Contents lists available at GrowingScience

International Journal of Industrial Engineering Computations

homepage: www.GrowingScience.com/ijiec

# Change point analysis of events in social networks: An online convex optimization approach

**Arya Karami[a,b*] and Seyed Taghi Akhavan Niaki[a]**

[a]*Department of Industrial Engineering, Sharif University of Technology, Tehran, Iran*
[b]*School of Mathematics and Statistics, University of New South Wales, Sydney, Australia*

**C H R O N I C L E**

*Article history*:
Received February 8 2024
Received in Revised Format March 30 2024
Accepted April 4 2024
Available online April 4 2024

Keywords:
*Social network events monitoring*
*Sequential Change Point detection*
*Convex Optimization*
*ADAM algorithm*

**A B S T R A C T**

Nowadays, online social networks play a crucial role in shaping human communication in various life activities. Social Network Analysis (SNA) provides valuable insights for businesses, authorities, and platform owners. One of the challenging tasks in SNA is detecting sequential change points in observed events in social networks when the parameters of statistical distribution of post-change networks are unknown. This challenging problem is particularly prominent in various real-world network systems, especially when the events in the networks can be modeled through a Hawkes process. Identifying change points in the stream of social network data, where the underlying statistical properties undergo significant changes, necessitates the development of adaptive online algorithms. Additionally, in cases where the use of maximum likelihood estimators is impractical or when no exact recursive function for likelihood is available, addressing this issue becomes more complex. This paper proposes likelihood estimators using online convex optimization methods, incorporating the adaptive moment estimation (ADAM) algorithm. The proposed method is seamlessly integrated into the sequential anomaly detection procedure for events in social networks. Experimental results on monitoring time between events demonstrate lower Expected Delay Detection (EDD), indicating the superiority of the proposed algorithm in both synthetic and real-world datasets such as Facebook and contact networks of individuals causing disease transmission. The proposed robust solution provides an efficient practical tool in situations where traditional methods face limitations in swift detection with high accuracy.



## 1. Introduction

Change point detection, or structural break analysis, is a statistical method to identify points in a dataset where a significant shift or change occurs in the underlying data process. These changes can manifest as alterations in the data's mean, variance, distribution, or other statistical properties (Aminikhanghahi & Cook, 2017). Change point detection aims to locate these transition points within a time series or sequence of data points, indicating the presence of shifts in the underlying dynamics. Detecting such changes is crucial in various fields, including finance, healthcare, environmental monitoring, and social networks, where identifying the timing and nature of changes can lead to improved decision-making (Drury et al., 2022; Kovács et al., 2023). Online change point detection is an adaptive extension of traditional change point detection methods, designed to continuously monitor and identify shifts in data patterns as new observations become available over time. Unlike offline methods that analyze the entire dataset simultaneously, online change point detection algorithms dynamically update and assess data in real time. This approach is precious in dynamic environments where changes can occur incrementally (Romano et al., 2023). Online change point detection is vital in applications such as real-time financial market analysis (Zhu et al., 2013), streaming healthcare data (Ali et al., 2021), and dynamic social networks (Song et al., 2016). It enables quick responses, enhancing the agility and responsiveness of decision-making processes in rapidly evolving scenarios. Moreover, as an approach to tackling online monitoring problems, sequential analysis is a fundamental area in statistics that deals with real-time inference from a continuous stream of observations (Lai & Xing, 2010; Karami & Niaki, 2024). While numerous studies have explored the sequential change-point detection problem under the assumption of knowing both pre-change and

* Corresponding author
E-mail karami_ariya@ie.sharif.edu (A. Karimi)
ISSN 1923-2934 (Online) - ISSN 1923-2926 (Print)
2024 Growing Science Ltd.
doi: 10.5267/j.ijiec.2024.4.001

post-change parameters, the current research focuses on a scenario where one possesses knowledge of the pre-change parameters but not the post-change parameters. Here, due to the intractability of maximum likelihood, online estimation of the likelihood becomes a critical issue that influences the overall effectiveness of the change point detection procedure. In this study, accurate knowledge of the parameters before the change point is presumed, which is often a reasonable assumption since obtaining reference data for the normal state enables precise parameter estimation. However, after the change point, the parameters shift to unknown values, indicating irregularities or uncharted circumstances that warrant further investigation. This challenging problem is worth investigating as it has diverse applications, including online anomaly detection, ensuring statistical quality, biosurveillance, identifying opportunities for financial arbitrage, and monitoring network security (Raginsky et al., 2012).

In cases where the post-change parameter values are unknown, two commonly employed approaches are the Cumulative Sum (CUSUM) procedure and the Generalized Likelihood Ratio (GLR) statistic-based procedure (Ruggieri & Antonellis, 2016; Wang et al., 2023). Due to its recursive nature, the CUSUM method is both memory and computation-efficient as it does not require storing historical data. On the other hand, the GLR statistic seeks the maximum likelihood estimate (MLE) of the post-change parameter and incorporates it into the likelihood ratio to create the detection statistic. The GLR statistic is more robust than CUSUM, making it particularly valuable when the post-change parameter may vary between different situations (Cao et al., 2018). MLE or constrained MLE methods lack a recursive structure and cannot be evaluated using simple summary statistics. In such cases, historical data must be stored, and the MLE must be recalculated whenever new data arrives, which is neither memory-efficient nor computationally efficient. Aside from CUSUM and GLR, various online change-point detection procedures employing one-sample updates have been explored (Xie et al., 2023). These procedures replace the MLE with a straightforward recursive estimator. Developing detection procedures for complex scenarios where the exact MLE lacks a recursive or closed-form expression is crucial. In this sense, the one-sample update method offers efficient computation since it can incorporate information from new data with low computational cost. It is also memory-efficient, requiring only the most recent sample. While one-sample update estimators may not correspond to the exact MLE, they often yield robust detection performance (Cao et al., 2018).

