x_{tij} - s)]^2 K_h(x_{tij} - s)(1 - \lambda)^{\tau-t}}{v^2(x_{tij})} \tag{9}$$

where x_{tij} is a p dimensional vector of explanatory variables, λ is a weighting parameter such as the scheme of weights in the EWMA method and $v^2(\zeta) = \gamma(\zeta, \zeta) + \sigma^2$ is the variance function. According to the local weighted negative log likelihood function, it combines the weighting scheme of EWMA through the term $(1 - \lambda)^{\tau-t}$ and a local linear kernel estimator, simultaneously. In addition, in phase II of monitoring, the population profile $g(\cdot)$ can be defined through the minimization of WL. The following equation presents the $g(\cdot)$ for phase II.

$$\widehat{g}_{\tau,h,\lambda}(\zeta) = \frac{\sum_{t=1}^{\tau} \sum_{i=1}^v \sum_{j=1}^v U_{tij}^{(\tau,h,\lambda)}(\zeta) y_{tij}}{\sum_{t=1}^{\tau} \sum_{i=1}^v \sum_{j=1}^v U_{tij}^{(\tau,h,\lambda)}(\zeta)} \tag{10}$$

where the following equations are established.

$$U_{tij}^{(\tau,h,\lambda)}(\zeta) = \frac{(1 - \lambda)^{\tau-t} K_h(x_{tij} - s)}{v^2(x_{tij})} \{m_2^{(\tau,h,\lambda)}(\zeta) - (x_{tij} - s)m_1^{(\tau,h,\lambda)}(\zeta)\} \tag{11}$$

$$m_l^{(\tau,h,\lambda)}(\zeta) = \sum_{t=1}^{\tau} (1 - \lambda)^{\tau-t} \sum_{i=1}^v \sum_{j=1}^v (x_{tij} - s)^l K_h(x_{tij} - s) v^{-2}(x_{tij}) \quad l = 0, 1, 2 \tag{12}$$

Considering that $v^2(\cdot)$ function and $g_0(\cdot)$ both known from the in-control dataset (Phase I), replacing the transformation of Eq. 13 in Eq. 10 leads to Eq. 14. Clearly, in this case, the distribution of $\widehat{\xi}_{\tau,h,\lambda}(\zeta)$ would not depend on $g_0(\cdot)$.

$$\xi_{tij} = y_{tij} - g_0(x_{tij}) \tag{13}$$

$$\widehat{\xi}_{\tau,h,\lambda}(\zeta) = \frac{\sum_{t=1}^{\tau} \sum_{i=1}^v \sum_{j=1}^v U_{tij}^{(\tau,h,\lambda)}(\zeta) \xi_{tij}}{\sum_{t=1}^{\tau} \sum_{i=1}^v \sum_{j=1}^v U_{tij}^{(\tau,h,\lambda)}(\zeta)} \tag{14}$$

If the process is in control, $|\widehat{\xi}_{\tau,h,\lambda}(\zeta)|$ should be small values. Finally, the following $T_{\tau,h,\lambda}$ is the novel statistic for the multivariate MENPC which can be used for monitoring purposes.

$$T_{\tau,h,\lambda} = c_{0,\tau,\lambda} \int \frac{[\widehat{\xi}_{\tau,h,\lambda}(\zeta)]^2}{v^2(\zeta)} \Gamma_1(\zeta) d\zeta \tag{15}$$

$$c_{t_0,t_1,\lambda} = \frac{a_{t_0,t_1,\lambda}^2}{b_{t_0,t_1,\lambda}} \tag{16}$$

$$a_{t_0,t_1,\lambda} = \sum_{i=t_0+1}^{t_1} (1 - \lambda)^{t_1-i} n_i \tag{17}$$

$$b_{t_0,t_1,\lambda} = \sum_{i=t_0+1}^{t_1} (1 - \lambda)^{2(t_1-i)} n_i \tag{18}$$

where Γ_1 is a prespecified density function. Moreover, practically the discretized version of the statistic $T_{\tau,h,\lambda}$ is in the form of the following equation.

$$T_{\tau,h,\lambda} = \frac{c_{0,\tau,\lambda}}{n_0} \sum_{k=1}^{n_0} \frac{[\widehat{\xi}_{\tau,h,\lambda}(\varphi_k)]^2}{v^2(\varphi_k)} \tag{19}$$

where $\{\varphi_k, k = 1, \dots, n_0\}$ are some prespecified design points from Γ_1 . Then, by setting a proper control limit L to achieve a specific ARL_0 the proposed multivariate mixed-effect nonparametric profile control (MENPC) chart can be used for social network monitoring. The Block Bootstrap resampling procedures (Galvao et al. 2024) are used here to obtain the limit L . If the $T_{\tau,h,\lambda} \geq L$ it demonstrates a change point in the social network.

A computational trick for online monitoring of social networks

In general, online social network monitoring consists of many profiles, where fast implementation is essential. Therefore, since computing the test statistic is time-consuming as it uses past social network data, the implementation of the following updating formulae simplifies the computation.

$$\widetilde{m}_l^{(\tau,h)}(s) = \sum_{i=1}^v \sum_{j=1}^v \frac{(x_{tij} - s)^l K_h(x_{tij} - s)}{v^2(x_{tij})} \quad l = 0,1,2 \tag{20}$$

$$\widetilde{q}_l^{(\tau,h)}(s) = \sum_{i=1}^v \sum_{j=1}^v \frac{(x_{tij} - s)^l K_h(x_{tij} - s) y_{tij}}{v^2(x_{tij})} \quad l = 0,1 \tag{21}$$

Then, the following recursive equations are obtained.

$$m_l^{(\tau,h,\lambda)}(s) = (1 - \lambda)m_l^{(\tau-1,h,\lambda)}(s) + \widetilde{m}_l^{(\tau,h)}(s) \quad l = 0,1,2 \tag{22}$$

where $m_l^{(0,h,\lambda)}(s) = 0$ for $l = 0,1,2$. Additionally, let

$$q_l^{(\tau,h,\lambda)}(s) = (1 - \lambda)q_l^{(\tau-1,h,\lambda)}(s) + \widetilde{q}_l^{(\tau,h)}(s) \quad l = 0,1 \tag{23}$$

where $q_l^{(0,h,\lambda)}(s) = 0$ for $l = 0,1$. Moreover, we can obtain $\widehat{g}_{\tau,h,\lambda}(\zeta)$

$$\widehat{g}_{\tau,h,\lambda}(s) = \left[M^{(\tau,h,\lambda)} \right]^{-1} \{ (1-\lambda)^2 M^{(\tau-1,h,\lambda)} \widehat{g}_{\tau-1,h,\lambda}(s) + \left[\widetilde{q}_0^{(\tau,h)} m_2^{(\tau,h,\lambda)} - \widetilde{q}_1^{(\tau,h)} m_1^{(\tau,h,\lambda)} \right] + (1-\lambda) [q_0^{(\tau-1,h,\lambda)} \widetilde{m}_2^{(\tau,h)} - q_1^{(\tau-1,h,\lambda)} \widetilde{m}_1^{(\tau,h)}] \} \tag{24}$$

where $M^{(\tau,h,\lambda)}$ is as follows.

$$M^{(\tau,h,\lambda)} = m_2^{(\tau,h,\lambda)} m_0^{(\tau,h,\lambda)} - [m_0^{(\tau,h,\lambda)}]^2 \tag{25}$$

Therefore, using the proposed updating formulae, the multivariate Mixed-Effect Nonparametric Profile Control (MENPC) chart can be constructed through the procedure shown in Algorithm 2.