The current study aims to introduce a highly effective online sequential change-point detection algorithm for monitoring time between events in social networks by employing an innovative likelihood estimation method. To achieve this, we utilize the adaptive moment estimation (ADAM) approach within a convex optimization to estimate the likelihood (Kingma & Ba, 2014). Our research presents likelihood estimators based on online convex optimization methods, with a central focus on utilizing the ADAM algorithm. This proposed estimation method is seamlessly incorporated into sequential monitoring of the events in social networks. Integrating the ADAM algorithm into the monitoring procedure enhances the adaptability and robustness of the approach, contributing to a more refined and efficient algorithm for detecting change points in dynamic social network data. The main contributions of this paper include 1) applying online convex optimization methods for likelihood estimation, 2) integrating ADAM as a stochastic approximation method for optimization, and 3) proposing a quick and more efficient algorithm for sequential change-point detection of events in social networks, demonstrated by improvements in Average Run Length (ARL) and Expected Delay Detection (EDD).

The subsequent sections of this paper unfold as follows: Section 2 undertakes an in-depth exploration of the pertinent literature in the field of sequential change-point detection. Section 3 provides the foundational knowledge surrounding Hawkes network processes. Section 4 delineates the proposed algorithm tailored for change-point detection in social networks, specifically focusing on Hawkes networks. Section 5 presents an exhaustive examination comprising diverse simulated scenarios within a social network, alongside observed events in prominent platforms such as Facebook, Ask Ubuntu, and the contact network associated with the spread of diseases within two distinct wild populations of Australian sleepy lizards, Tiliqua rugosa, thereby serving as a representative depiction of real-world social networks. Finally, Section 6 explains the findings and presents insightful conclusions and recommendations for future studies. These recommendations are aimed at advancing the understanding and exploration of social network monitoring.

## 2. Related works

Sequential change-point detection is a fundamental challenge within statistical theory and signal processing, holding relevance across diverse domains such as quality control, finance, environmental monitoring, and social networks. Existing approaches addressing the problem of online anomaly detection, concerning the inclusion of information pertaining to both pre-change and post-change parameters, can generally be classified into three main categories: 1) where distribution parameters are known before and after the change points, 2) where pre-change parameters are known but post-change parameters are unknown, and 3) where both pre-change and post-change parameters remain unknown. While a substantial body of literature is dedicated to sequential change-point detection, much progress has been made under the assumption of known distribution parameters before and after the change points (Aminikhanghahi & Cook, 2017). For instance, procedures like the Shiryaev-Roberts (SR) (Shiryaev, 2010) method depend on having prior knowledge of pre and post-change parameters, enabling more effective change-point detection. The Shiryaev (2010) method is designed to swiftly detect changes while minimizing the risks associated with false alarms (declaring a change when none exists) or delays (detecting a change too late). Unlike conventional approaches that await the accumulation of a fixed number of observations before making decisions, the Shiryaev (2010) method enables a sequential examination of incoming data points. This adaptive approach allows for the potential rapid detection of data changes. The technique employs a likelihood ratio test to compare the likelihood of observed data under two competing hypotheses: the null hypothesis (no change has occurred) and the alternative hypothesis (a change has occurred). Continuously updated, the likelihood ratio evolves with the arrival of new data points. A critical aspect of the method is its establishment of a stopping rule grounded in the likelihood ratio. This rule dictates when to conclude the sequential examination and decide. It serves to strike a balance between the imperative of swift change detection and the necessity of avoiding false alarms (Cao et al., 2018). Besides, Shiryaev's (1963) contributions rely upon deriving optimal stopping rules that minimize expected detection delays while adhering to constraints on the probability of false alarms. In essence, it represents a sequential statistical approach aimed at detecting changes in data over time. Its core objective is to achieve a delicate equilibrium between rapid change detection and mitigating false alarms or detection delays.