Algorithm 2. Online Social Network Monitoring

1	Specify L, h, λ , and compute the quantities $\widetilde{m}_l^{(\tau,h)}(s)$ for $l = 0, 1, 2$ and $\widetilde{q}_l^{(\tau,h)}(s)$ for $l = 0, 1$ by the historical data, based on Eqs. 20 and 21
2	Update $m_l^{(\tau,h,\lambda)}(s)$ for $l = 0, 1, 2$ and $q_l^{(\tau,h,\lambda)}(s)$ for $l = 0, 1$ through Eqs. 22 and 23
3	Compute $\widehat{g}_{\tau,h,\lambda}(s)$ from Eq. 24
4	Compute the test statistic $T_{\tau,h,\lambda}$ from Eq. 19, by $\xi_{tij} = y_{tij} - g_0(x_{tij})$, and compare it to the control limit L
5	If $T_{\tau,h,\lambda} \geq L$ alert that a change point is detected

The following flowchart illustrates the workflow of the proposed MENPC method.

Figure 2 illustrates the workflow of MENPC. Initially, parameters such as the control limit L , window size h , and decay factor λ , are specified. Then, historical data is used to compute $m_l^{(\tau,h,\lambda)}(s)$ and $q_l^{(\tau,h,\lambda)}(s)$ for different levels l using Eqs. 20 and 21, respectively. Following this, $m_l^{(\tau,h,\lambda)}(s)$ and $q_l^{(\tau,h,\lambda)}(s)$ are updated using Eqs. 22 and 23. It should be noted that the temporary variables are proposed for the efficient calculation of population profile $g(\cdot)$, obtained by minimizing the multivariate local weighted negative log likelihood function presented in Eq. 9. Subsequently, the estimated function $\widehat{g}_{\tau,h,\lambda}(s)$ is calculated using Eq. 24. The test statistic $T_{\tau,h,\lambda}$ is then computed from Eq. 19. Finally, if the test statistic exceeds the control limit L , an alert is raised to indicate the detection of a change point; otherwise, normal network data at time τ is observed. It is assumed that the user can set conditions for updating the control limit; for example, it can be updated periodically. Additionally, after detecting a change point and informing the user, a decision based on the user’s utility is made regarding whether to continue the monitoring process or stop it for further consideration. This systematic process facilitates effective monitoring of online social networks for potential changes.

MENPC chart parameters

The MENPC chart requires some parameters before implementing the control limits, similar to different change point detection approaches. Therefore, the bandwidth h , the smoothing weight λ , the Kernel function $K(\cdot)$, and the determination of the control limit L should be specified by the user. Additionally, there are some guidelines based on the methods in the literature and the obtained results of conducted experiments.

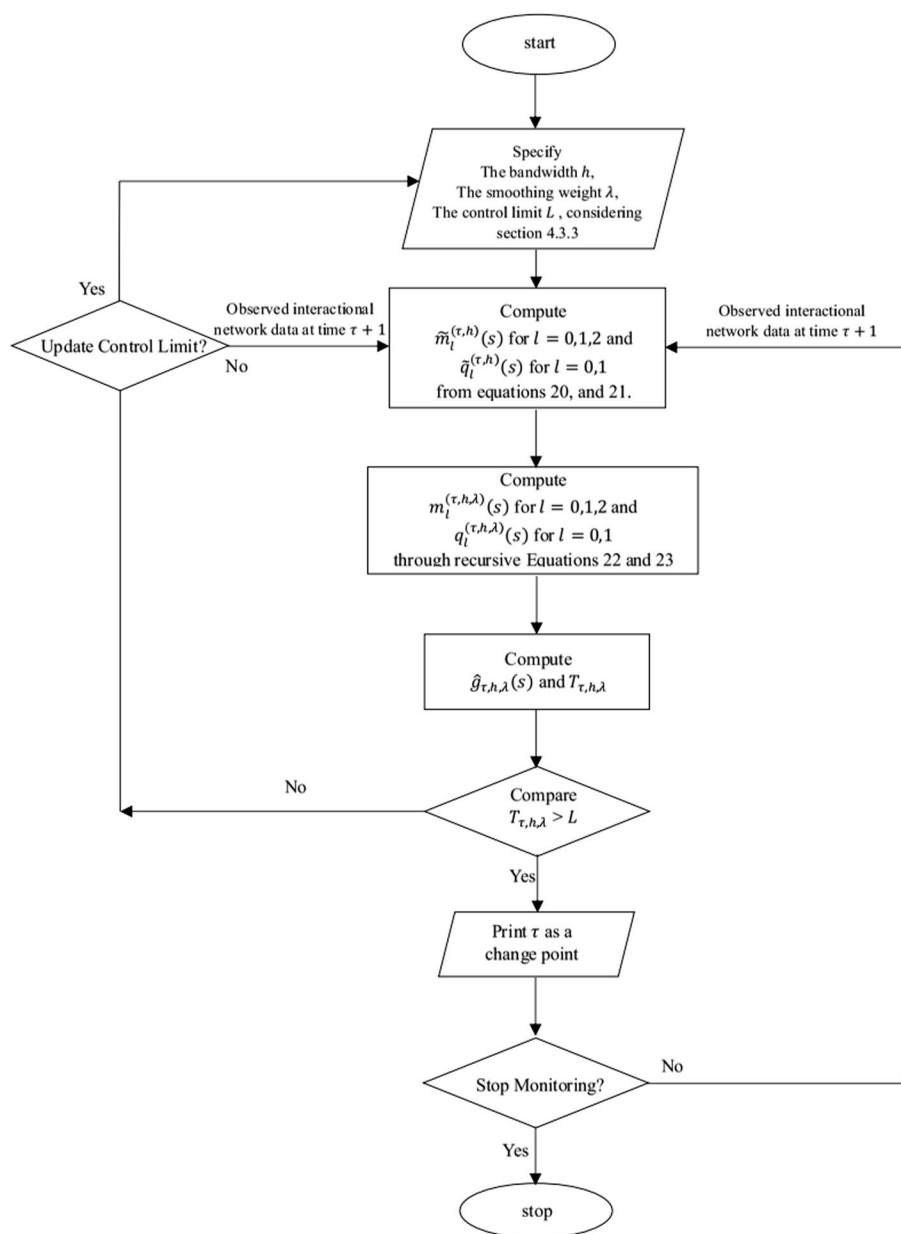


Fig. 2 Workflow of the proposed MENPC

Similar to univariate cases, we followed (Qiu et al. 2010) to specify Kernel, smoothing weight, and bandwidth. Usually, in Phase II SPC, the distribution of in-control data is assumed known, then the control limit can be obtained by simulation. However, in practice, the distribution of in-control data is unknown. Instead, a large number of in-control data can be used by a resampling procedure to obtain the control limit. Here, we use the Block Bootstrap resampling procedures (Sroka 2022) to set the control limit L . In this method, for each run, the in-control dataset is resampled by randomly choosing a sequence of profiles from blocks, until a signal of shift is alerted. Then, an estimated

ARL₀ value is computed based on the simulation runs and L is adjusted based on a comparison to the nominal value (e.g., 500). In this study, we use $B = 10,000$ as the number of simulations runs.

Experiments

This section will investigate the problem of change point detection in real-world social data and synthetic data using the proposed method. We use the number of vertices, the number of edges, the degree of centrality, network density, the number of triangles, betweenness, indegree, outdegree, and total connections as our p dimensional vector of explanatory variables in x_{tij} for constructing the profiles.

Datasets

Two types of datasets are used in this section to evaluate the applicability and performance of the proposed control chart; synthetic and real-world datasets.

Synthetic data

Although there are varying real-world datasets in the literature for experiments, in this subsection, we are experimenting with generated datasets to show the applicability of the proposed method. The decision to use synthetically generated datasets in our study was primarily driven by the following reasons:

Ground Truth Availability: Synthetically generated datasets allow us to have complete control over the ground truth, which is crucial for evaluating the performance of change point detection methods. This control enables us to precisely define the locations of change points, facilitating a thorough assessment of the method's accuracy.

Evaluation of Methodology under various scenarios: Synthetic datasets provide a standardized and reproducible means of evaluating the performance of the MENPC method under various scenarios and conditions. By systematically varying parameters through varying scenarios, we can assess the performance of the method across different settings.

Comparison with existing Methods: Synthetic datasets allow for a fair comparison with existing approaches in the literature. By employing the same set of synthetic datasets, we can evaluate the relative performance of MENPC against alternative methods.

To generate sample networks, here the proposed DCSBM procedure by Wilson et al. (2019) is implemented. This procedure provides flexibility for analyzing shift changes in communication rates among nodes and changes in community structures. The following equation presents the probability function:

$$\mathbb{P}(\cdot | \theta, \pi, P) \sim$$

$$\sum_{c \in \text{all permutation}} \left(\prod_{r \in [k]} \pi_r^{n_r} \prod_{u \in [n]} \theta_u^{d_u} \prod_{u < v \in [n]} \frac{1}{w_{u,v}!} \prod_{r,s \in [k]} P_{r,s}^{\frac{m_{r,s}}{2}} e^{-\frac{n_r n_s P_{r,s}}{2}} \right) \tag{26}$$

where n and k show the number of network nodes and communities, respectively. In addition, $\theta = (\theta_1, \dots, \theta_n)$ is the nonnegative connection tendency for nodes, which is analogous to the node's degree vector representing the intensity of all nodes for having connections, $\pi = (\pi_1, \dots, \pi_k)$ is the nonnegative probability vector representing the probability of community members for a node, in which the summation over all communities equals 1. Also, $P_{k \times k}$ is a nonnegative matrix showing the intensity of connection between communities. Moreover, $c = (c_1, \dots, c_n)$ is a random community label vector assigned to nodes, identically and independently distributed from a multinomial distribution.

$$c \sim \text{Multinomial}(1, \pi) \tag{27}$$

Additionally, n_r is the number of nodes in the community r which follows from the following equation:

$$\sum_{u: c_u=r} \theta_u = n_r \tag{28}$$

And $m_{r,s}$ shows the total weight of edges for communications between community r and s . It can be obtained from Eq. 29.

$$m_{r,s} = \sum_{u,v} w_{u,v} \mathbb{I}(c_u = r, c_v = s) \tag{29}$$

In the simulation, it is common to specify node memberships deterministically. Thus, given c, P , and θ the edge weights should be sampled according to the following Poisson distribution.

$$w_{u,v} \sim \text{Poisson}(\theta_u \theta_v P_{c_u, c_v}) \tag{30}$$

Here, similar to Wilson et al. (2019) the dynamic network is simulated according to the above explanations with $n = 100, k = 2$, and $T = 50$ time points. In addition, the node's connection tendency θ^0 is randomly generated from the following uniform distribution.

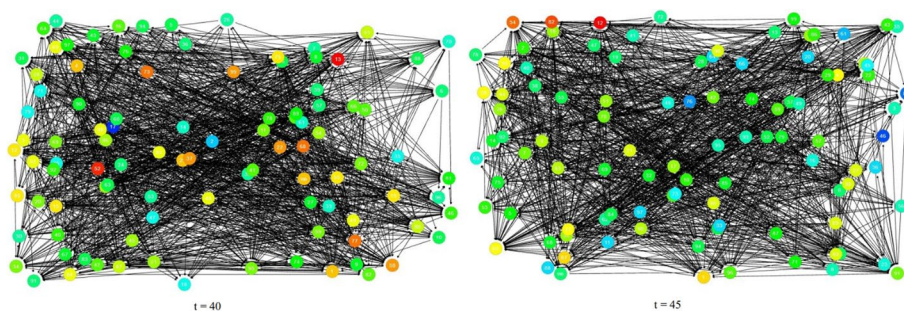


Fig. 3 DCSBM generated dynamic social network with a consistent evolving pattern

$$\theta_u^0 \sim \text{Uniform}(1 - \delta_{c_u}^0, 1 + \delta_{c_u}^0) \tag{31}$$

Furthermore, similar to Wilson et al. (2019), we set $P^0 = \begin{bmatrix} 0.2 & 0.1 \\ 0.1 & 0.2 \end{bmatrix}$, $\delta_{c_u}^0 = 0.5$. Figure 3. shows two snapshots of the generated dynamic social networks with the mentioned initial values for the DCSBM procedure, in time 40 and 45 with no changes over the time.

Enron email dataset

Enron Company, established in 1985, was an American corporation mainly working in the electricity and communication sectors till 2001. According to the Fortune report, the company had reached approximately 101 billion USD in revenue in 2000 and was named the most innovative corporation in the US. Enron’s bankruptcy occurred in 2001 because of accounting fraud by Arthur Andersen LLP firm. After the dissolution of the company, the Enron email dataset was collected in the CALO (A Cognitive Assistant that Learns and Organizes) project. Federal Energy Regulatory Commission made the data publicly available, especially for social network researchers for analysis purposes and developing the research for earlier detection of such frauds in the future. There are different versions of the dataset publicly available on the web. The first one is the March 2, 2004 version, and after that 2009 version and 2011 versions were distributed through the web (William W. Cohen 2015). These versions have problems in data integration and differences in some special messages that are removed from the datasets due to privacy

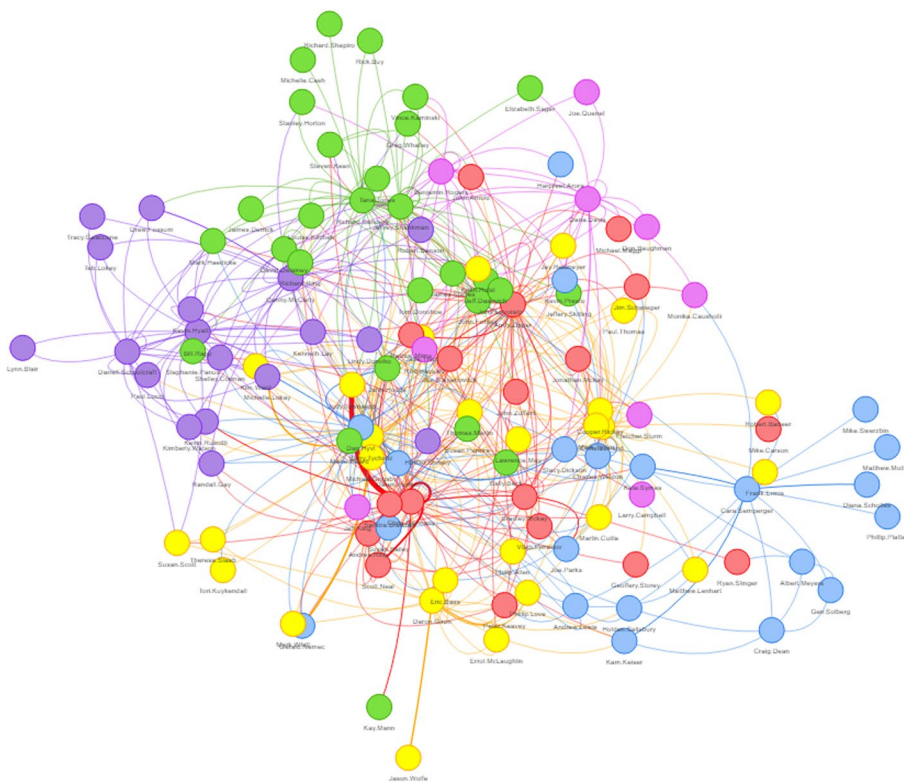


Fig. 4 A snapshot of the evolving Enron email network

concerns. The last version of the dataset we are using in this research is the May 7, 2015 version. It contains the emails between 184 employees of the Enron corporation from 1979 to 2002, including sender and receiver email addresses, sent and received date and time, subject of emails, and email body. Figure 4 shows a snapshot of the interactions in Enron email network, wherein the weight of edges is proportional to the number of interactions.

Performance evaluation

In this section, experimental analysis is carried out using synthetic and real-world datasets to show the superiority of the proposed online social network monitoring tool. We have constructed two proposed control charts in the literature on each dataset and compared them to the MENPC method for online monitoring.