Additionally, among the researches addressing known pre-change and post-change parameters, Wang et al. (2023) introduced an innovative CUSUM procedure designed for the sequential change-point detection within self- and mutually-exciting point processes, with a focus on Hawkes networks and the utilization of discrete events data. Their work proposes an online recursive implementation of the CUSUM statistic tailored specifically for Hawkes processes. This implementation stands out for its computational efficiency and memory-friendliness, which is particularly suitable for decentralized computing environments. One noteworthy aspect is that the CUSUM procedure typically necessitates the specification of post-change distribution parameters. However, researchers try to improve it and demonstrate its impressive robustness in cases of parameter misspecification, particularly when it becomes feasible to estimate both the topology and magnitude of a potential abrupt change (Kovács et al., 2023). While extensive research has been conducted on sequential change-point detection under the assumption of known pre-change and post-change parameters, considerable attention has been directed towards Bayesian approaches to tackle the challenge of unknown post-change parameters (Ruggieri & Antonellis, 2016). Tartakovsky et al. (2006) developed adaptive sequential and batch-sequential methods for the early detection of attacks causing changes in network traffic. This algorithm applies the change-point detection theory by using thresholding to test statistics, aiming to maintain a fixed rate of false alarms while detecting changes in statistical models as promptly as possible. The proposed approach boasts three attractive features. First, the algorithms are self-learning, enabling them to adapt to varying network loads and usage patterns. Second, they facilitate the detection of attacks with minimal average delay while maintaining a given false alarm rate. Third, their computational simplicity allows for practical online implementation. Furthermore, an introduced sequential Bayesian change-point algorithm provides uncertainty bounds on the number and location of change points. This algorithm updates itself efficiently in linear time as each new data point is recorded and utilizes the exact posterior distribution to infer the presence of a change point (Tartakovsky et al., 2006; Ruggieri & Antonellis, 2016).

In addition, unlike many existing sequential change-point algorithms that only estimate the posterior distribution of change-point locations, the Sequential Bayesian change-point algorithm stands out by directly sampling from the posterior distribution. It makes no assumptions about the distance between adjacent change points, offering a more accurate and precise approach. This algorithm can rapidly update each new observation and provides uncertainty bounds for the number and locations of change points within a dataset (Ruggieri & Antonellis, 2016). Zarepour and Habibi (2023) derived a Quasi-Bayesian change-point test statistic considering fixed and random exchangeable priors. Under null and alternative hypotheses, the asymptotic behavior of these quasi-Bayesian test statistics is expressed in terms of stochastic integrals. It also discusses M-estimate approaches for detecting changes in mean, covering scenarios with both finite and infinite variance observations. The random prior can be updated sequentially for future change points at each point during the change detection process. The quasi-Bayesian test statistic is introduced to assess the null hypothesis of no change point. It is formulated as a weighted combination of score functions, assigning higher probability to potential change point locations. The choice of weight (prior) function holds significant importance. Various approaches such as empirical arguments, Bayesian bootstrapping, and updating rules are considered for selecting weights from the available sample, with random weights, particularly exchangeable weights (such as Dirichlet priors), being favored over deterministic weights. The test statistic is represented as stochastic integrals involving empirical processes, and its limiting distributions are studied using point process techniques. Various random weight assumptions are examined under null and alternative hypotheses (Zarepour & Habibi, 2023). Additionally, Beyond the Bayesian approaches, there is a body of research addressing change-point detection in scenarios where both pre-change and post-change parameters are unknown by using convex optimization methods to iteratively update unknown post-change parameters and obtain an estimation of likelihood and proceed with the change-point detection procedure with a reasonable and acceptable imposed error. Mei (2006) presented asymptotically optimal procedures tailored for one-parameter exponential families. Furthermore, a comprehensive theory is developed in this research to tackle change-point problems when both the pre-change and post-change distributions involve unknown parameters. In the work described by Lai and Xing (2010), the focus is on achieving asymptotic optimality from both Bayesian and frequentist perspectives within sequential change-point detection in multiparameter exponential families when the parameters before and after the change are unknown. It proposes an extension of Shiryaev's model to encompass situations where both the pre-and post-change parameters remain unknown. This is accomplished by introducing conjugate priors for the unknown parameters within the framework of exponential families. In this scenario, while the Bayesian model provides explicit formulas for posterior distributions, the associated optimal stopping problem, governed by the Bayes rule, becomes notably more challenging than Shiryaev's original problem. In contrast, the researchers introduce the concept of a sliding window approach, which furnishes an asymptotically optimal solution to the Bayesian problem (Lai & Xing, 2010).

The primary focus of this paper centers on a specific scenario in which the pre-change parameters are known. In contrast, the post-change parameters remain unknown, particularly emphasizing the Shiryaev approach. Within this context, three primary approaches exist for addressing such cases: the Generalized Likelihood Ratio (GLR) method, the mixture approach, and parameter estimation. GLR approach represents a method employed for online anomaly detection within sequential data (Mariani & Cawley, 2021). It extends the traditional likelihood ratio test to scenarios where the underlying distributional properties of the data might change, rendering it highly adaptable and applicable across various domains. The GLR approach maintains a continuous update of a likelihood ratio statistic as new data arrives, facilitating early detection of changes in the data patterns. Unlike recursive methods, the GLR approach does not adhere to a repetitive sequence of steps. Instead, it operates as a sequential analysis technique, evaluating data as it becomes available. To address situations where handling large volumes of data becomes computationally burdensome, researchers frequently employ a modified version of GLR known as the window-limited GLR (He et al., 2018). This variant considers only a limited recent history of data, reducing computational demands while offering meaningful change detection capabilities. However, a notable drawback of GLR-based methods lies in the need to solve optimization problems to determine when to declare a change. These optimization problems often lack closed-form solutions, necessitating computationally intensive iterative numerical techniques for resolution (Cao et al., 2018). The mixture approach in change-point detection involves the fusion of multiple change-point detection methods to bolster the overall reliability and adaptability of the detection system. Siegmund & Yakir's (2008) work centers on the minimax optimality of the Shiryaev-Roberts (Shiryaev, 2010) change-point detection rule, with the primary objective of minimizing detection delay while controlling the risk of false alarms. In contrast, Pollak's research (Pollak, 1987) delves into the performance characteristics of an alternative change-point detection method. Incorporating a mixture approach would entail the integration of these methods. Specifically, the Shiryaev-Roberts (Shiryaev, 2010) rule would be employed when prior information about the data distribution is available. At the same time, Pollak (1987) method would come into play when such information is lacking. This blending strategy enables the creation of a more versatile change-point detection system, capable of adapting to diverse scenarios and striking a balance between detection speed, accuracy, and robustness. The specifics of combining these methods would depend on the application context and research objectives. Utilizing a mixture approach, researchers aim to develop a more potent and flexible tool for detecting changes in data sequences, ensuring that the system performs effectively under various conditions and data distributions. However, one of the challenges inherent in the mixture approach lies in the selection of weights for functions when no conjugate prior is available for the distributions of unknown post-change parameters. This aspect adds complexity to the approach and underscores the need for careful consideration in its implementation.

Another viable approach involves substituting the unknown parameters with non-anticipating estimates, a methodology in which the estimates exclusively depend on the current historical information (Cao et al., 2018). These estimates do not incorporate or anticipate future data when making parameter estimations. This approach ensures that the estimations are solely grounded in the information known up to the current time, with no assumptions regarding foreknowledge of future observations. A prior distribution is developed for unknown parameters to implement this approach, and a recursive nature is introduced by estimating non-anticipating parameter values.

In this study, we employ stochastic approximation techniques for convex optimization when estimating these non-anticipating estimates. Significant progress has been achieved in parameter estimation within the online setting. These advancements encompass a wide range of topics, such as online density estimation within the exponential family, achieved through regret minimization (Raginsky et al., 2009; Raginsky et al., 2012; Azoury & Warmuth, 2001). Additionally, there has been substantial research in sequential prediction for individual sequences employing the logarithmic loss function (Cesa-Bianchi & Lugosi, 2006; Kotłowski & Grünwald, 2011), as well as online prediction methods tailored for time series data. The design of sequential change-point detection procedures revolves around optimizing a critical trade-off between two key performance metrics. One metric quantifies the detection delay, while the other assesses the frequency of false alarms. Two standard mathematical formulations exist for addressing this optimal trade-off problem. The first formulation is the minimax approach, initially proposed by Lorden (1971) and later explored by Pollak (1985). This framework's primary objective is to minimize the worst-case delay while adhering to a prescribed lower bound on the mean time between false alarms. The second formulation is Bayesian, introduced by Shiryaev (1963). Within this framework, the change point is assumed to follow a prior distribution, and the primary goal is to minimize the expected delay while maintaining an upper bound on the false-alarm probability. As a result, both Bayesian and non-Bayesian sequential detection methods are assessed based on two fundamental performance criteria: the rate of false alarms and the detection delay (Tartakovsky et al., 2006).

In the proposed methodology in (Raginsky et al., 2012), a specialized loss function is utilized for online estimation. The work introduces and analyzes a sequential (or online) anomaly detection method in individual sequences of potentially noisy observations. The proposed approach is particularly relevant when the anomaly detection system can receive external feedback to validate or dispute its inference regarding the current observation's anomalous nature relative to its historical context. This approach's foundation rests on exponential family models that can be applied across various contexts. Importantly, it achieves sublinear per-round regret concerning static and slowly changing product distributions with marginals from the same exponential family. Furthermore, the regret concerning stationary distributions aligns with the minimax value of the corresponding online strongly convex game (Raginsky et al., 2012).

## 3. Preliminaries

User interactions like comments, likes, and resharing are ubiquitous in social networks, offering valuable insights into their underlying mechanisms. Analyzing engagement patterns unveils user preferences, while reshare data reveals viral content dynamics. Influencer impact becomes evident through likes and shares, emphasizing their role in content amplification. Community-specific trends are discernible, highlighting the intricate interplay of user behavior. In essence, scrutinizing these events provides a nuanced understanding of social network dynamics, crucial for comprehending the intricate tapestry of user engagement and content dissemination.

Assume a sequence of times $X_1, X_2, \dots, X_T$ of observed events in a social network with a probability density function following a parametric form $f_\theta$, where the parameter $\theta$ may be unknown, the main question arises is the sequential change point detection of observed events. The provided Fig. 1 illustrates the chronological progression of events within a social network. While the parameters before the change $\theta_0$ are well-understood, the parameters of the distribution after the change $\theta_1$ remain unknown.