Synthetic dataset

To evaluate the performance of the proposed MENPC on a synthetic dataset, three types of changes through six scenarios in the simulated stream of networks are considered. These are changes in the community structure of the networks, changes in the communications rates, and simultaneous changes in both. We generated the network stream of 100 nodes using a DCSBM procedure, where the DCSBM parameters are similar to Wilson et al. (2019) and drawn from the Poisson distribution, using Eq. (31). Table 3 shows six different scenarios.

Figure 5 illustrates examples of small networks representing six distinct changing scenarios: (a) Global outbreak, (b) Local outbreak in a community, (c) Global variability increase, (d) Local variability increase, (e) Merge communities, and (f) Split a community. Simulations 1–2 relate to changes in the global and local mean interaction rates. We simulate networks based on three different values for $\epsilon = 0.01, 0.05,$ and 0.1 . These scenarios will show the power of the proposed method to identify small and large changes in the interaction rates of the communities and the whole network. Similarly, simulations 3–4 depict scenarios where global and local changes occur in the variance of the interaction rates. In addition, the scenario when θ percent of individuals in the network are involved in the outbreak and they change their communication rates by δ in positive or negative directions is conducted in these simulations. Lastly, simulations 5–6 capture changes in the community structures. In the merging scenario, we set the connection probability value to the average of communities,

Table 3 Synthetic dataset change scenarios

Simulation	Change	Description
1	$P_{ij}^* = P_{ij}^0 + \epsilon$	Global outbreak ($i = 1, 2, j = 1, 2$)
2	$P_{11}^* = P_{11}^0 + \epsilon$	A local outbreak in a community
3	$\delta_i^* = \delta_i^0 + \tau$	Global variability increase
4	$\delta_1^* = \delta_1^0 + \tau$	Local variability increase
5	$C^0 = \{c_1, c_2\} \rightarrow C^* = \{c_1\}$	Merge communities
6	$C^0 = \{c_1, c_2\} \rightarrow C^* = \{c_1, c_2, c_3\}$	Split a community

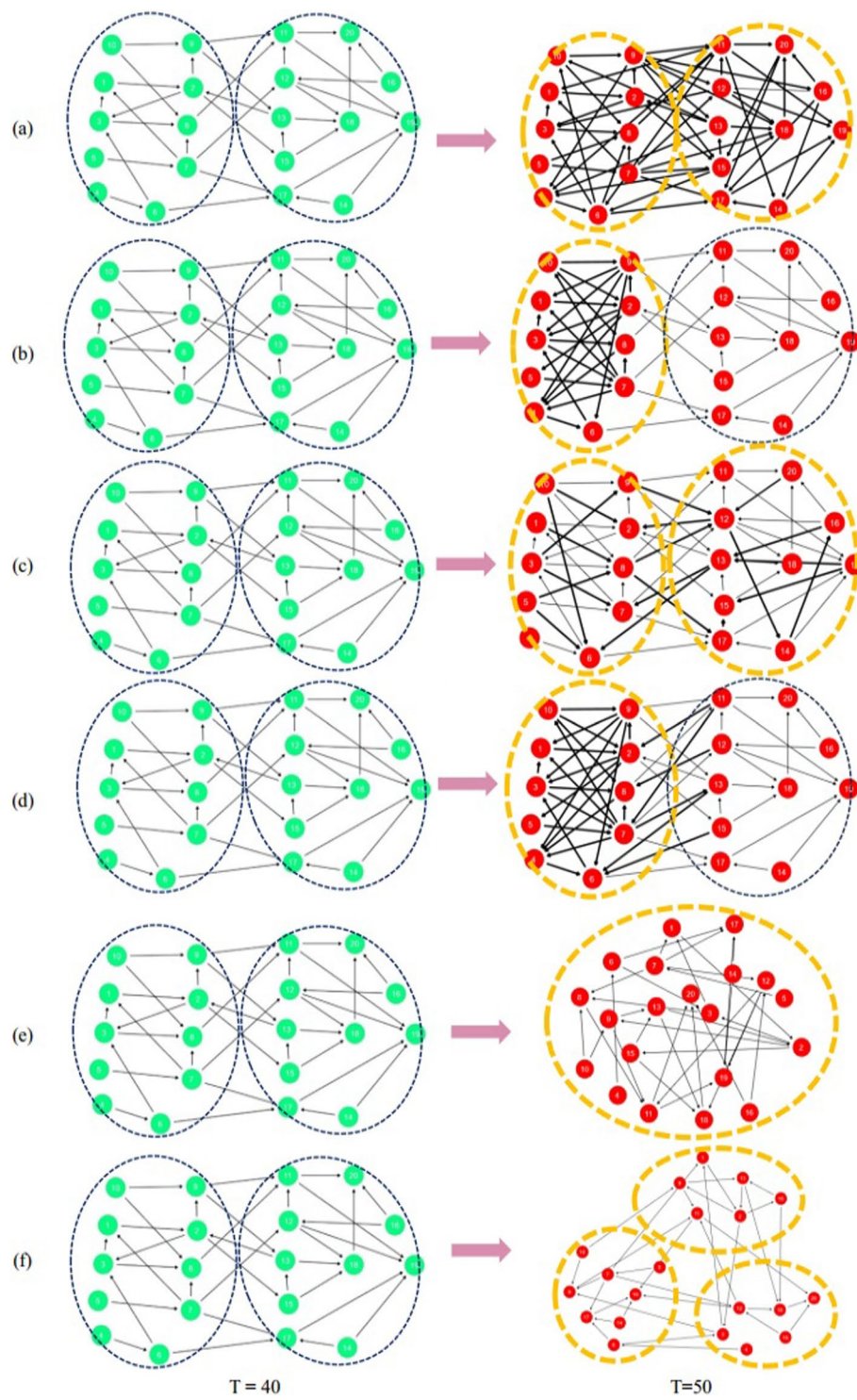


Fig. 5 Sample networks for the changing scenarios

Table 4 Comparison results for the synthetic dataset

Sim	Change	Values	Wilson Method		Eigenvalue Method		MENPC Method	
			ARL	SDRL	ARL	SDRL	ARL	SDRL
0	None	–	501.91	23.39	504.72	11.67	503.77	16.74
1	$P_{ij}^* = P_{ij}^0 + \epsilon$	$\epsilon = 0.01$	81.04	18.18	86.47	8.1	34.92	6.32
		$\epsilon = 0.05$	83.04	26.54	58.27	9.96	28.28	5.92
		$\epsilon = 0.10$	73.72	25.78	54.97	4.87	19.39	7.09
2	$P_{11}^* = P_{11}^0 + \epsilon$	$\epsilon = 0.01$	88.98	19.81	58.72	7.17	44.42	6.27
		$\epsilon = 0.05$	82.61	20.11	50.02	5.41	39.78	5.96
		$\epsilon = 0.10$	80.78	18.07	36.97	4.66	34.45	5.2
3	$P_{12}^* = P_{12}^0 + \epsilon$ $P_{21}^* = P_{21}^0 + \epsilon$	$\epsilon = 0.01$	93.98	23.43	53.02	5.06	42.08	5.64
		$\epsilon = 0.05$	87.06	22.27	45.52	6.11	35.51	5.85
		$\epsilon = 0.10$	74.69	17.37	43.73	4.64	24.59	6.43
4	$\delta_i^* = \delta_i^0 + \tau$	$\tau = 0.01$	106.09	22.48	45.75	4.39	35.46	5.49
		$\tau = 0.05$	102.96	20.96	45.11	5.64	27.71	5.74
		$\tau = 0.10$	95.98	21.16	41.25	5.55	18.92	3.9
5	$\delta_1^* = \delta_1^0 + \tau$	$\tau = 0.01$	98.72	21.89	60.52	6.23	33.55	6.8
		$\tau = 0.05$	95.31	21.31	57.97	5.61	16.26	6.11
		$\tau = 0.10$	78.87	23.8	38.25	4.5	14.76	3.67
6	$\delta_i^* = \delta_i^0 + \tau$ <i>$\theta\%$ of individuals in outbreak</i>	25%	100.61	24.27	53.02	7.62	49.06	6.57
		50%	96.65	22.96	44.25	4.71	35.32	5.67
		75%	89.02	24.44	34.5	4.26	35.21	6.26
7	$C^0 = \{c_1, c_2\} \rightarrow$ $C^* = \{c_1\}$	$n = 50$	49.85	23.88	47.25	6.01	42.12	6.88
		$n = 100$	43.48	23.76	44.18	5.89	38.25	6.69
		$n = 250$	41.21	23.11	39.62	5.94	34.13	6.75
8	$C^0 = \{c_1, c_2\} \rightarrow$ $C^* = \{c_1, c_2, c_3\}$	$n = 50$	72.22	22.68	51.22	7.19	50.17	6.74
		$n = 100$	68.16	23.14	48.27	7.92	47.34	7.01
		$n = 250$	56.28	23.74	47.11	6.89	45.39	6.58