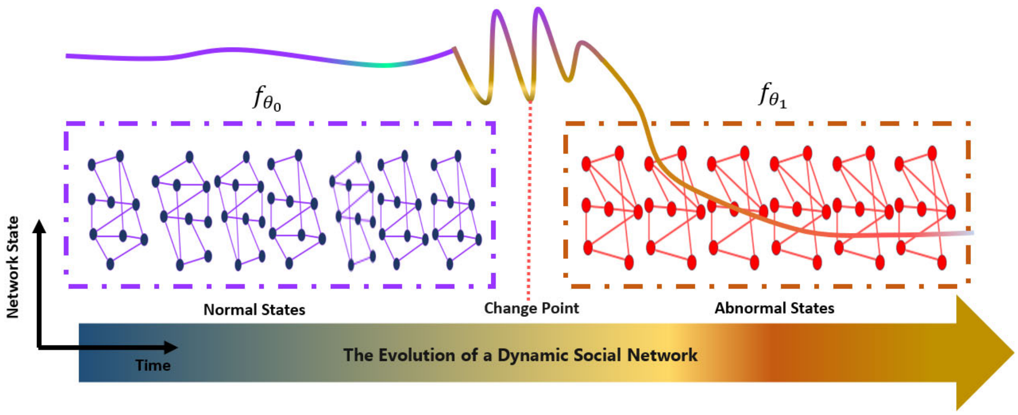


**Fig. 1.** The evolution of events in a social network

In this research, the detection statistic relies on the sequence of estimates $\{\hat{\theta}_t\}$ constructed using online stochastic approximation approaches, such as the adaptive moment estimation (ADAM) algorithm. In the proposed algorithm, the update from $\hat{\theta}_{t-1}$ to $\hat{\theta}_t$ only utilizes the current observation $X_t$. This marks a significant departure from the traditional generalized likelihood ratio (GLR) statistic, where each $\hat{\theta}_t$ is estimated using all historical data. In the following subsections, we first provide a detailed explanation of the problem of sequential change point detection, followed by the proposed Adaptive Cumulative Likelihood Ratio (ACLR) procedure. Then, to apply ACLR for monitoring events in social networks, we begin with a review of Hawkes Point processes and their log likelihood functions, which serves as a foundation for subsequent discussions. Next, we present the ACLR procedure for network event monitoring. Given that the procedure involves convex optimization problems, we illustrate these challenges and discuss stochastic approximation approaches such as the Adaptive Moment (ADAM) method. Subsequently, we incorporate our estimator based on one-sample estimation and propose a sequential change point detection procedure for observed events in a Hawkes process within a social network.

### *3.1 Sequential change point detection*

Sequential change points pertain to detecting moments within a sequential dataset where a substantial change or shift emerges in the fundamental statistical properties. This technique holds immense importance in fields such as finance, as it enables the continuous monitoring of data streams and the timely identification of transitions or anomalies. In this scenario, a change might occur at an unknown time, denoted as $\nu$, which alters the underlying data distribution. The primary objective is to detect such changes promptly. Formally, sequential change point detection can be framed as the following hypothesis test, where the effort is to find changing time points as quickly as possible:

$$\begin{cases} H_0: X_1, X_2, \dots, X_\nu \sim f_{\theta_0} \\ H_1: X_1, X_2, \dots, X_\nu \sim f_{\theta_0}, \quad X_{\nu+1}, X_{\nu+2}, \dots \sim f_\theta \end{cases} \tag{1}$$

Where $\theta$ presents the unknown model parameters after the anomaly. The goal is to detect the changes as quickly as possible after they occur under a false alarm constraint. We will propose likelihood ratio-based detection procedures adapted from the CUSUM method. In the process of detecting change points, the estimation of the parameter after the change occurs relies on the samples collected after the suspected change point. Specifically, when examining putative change points occurring prior to the current time $k < t$, the samples considered for estimating the post-change parameter consist of the sequence

$\{X_k, X_{k+1}, \ldots, X_t\}$. This approach ensures that the parameter estimation is based on the data gathered after the potential change, thus facilitating accurate detection of shifts in the underlying distribution.

$$\hat{\theta}_{k,i} = \hat{\theta}_{k,i}(X_k, X_{k+1}, \ldots, X_i), \qquad i \geq k \tag{2}$$

Therefore, for $k = 1$, $\hat{\theta}_{k,i}$ becomes $\hat{\theta}_i$ serving as a statistic for offline monitoring which uses all observed events until the last one. Initializing with $\hat{\theta}_{k,i} = \theta_0$, the likelihood ratio at time $t$ for a hypothetical change point location $k$ is given by:

$$\Lambda_{k,t} = \prod_{i=1}^{t} \frac{f_{\hat{\theta}_{k,i-1}}(X_i)}{f_{\theta_0}(X_i)}, \qquad i \geq k \tag{3}$$

where $\Lambda_{k,t}$ can be computed recursively, as follows:

$$\Lambda_{k,t} = \Lambda_{k,t-1} \frac{f_{\hat{\theta}_{k,i-1}}(X_i)}{f_{\theta_0}(X_i)}, \qquad i \geq k \tag{4}$$

Although we do not know the change-point location $v$ in equation 1, from the maximum likelihood principle, we take the maximum of the statistics over all possible values of $k$. This will give the Adaptive Cumulative Likelihood Ratio (ACLR) procedure as follows:

$$\begin{aligned} T_{ALR}(CL_1) &= \inf\left\{t \geq 1 : \max_{1 \leq k \leq t} \log \Lambda_{k,t} > CL_1\right\} \\ &= \inf\left\{t \geq 1 : \min_{1 \leq k \leq t} - \log \Lambda_{k,t} < -CL_1\right\} \end{aligned} \tag{5}$$

Where $CL_1$ is a prespecified threshold. In practice, to prevent memory and computational complexity, we can use a window-limited version of the detection procedure, which can be obtained as follows:

$$T_{ALR,\omega}(CL_2) = \inf\{t \geq 1 : \max_{t-\omega \leq k \leq t} \log \Lambda_{k,t} > CL_2\} \tag{6}$$