$P^* = 0.15$. In these two scenarios, comparisons occurred on two equally sized communities.

For comparison, we also applied the EWMA monitoring scheme introduced by (2019) Wilson et al. (2019) and Hazrati-Marangaloo and Noorossana (2021) called Wilson’s method and the Eigenvalue method hereafter, respectively. In Eq. 13, x_{tij} is a nine-dimensional vector consists of the topological measures of the networks over time including, the number of vertices, the number of edges, the degree of centrality, network density, the number of triangles, and network betweenness are considered. Also, as the characteristics of each node in network interactions, we considered node indegree, node outdegree, and node total connections in our explanatory variables. Table 4 shows the obtained results over 100 runs for each row, in terms of expected detection delay (EDD) (average run length (ARL₁) in the process control terminology) and standard deviation of run length (SDRL) (Figs. 6, 7, 8, 9).

Enron email dataset

In this experiment, the MENPC algorithm is applied for change point detection in the Enron Email dataset. Although we can improve the precision of results by considering daily or hourly data, the weekly intervals can reveal interesting events that had

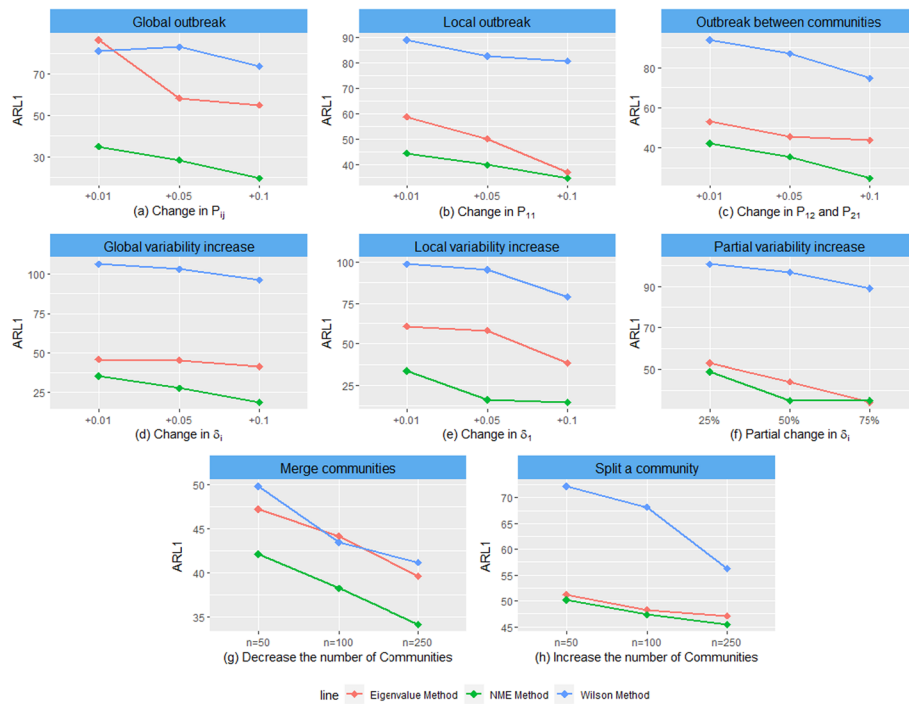


Fig. 6 Analyzing ARL_1 across various scenarios



Fig. 7 Boxplot analysis of change scenarios 1 to 3



Fig. 8 Boxplot analysis of change scenarios 4 to 6

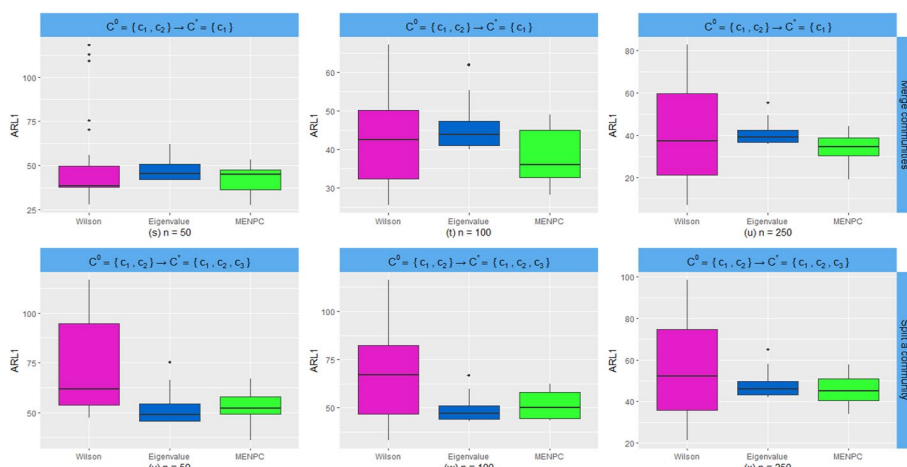


Fig. 9 Boxplot analysis of change scenarios 7 and 8

influences on Enron’s lifetime. Figure 10 shows the obtained statistics in 187 weeks, from 13 November 1998 to 21 Jun 2002. According to the control limit, the points in weeks 96, 100 to 102, 135, and 136 are recognized as change points. Those weeks are related to early September and October 2000, and Jun 2001. Our investigation of these intervals shows that point 96 is around of hitting Enron Stock to an all-time high of \$90.56, while then Enron used "aggressive" accounting to declare \$53 million in earnings on a collapsing deal that hadn’t earned anything at all in profit. In addition, FERC (the Federal Energy Regulatory Commission) orders an investigation into strategies designed to drive electricity prices up in California. After that, change points in