*3.2 Hawkes Point process*

A Hawkes process is a type of point process used in statistics and applied mathematics to model the occurrence of events or points in time that exhibit self-excitatory behavior. In other words, events in a Hawkes process tend to trigger further events. Hawkes processes have various applications in finance, epidemiology, and social network analysis. They are used to model a wide range of phenomena where events or interactions between entities occur in a self-reinforcing manner. Mathematically, a Hawkes process can be described using intensity functions, which represent the instantaneous rate of event occurrence at any given point in time. The intensity function in a Hawkes process is often expressed as a sum of a baseline rate and contributions from past events, where the contributions decay exponentially over time. One approach to characterize this process is by examining the sequence of random arrival times, specifically when the counting process $N(t)$ experiences jump. Fig. 2 depicts both the point process and its corresponding counting process. Furthermore, it illustrates how the conditional intensity function of the process increases subsequent to the observation of an event.

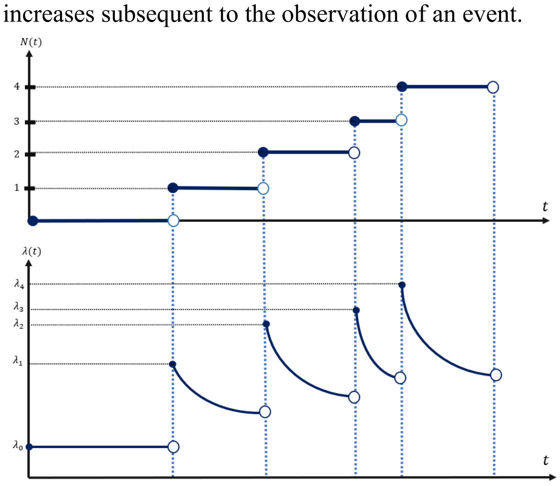


**Fig. 2.** An example of a Hawkes process, along with its corresponding counting process and conditional intensity function

In a one-dimensional homogeneous Hawkes process, the intensity function is given by Equation (12) (Li et al., 2018):

$$\lambda(t|s_t) = \lambda_0 + \sum_{t_i \in s_t} \exp\left(-(t - t_i)\right). \tag{7}$$

This Equation reflects how the occurrence of past events enhances the likelihood of future occurrences. Using this intensity function, one can define the survival function denoted as $S(t|s_t) = \exp\left(-\int_{t_n}^{t} \lambda(\tau)d\tau\right)$, which represents the conditional probability of no events occurring in the interval $[t_n, t)$. The likelihood of observing an event at time $t$ can be defined as $f(t|s_t) = \lambda(t|s_t)S(t|s_t)$. Consequently, the joint likelihood of observing a sequence of events $s_T = \{t_1, t_2, t_3, \dots, t_n | t_n < t\}$ can be formulated within an observation window $T$ as follows:

$$p(\{t_1, t_2, \dots, t_n | t_n < T\}) = \prod_{t_i \in s_T} \lambda\left(t_i \middle| s_{t_i}\right). \exp\left(-\int_0^T \lambda(\tau|s_\tau)d\tau\right) \tag{8}$$

Performing integral normalization within the likelihood function can pose computational challenges, especially when the expression for $\lambda(t|s_t)$ is complex. In such situations, resorting to numerical approximations becomes a common practice. In addition, in the context of an inhomogeneous Hawkes process, the intensity function is represented as follows (Wang et al., 2023):

$$\lambda(t) = \mu(t) + \alpha \int_0^t \varphi(t - \tau) dN_\tau, \tag{9}$$

In this Equation, $\mu(t)$ represents the base intensity, $\alpha$ is the influence parameter, and $\varphi(t)$ is a kernel function satisfying the condition $\int \varphi(t) = 1$. The exponential kernel, $\varphi(t) = \beta e^{-\beta t}$, with $\beta > 0$, is often employed. Additionally, we assume $0 \leq \alpha < 1$ to maintain a stationary process. The likelihood function and the corresponding log-likelihood function are provided as

$$p(\{t_1, t_2, \dots, t_n | t_n < T\}) = \prod_{i=1}^{n} \lambda(t_i) . \exp\left(-\int_0^t \mu(s) ds - \alpha \sum_{i=1}^{n} \int_{t_i}^{t} \varphi(s - t_i)\, ds\right) \tag{10}$$

$$l_t = \sum_{i=1}^{n} \log\left(\mu(t_i) + \alpha \sum_{t_j < t_i} \varphi(t_i - t_j)\right) - \int_0^t \mu(s) ds - \alpha \sum_{i=1}^{n} \int_{t_i}^{t} \varphi(s - t_i)\, ds \tag{11}$$

Furthermore, the log-likelihood is pivotal in the change point detection procedure when employing an exponential kernel alongside a constant base intensity, as follows in equation 12.

$$l_t = \sum_{i=1}^{n} \log\left(\mu + \alpha \sum_{t_j < t_i} \beta e^{-\beta(t_i - t_j)}\right) - \mu t - \alpha \sum_{i=1}^{n} \left[1 - e^{-\beta(t - t_i)}\right] \tag{12}$$

*3.3 Hawkes Process change point detection*

Change point detection in Hawkes processes involves identifying points in time where the underlying process poses a shift in behavior. These changes could indicate shifts in the intensity function, representing the rate of event occurrences over time. Detecting change points in Hawkes Processes is crucial in various applications, such as online social network platforms where the individual interactions can be modeled via the Hawkes stochastic process. Assume the intensity of the Poisson process is $\lambda_s, s \in (0, T)$ and there may exist a change point $\kappa \in (0, T)$ such that the process changes. The null and alternative hypothesis tests adapted from equation 1 are as follows:

$$\begin{cases} H_0: \lambda_s = \mu, \quad 0 < s < T \\ H_1: \lambda_s = \mu, \quad 0 < s < \kappa \\ \qquad \lambda_s = \mu + \theta \displaystyle\sum_{\kappa < t_j < s} \varphi(s - t_j), \quad \kappa < s < T \end{cases} \tag{13}$$

where $\mu$ is a known baseline intensity, $\theta > 0$ is the unknown magnitude of the change, and $\varphi(s) = \beta e^{-\beta s}$ is the normalized kernel function with prespecified parameter $\beta > 0$, which captures the influence from past events. We treat the post-change influence parameter $\theta$ as unknown since it represents an anomaly. Here, we use a sliding window to convert the event times into a sequence of vectors with overlapping events. Assuming a sliding window with size $L$, for a given scanning time $T_i \leq T$, we map all the events in $[T_i - L, T_i]$ to a vector $X_i = [t_{(1)}, \dots, t_{(m_i)}]^T$, $t_{(i)} \in [T_i - L, T_i]$, where $m_i$ is the number of events falling into the window. Considering a set of scanning times $T_1, T_2, \dots, T_t$, this maps the event times into a sequence of vectors

$X_1, X_2, \dots, X_t$ of lengths $m_1, m_2, \dots, m_t$. These scanning times can be arbitrary; here, we set them to event times so that at least one sample per sliding window exists.

Given that the intensity function adheres to the Hawkes Point process, we can employ the proposed monitoring statistic in equation 4 and the corresponding window-limited version of the detection procedure outlined in equation 5. However, to ensure an efficient optimization process for minimization of the expression $\min_{1\leq k\leq t} - \log \Lambda_{k,t}$, we enhance the procedure by incorporating the ADAM method as a stochastic approach for addressing the underlying convex optimization problem.

*3.4 Online convex optimization: Stochastic approaches*

Online convex optimization (OCO) algorithms can be conceptualized as a player who sequentially makes decisions. At each decision point, the player is unaware of the outcomes. The player incurs a loss once a decision is made, which can be chosen adversarially. An OCO algorithm's objective is to make decisions that, based on observed outcomes, minimize regret. Regret is defined as the difference between the total incurred loss and the best-fixed decision in hindsight (Wang et al., 2023; Karimi & Sadjadi, 2024). We focus on OCO algorithms with likelihood-based regret functions to design non-anticipating estimators. The approach involves iteratively estimating parameters as new observations become available using the maximum likelihood principle. Consequently, the incurred loss corresponds to the negative log-likelihood of the new sample, evaluated using the estimator $l_t(\theta) \coloneqq -\log\, f_\theta(X_t)$, which aligns with the log loss function.

Given the observations $\{X_1, X_2, \dots, X_t\}$, the regret for a sequence of estimators $\{\hat{\theta}_i\}_{i=1}^{t}$, generated by a likelihood-based OCO algorithm is as follows:

$$Loss_t = \sum_{i=1}^{t}\{-logf_{\hat{\theta}_{i-1}}(X_i)\} - \inf_{\tilde{\theta}\in\Theta}\sum_{i=1}^{t}\{-logf_{\tilde{\theta}}(X_i)\} \tag{14}$$

At each time step, the estimator $\hat{\theta}_{t-1}$ is updated using the new sample $X_t$, by balancing the tendency to stay close to the previous estimate against the tendency to move toward the greatest local decrease of the loss function. For the loss function defined above, a sequence of ADAM estimators would be constructed to estimate $\hat{\theta}_t$ at each step. For strongly convex loss function, the regret of OCO algorithms has the property that $loss\ \leq Clog(n)$ for some constant $C$ (which depends on $f_\theta$), and any positive integer $n$.

*3.5 Adaptive Moment (ADAM) method*

The Adaptive Moment Estimation (ADAM) algorithm is a versatile tool designed to optimize stochastic objective functions by using first-order gradients and adaptive estimates of lower-order moments. This method is well known in the learning process of neural networks because of its high efficiency, requiring minimal memory resources, and has proven to be remarkably effective in addressing large-scale, high-dimensional problems. Widely recognized for its robustness and applicability across a wide range of non-convex optimization challenges within the field of machine learning, ADAM stands out for its ability to compute individual adaptive learning rates for different parameters based on estimates of gradient moments (Kingma & Ba, 2014).