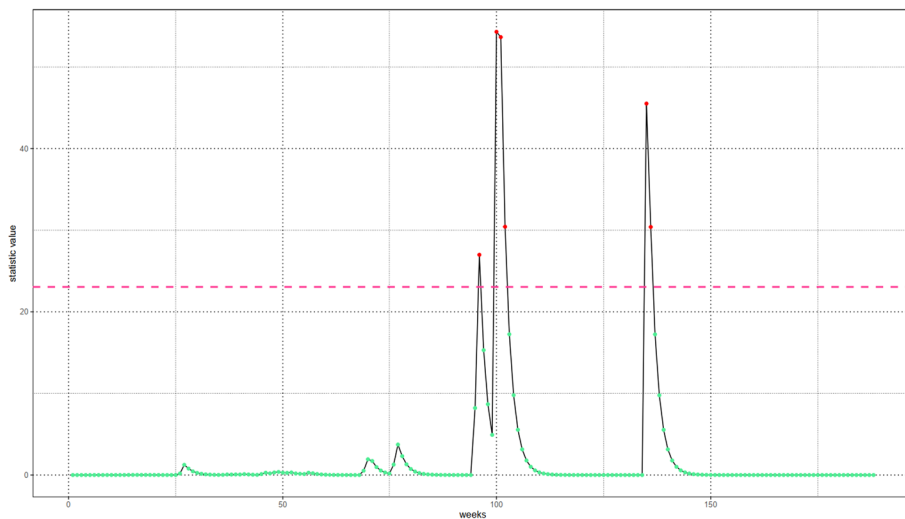


Fig. 10 The proposed chart statistics in 187 weeks on the Enron dataset

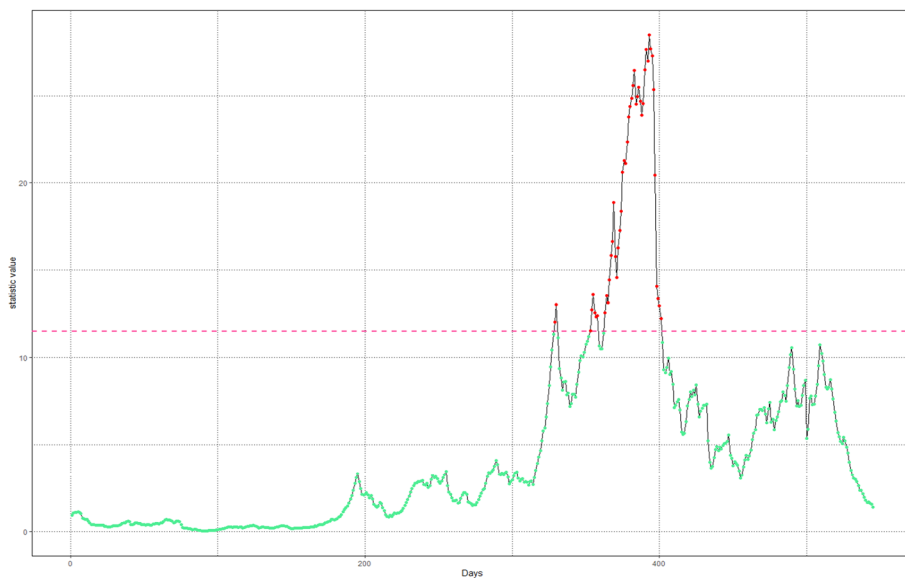


Fig. 11 The proposed control chart for Ask Ubuntu dataset

weeks 100 to 102, are related to the period that Enron attorney Richard Sanders travels to Portland to discuss Timothy Belden’s strategies. Also, the FERC investigation exonerates Enron for any wrongdoing in California. In addition, the change point in weeks 135, and 136 are in Jun 2001, in which FERC finally institutes price caps across the western states, and the California energy crisis ends.

Ask Ubuntu dataset

Ask Ubuntu functions as a collaborative platform within the broader Stack Exchange Network, where members of the Ubuntu community can share knowledge, seek assistance, and provide solutions. It serves as a central hub for discussions, inquiries, and

responses pertaining to Ubuntu operating systems. Utilizing a voting mechanism, the platform evaluates the quality of content, thereby fostering active engagement from users. In this analysis, our focus is on the Ask Ubuntu event datasets spanning from July 2010 to March 2016. Specifically, we narrow our scope to the interactions that transpired within the final 18 months, from October 2014 to March 2016, segmenting the data into daily intervals, which offer insights into evolving trends and notable shifts over time. Each edge in the dataset, denoted as (u, v, t) , signifies an interaction event at time t between user u and user v on Ask Ubuntu. These interactions encompass various forms, including responding to queries, commenting on questions, and engaging in discussions by commenting on responses. To construct the network for analysis, we aggregate all observed interactions within a given day. The accompanying chart displays the obtained control chart, illustrating the dynamics of user interactions over time on Ask Ubuntu.

The obtained control chart for the observed events on the Ask Ubuntu website is shown in Fig. 11. After a thorough analysis of the found change points, it was discovered that they were in line with the release of new version modifications that came in September, and October 2015 of Ubuntu (Miller and Mokryn 2020). The suggested chart, which is based on the MENPC technique, shows how effective it is at detecting changes.

Discussion

The results obtained from running MENPC and competing algorithms on both synthetic and real-world datasets yield three significant insights into nonparametric profile monitoring and social network change point analysis. Firstly, according to the simulation results in Fig. 6, MENPC demonstrates superior performance across various change scenarios, consistently achieving lower ARL_1 values, indicative of higher accuracy. Secondly, MENPC proves to be scalable, effectively handling increasingly larger dynamic network datasets. Even when processing thousands of interactions in a network, it maintains consistent performance, thanks to the innovative trick of updating formulas. This capability makes it well-suited for real-world applications where datasets are often large and complex. Thirdly, as a multivariate approach, MENPC uses both nodal and network topological measures over time without assuming internal network independence across time frames. This versatility highlights MENPC's ability to address change point detection in dynamic social networks effectively. Furthermore, its generality for network data is evident through the development of a multivariate Mixed Effect model, coupled with the capability to utilize both nodal and network-level attributes.

However, it is important to note some limitations of MENPC. Firstly, the performance of the algorithm may be influenced by the choice of hyperparameters h and λ , necessitating further optimization to fine-tune these parameters for specific applications. Additionally, while our algorithm achieved high overall accuracy rates, there was an instance where the Wilson method exhibited a slightly lower ARL_1 due to partial changes in node connection tendencies δ_i^* in scenario 6. This can be attributed to the fact that the Wilson method relies on myopic estimation of maximum likelihood and computing a set of monitoring statistics, leading to faster alerts even when the outbreak involves 75% of individuals. Furthermore, this case highlights MENPC's capability to avoid false alarms

by waiting a bit longer to gather more information and issue accurate alerts. However, it is important to emphasize that the ARL_1 difference between the Wilson method and MENPC was negligible.

In short, MENPC's scalability, accuracy, and versatility make it a valuable tool for analyzing complex network data in various real-world applications. The following subsections provide a more detailed discussion about the computational efficiency and robustness of MENPC.

Computational efficiency

When evaluating the computational efficiency of an algorithm, particularly for change point detection algorithms, it is crucial to consider several aspects. Specifically, factors such as the processing power of the hardware, available memory capacity, the allotted time resources deemed reasonable for real-world applications, and the level of precision accepted in the results play pivotal roles. Indeed, these factors are contingent upon user preferences and utilities, as users must weigh the costs associated with acquiring hardware, the patience required to achieve the desired level of accuracy, and the trade-offs involved in waiting for results. Given the significance of the factors mentioned, the computational efficiency of MENPC can be evaluated from two perspectives: absolute and relative computational efficiency analysis. In absolute analysis, the focus is on investigating the time complexity of MENPC. In relative analysis, however, the algorithm's accuracy and time cost are compared with those of competing methods, illustrating how MENPC performs relative to existing techniques addressing the same problem.

As an extension of the NME model for multivariate cases, the complexity analysis of MENPC demonstrates that the algorithm only requires $O(n_0 p h n_t)$ operations for monitoring the t th profile (Qiu et al. 2010). This complexity is comparable to the computation involved in conventional local linear kernel smoothing. In section "Phase II monitoring using the charting statistic based on the proposed NME model", the process involves computing temporary variables $\tilde{m}_l^{(\tau, h)}(s)$ and $q_l^{(\tau, h, \lambda)}(s)$ for different levels l using Eqs. 20 and 21. This computation is carried out for n_t interactions in the observed network at time t , at n_0 predetermined p -dimensional s points. Subsequently, $m_l^{(\tau, h, \lambda)}(s)$ and $q_l^{(\tau, h, \lambda)}(s)$ are updated through the recursive Eqs. 22 and 23. Finally, the monitoring statistic can be obtained from Eq. 19. Using the proposed recursive updating formulas, it becomes clear that storage does not grow sequentially with time t , which helps in efficiently monitoring a profile at time t .

Furthermore, as a nonparametric profile monitoring technique, MENPC demonstrates superior performance compared to competing algorithms such as the Wilson and Eigenvalue method, as outlined in Table 3. MENPC exhibits lower ARL_1 (expected delay after the occurrence of a change point to detecting changes), indicating its effectiveness in detecting changes more promptly. Additionally, it should be noted that the Wilson method is based on maximum likelihood estimation (Wilson et al. 2019), which requires the use of all historical interactions observed in the dynamic network, as the likelihood function lacks recursiveness. In this regard, the Wilson method utilizes only the observed data at time t , resulting in a myopic estimation that compromises accuracy. Furthermore, while the Wilson method proposes a set of monitoring statistics, selecting appropriate ones poses a challenge for efficient dynamic network monitoring. In short,

the Wilson method sacrifices accuracy to expedite the calculation of monitoring statistics with linear time complexity by relying solely on the observed data at time t rather than historical data.

Furthermore, the eigenvalue method is based on computing the eigenvalues of adjacency matrices of observed social networks over time and subsequently calculating the energy distance between these matrices (Hazrati-Marangaloo and Noorossana 2021). These distances are then monitored using an exponentially weighted moving average (EWMA) weighting scheme. However, it is important to note that computing eigenvalues can be computationally demanding, particularly for large matrices. The computational expense is influenced by factors such as the matrix's size, sparsity, and the specific algorithm employed for eigenvalue computation. For matrices of small to moderate sizes, algorithms typically exhibit a computational complexity of $O(n^3)$, where n represents the matrix's size (Lin et al. 2020). While manageable for smaller matrices, this complexity can become impractical for larger ones. To address this, iterative methods are often employed for large matrices, which are more scalable and memory-efficient, particularly for sparse matrices. Nonetheless, even these iterative methods can pose computational challenges, especially when high precision is necessary or when dealing with exceptionally large matrices.

Overall, MENPC offers favorable time complexity as it extends the NME model, aided by recursive updating formulas that maintain efficiency without a sequential increase in storage over time. In contrast, the Wilson method prioritizes speed over accuracy by relying solely on current data, while the eigenvalue method faces computational hurdles, particularly with large matrices. Thus, MENPC emerges as a promising approach for efficient and accurate change point detection in dynamic social networks.

Robustness and reliability

Figures 7, 8 and 9 depict boxplots illustrating the results of replications in each change scenario for the simulated networks. Concerning the central tendency of the results in each method, the figures consistently indicate that the median values in the boxplots for the MENPC method generally exhibit a lower central location for ARL_1 , implying better performance in expected delay detection. Moreover, when assessing the spread or variability of data within each method, as portrayed by the interquartile range (IQR) in the boxplots, the MENPC method demonstrates a narrower IQR, suggesting a more robust performance. This underscores the MENPC method's suitability in real-world scenarios where variations in network dynamics are prevalent.

Furthermore, due to the λ parameter serving as a weighting parameter in Eq. 9, similar to the weighting scheme in the EWMA method (Noorossana et al. 2011), MENPC inherently exhibits robustness akin to that of EWMA. In essence, it assigns exponentially decreasing weights to past observations, thereby placing greater emphasis on recent data points. This sensitivity enables the method to promptly adapt to changes or trends in the data, rendering it particularly adept at detecting shifts in process behavior. Thus, the derived equation for the population profile $g(\cdot)$ in Eq. 10 aligns with this notion. Therefore, it is reasonable to conclude that the resulting statistic for nonparametric profile monitoring in Eq. 19 showcases robustness to variations in real-world applications.

Additionally, it is worth noting that every change point detection method is prone to false alerts, which need to be addressed by adjusting the Type I error and selecting an appropriate ARL_0 to calculate the control threshold L . In this study, as observed in numerous research works in the literature and real-world applications (Noorossana et al. 2011), the ARL_0 value is set to 500 for conducting the simulation runs, and L is adjusted accordingly based on this nominal value. Choosing a suitable value for ARL_0 and, consequently, deriving an appropriate control limit L ensures the method's resilience against false alarms triggered by fluctuations or noise in the recorded data when no genuine change is present in the real-world scenario. This capability enables the method to accurately differentiate between actual changes in the network data and false alerts stemming from noise in the recorded data.

Practical implications and insights

MENPC is a flexible and effective tool for monitoring online social networks in various settings, including marketing campaigns, corporate operations, and public safety and emergency response plans. By addressing the limitations of existing change point detection methods in online social networks, MENPC enables a wide range of applications to access crucial information necessary for decision-making.

Businesses can utilize MENPC to track user patterns and behaviors in online social networks, identifying new trends, anomalies in user behavior, and optimizing advertising campaigns for maximum impact. Effective monitoring provides vital insights into customer preferences, sentiments, and interaction patterns through accurate change point detection. This information is essential for improving product development efforts, developing customer relationships, and fine-tuning marketing strategies. Additionally, MENPC aids in the early detection of criminal activities and cyber risks by identifying changes in user behaviors or network structures, benefiting law enforcement agencies and governmental entities (Dey et al. 2023).

Moreover, MENPC can play a critical role in managing public health emergencies and tracking the spread of diseases through online social networks. By closely examining changes in user behavior, information distribution patterns, and community dynamics, MENPC provides early warning signs of possible outbreaks or new public health issues. With this information, governmental organizations and healthcare facilities can prevent the spread of diseases and protect public health by taking prompt actions, allocating resources wisely, and developing efficient communication plans (Sterchi et al. 2021).

Conclusion and recommendations for future research

Recently, social network change point detection has attracted much attention among researchers. There are two streams of approaches for analysis; model-based and free-model analysis leading to parametric and nonparametric methods. Although parametric methods are useful in many applications of change point detection, the adequacy of these methods' assumptions and the unknown influence of misspecification on the performance of parametric profile monitoring always is questioning. For example, the assumption of i.i.d errors in social network modeling and the monitoring methods are often invalid in practice, since the social networks have the nature of dynamic and some sort of network autocorrelation during the time. This study proposes a novel approach

for social network monitoring through the design of a multivariate Mixed-Effects Non-parametric Profile Control (MENPC) chart. The proposed charting statistic is based on nonparametric mixed-effects (NME) modeling, local linear kernel smoothing, and the EWMA process weighting scheme. We used local linear kernel smoothing of profile data, through mixed-effects modeling. The experimental results on simulated data and real-world datasets indicate that the proposed method has superiority over the Eigenvalue method and Wilson's method. Industries can use the proposed tool in various applications ranging from obtaining business insights and marketing to criminal monitoring, and monitoring of disease spreads. Future studies can shed more light on the monitoring of online social networks by considering different categories of changes, nodal attributes, and types of connections among nodes.

Abbreviations

IoT	Internet of Things
SNA	Social network analysis
NP	Nonparametric profiles
NME	Nonparametric mixed effects
MENPC	Mixed-effects nonparametric profile control
DCSBM	Degree correlated stochastic block model
ARL	Average run length
SDRL	Standard deviation of run length
EDD	Expected detection delay
CUSUM	Cumulative sum chart
MCUSU	Multivariate cumulative sum chart
EWMA	Exponentially weighted moving average
MEWMA	Multivariate exponentially weighted moving average
LS	Least squares
OLS	Ordinary least squares
WLS	Weighted least square
MLE	Maximum likelihood estimator
BPC	Between-profile correlation
WPA	Within-profile autocorrelation
GLR	Generalized likelihood ratio
GEE	Generalized estimating equations
LLKE	Local linear kernel estimator
SSS	Spectral scan statistic
CCID	Cross-covariate isolate detection
FC	Functional connectivity
FERC	Federal energy regulatory commission

Acknowledgements

Not applicable.

Author contributions

All authors have contributed with ideas for the approach described in the article. AK implemented the coding part for the computational experiments. Both authors read and approved the final manuscript.

Funding

This research was conducted independently without any financial support or sponsorship.

Availability of data and materials

The datasets used in this study are available from the corresponding author upon request. The Enron dataset used here is available at <https://www.cs.cmu.edu/~enron/>. Additionally, the Ask Ubuntu dataset can be accessed at <https://snap.stanford.edu/data/sx-askubuntu.html>.

Declarations

Competing interests

The authors declare that they have no competing interests.

Received: 7 September 2023 Accepted: 25 June 2024

Published online: 03 July 2024

References

- Anastasiou A, Cribben I, Fryzlewicz P (2022) Cross-covariance isolate detect: a new change-point method for estimating dynamic functional connectivity. *Med Image Anal* 75:102252
- Azarnoush B, Paynabar K, Bekki J, Runger G (2016) Monitoring temporal homogeneity in attributed network streams. *J Qual Technol* 48:28–43
- Barigozzi M, Cho H, Fryzlewicz P (2018) Simultaneous multiple change-point and factor analysis for high-dimensional time series. *J Econom* 206:187–225
- Camacho D, Panizo-Lledot A, Bello-Organ G et al (2020) The four dimensions of social network analysis: an overview of research methods, applications, and software tools. *Inf Fusion* 63:88–120
- Cho H, Fryzlewicz P (2015) Multiple-change-point detection for high dimensional time series via sparsified binary segmentation. *J R Stat Soc Ser B Stat Methodol* 77:475–507
- Cohen A, Migliorati G (2017) Optimal weighted least-squares methods. *SMAI J Comput Math* 3:181–203
- William W. Cohen (2015) Enron Email Dataset. <http://www.cs.cmu.edu/~enron/>. Accessed 8 May 2015
- Craven BD, Islam SMN (2011) Ordinary least-squares regression. *SAGE Dict Quant Manag Res* 1:224–228
- Dey A, Kumar BR, Das B, Ghoshal AK (2023) Outlier detection in social networks leveraging community structure. *Inf Sci (NY)* 634:578–586
- Eswaran D, Faloutsos C, Guha S, Mishra N (2018) Spotlight: Detecting anomalies in streaming graphs. In: Proceedings of the 24th ACM SIGKDD international conference on knowledge discovery & data mining, pp 1378–1386
- Farahani EM, Baradaran Kazemzadeh R, Noorossana R, Rahimian G (2017) A statistical approach to social network monitoring. *Commun Stat Methods* 46:11272–11288
- Fotuhi H, Amiri A, Taheriyoun AR (2023) Phase II monitoring of autocorrelated attributed social networks based on generalized estimating equations. *Commun Stat Comput* 52(4):1504–1522
- Galvao AF, Parker T, Xiao Z (2024) Bootstrap inference for panel data quantile regression. *J Bus Econ Stat* 42(2):628–639
- Ghoshal AK, Das N, Das S (2022) A fast community-based approach for discovering anomalies in evolutionary networks. In: 2022 14th international conference on communication systems & networks (COMSNETS). IEEE, pp 455–463
- Hazrati-Marangaloo H, Noorossana R (2021) A nonparametric change detection approach in social networks. *Qual Reliab Eng Int* 37:2916–2935
- He Q, Wang J (2023) Monitoring networks with overlapping communities based on latent mixed-membership stochastic block model. *Expert Syst Appl* 229:120432
- Jeske DR, Stevens NT, Tartakovsky AG, Wilson JD (2018) Statistical methods for network surveillance. *Appl Stoch Model Bus Ind* 34:425–445
- Khalilzadeh M, Karami A, Hajikhani A (2020) The multi-objective supplier selection problem with fuzzy parameters and solving the order allocation problem with coverage. *J Model Manag* 15(3):705–725
- Klyushin D, Martynenko I (2021) Nonparametric test for change point detection in time series. In: CEUR workshop proceedings, pp 117–127
- Kumar C, Bharti TS, Prakash S (2023) A hybrid data-driven framework for spam detection in online social network. *Procedia Comput Sci* 218:124–132
- Lin RM, Mottershead JE, Ng TY (2020) A state-of-the-art review on theory and engineering applications of eigenvalue and eigenvector derivatives. *Mech Syst Signal Process* 138:106536
- Meeker WQ, Escobar LA (1994) Maximum likelihood methods for fitting parametric statistical models. *Methods Exp Phys* 28:211–244
- Miller H, Mokryn O (2020) Size agnostic change point detection framework for evolving networks. *PLoS ONE* 15:e0231035
- Noorossana R, Saghaei A, Amiri A (2011) Statistical analysis of profile monitoring. Wiley, New York
- Peixoto P, Gauvin L (2018) Change points, memory and epidemic spreading in temporal networks. *Sci Rep* 8:1–10
- Qiu P, Zou C, Wang Z (2010) Nonparametric profile monitoring by mixed effects modeling. *Technometrics* 52:265–277
- Rajabi F, Sadinejad S, Saghaei A (2020) Monitoring of social network and change detection by applying statistical process: ERGM. *J Optim Ind Eng* 13:131–143
- Reis MS, Gins G (2017) Industrial process monitoring in the big data/industry 4.0 era: From detection, to diagnosis, to prognosis. *Processes* 5:35
- Salmasnia A, Mohabbati M, Namdar M (2020) Change point detection in social networks using a multivariate exponentially weighted moving average chart. *J Inf Sci* 46:790–809
- Schweinberger M, Stingo FC, Vitale MP (2021) Special issue on statistical analysis of networks. *Stat Methods Appl* 30:1285–1288
- Sharpnack J, Singh A, Rinaldo A (2013) Changepoint detection over graphs with the spectral scan statistic. In: Artificial intelligence and statistics. PMLR, pp 545–553
- Sroka Ł (2022) Applying block bootstrap methods in silver prices forecasting. *Econometrics* 26:15–29
- Sterchi M, Sarasua C, Grütter R, Bernstein A (2021) Outbreak detection for temporal contact data. *Appl Netw Sci* 6:1–21
- Stevens NT, Wilson JD, Driscoll AR et al (2021a) Foundations of network monitoring: definitions and applications. *Qual Eng* 33:719–730
- Stevens NT, Wilson JD, Driscoll AR et al (2021b) Broader impacts of network monitoring: its role in government, industry, technology, and beyond. *Qual Eng* 33:749–757
- Todeschini A, Miscouridou X, Caron F (2020) Exchangeable random measures for sparse and modular graphs with overlapping communities. *J R Stat Soc Ser B (stat Methodol)* 82:487–520
- Wang T, Wang Y, Zang Q (2022) Outlier detection in non-parametric profile monitoring. *Statistics (BER)* 56:805–822
- Wang H, Xie L, Xie Y et al (2023) Sequential change-point detection for mutually exciting point processes. *Technometrics* 65:44–56
- Wasserman S, Pattison P (1996) Logit models and logistic regressions for social networks: I. An introduction to Markov graphs and p. *Psychometrika* 61:401–425
- West BT, Welch KB, Galecki AT (2022) Linear mixed models: a practical guide using statistical software. CRC Press, Boca Raton

- 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* 35:1363–1378
- Wu H, Zhang J-T (2002) Local polynomial mixed-effects models for longitudinal data. *J Am Stat Assoc* 97:883–897
- Yeganeh A, Abbasi SA, Shongwe SC (2022) Monitoring non-parametric profiles using adaptive EWMA control chart. *Sci Rep* 12(1):14336
- Yoon M, Hooi B, Shin K, Faloutsos C (2019) Fast and accurate anomaly detection in dynamic graphs with a two-pronged approach. In: *Proceedings of the 25th ACM SIGKDD international conference on knowledge discovery & data mining*, pp 647–657
- Zara A, Rehman SU, Ahmad F et al (2022) Numerical approximation of modified Kawahara equation using Kernel smoothing method. *Math Comput Simul* 194:169–184
- Zhang J, Ren H, Yao R et al (2015) Phase I analysis of multivariate profiles based on regression adjustment. *Comput Ind Eng* 85:132–144

Publisher's Note

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