In practical terms, ADAM aims to minimize the expected value of a noisy objective function, denoted as $f(\theta)$, with respect to the parameters $\theta$. This function is stochastic and may result from the evaluation of random subsamples (minibatches) of data points or intrinsic characteristics of the function itself. At each timestep from 1 to $T$, denoted as $f_1(\theta), \dots, f_T(\theta)$, the gradient, represented as $g_t = \nabla_\theta f_t(\theta)$, captures the partial derivatives of $f_t$ with respect to $\theta$. ADAM updates exponential moving averages of two quantities: the gradient, denoted as $m_t$, and the squared gradient, $v_t$. These averages' decay rates are controlled by hyperparameters $\beta_1$ and $\beta_2$, both belonging to the interval $[0,1)$. They serve as estimations for the first moment (mean) and the second raw moment (uncentered variance) of the gradient. However, initializing these moving averages as vectors of 0's leads to biased moment estimates towards zero, especially during initial timesteps and when the decay rates are small (i.e., $\beta s$ close to 1). Fortunately, this bias can be mitigated through effective initialization, resulting in bias-corrected estimates denoted as $\hat{m}_t$ and $\hat{v}_t$. The updating formula for $\hat{m}_t$ and $\hat{v}_t$ presented in equations 15 to 18, as outlined by (Heusel et al., 2017), plays a crucial role in this correction process.

$$m_t = \beta_1 . m_{t-1} + (1-\beta_1) . g_t \tag{15}$$

$$\hat{m}_t = \frac{m_t}{1-\beta_1^t} \tag{16}$$

$$v_t = \beta_2 . v_{t-1} + (1-\beta_2) . g_t^2 \tag{17}$$

$$\hat{v}_t = \frac{v_t}{1-\beta_2^t} \tag{18}$$

Good default settings for the tested machine learning problems typically include $\alpha = 0.001, \beta_1 = 0.9, \beta_2 = 0.999$, and $\epsilon = 10^\wedge(-8)$. All vector operations are conducted element-wise. Here, $\beta_1^t$ and $\beta_2^t$ denote the values of $\beta_1$and $\beta_2$ raised to the power $t$, respectively. Finally, the updating formula for parameters is as follows:

$$\theta_t = \theta_{t-1} - \alpha . \frac{\widehat{m}_t}{\sqrt{\hat{v}_t + \epsilon}} \tag{19}$$

In each iteration, the objective is to predict the parameter $\theta_t$ and assess it on a previously unknown cost function. This involves computing the vector of partial derivatives with respect to the stochastic objective function. Subsequently, based on these gradients, the bias-corrected first moment estimate and bias-corrected second raw moment estimate are calculated. Iterations persist until the convergence criteria are satisfied (Kingma & Ba, 2014). Furthermore, Chen et al. (2019) offer a theoretical analysis of Adam's convergence in online convex programming.

## 4. Proposed approach

For a hypothetical change point location $k$, it can be shown that the log-likelihood ratio, between the Hawkes process and the Poisson process, as a function of $\theta$, is given by

$$l(\theta|X_i) = \sum_{t_q \in (T_i - L, T_i)} \log \left[\mu + \theta \sum_{t_j \in (T_i - L, t_q)} \beta e^{-\beta(t_q - t_j)}\right] - \mu L - \theta \sum_{t_q \in (T_i - L, T_i)} \left[1 - e^{-\beta(T_i - t_q)}\right] \tag{20}$$

According to the sliding window approach, we can approximate the original change point detection problem as follows. With the assumption that $X_1, X_2, \dots, X_t$ are sampled from a Poisson process, under the alternative circumstances, the change occurs at some time such that $X_1, X_2, \dots, X_\kappa$ are sampled from a Poisson process, and $X_{\kappa+1}, X_{\kappa+2}, \dots, X_t$ are sampled from a Hawkes process with parameter $\theta$, for assumed change point location $\kappa = k$ as follows.

$$\hat{\theta}_{k,i} = \hat{\theta}_{k,i}(X_\kappa, X_{\kappa+1}, \dots, X_i) = \hat{\theta}_{k,i}\ (t_l \in [T_k, T_i]) \tag{21}$$

Now, consider $k \in [i-w, i-1]$, and keep $w$ estimators: $\hat{\theta}_{i-w,i}, \dots, \hat{\theta}_{i-1,i}$. The update for each estimator is based on the adaptive moment estimation (ADAM) stochastic method. By taking the derivative with respect to $\theta$, we have:

$$g(.) = \frac{\partial l(\theta|X_i)}{\partial \theta} = \sum_{t_q \in (T_i - L, T_i)} \frac{\sum_{t_j \in (T_i - L, t_q)} \beta e^{-\beta(t_q - t_j)}}{\mu + \theta \sum_{t_j \in (T_i - L, t_q)} \beta e^{-\beta(t_q - t_j)}} - \sum_{t_q \in (T_i - L, T_i)} \left[1 - e^{-\beta(T_i - t_q)}\right] \tag{22}$$

Note that there is no close form expression for the MLE, which is the solution to the above Equation. We perform ADAM algorithm estimate to update each of $w$ estimators ($\hat{\theta}_{i-w,i}, \dots, \hat{\theta}_{i-1,i}$).

$$\hat{\theta}_{k+1,i+1} = \hat{\theta}_{k,i} - \alpha \frac{\widehat{m}_{k,i}}{\sqrt{\hat{v}_{k,i}} + \varepsilon} \qquad k = i-w, i-w+1, \dots, i-1 \tag{23}$$

$$m_{k+1,i+1} = \beta_1 m_{k,i} + (1-\beta_1) g_{k,i} \qquad k = i-w, i-w+1, \dots, i-1 \tag{24}$$

$$v_{k+1,i+1} = \beta_2 v_{k,i} + (1-\beta_2) g_{k,i}^2 \qquad k = i-w, i-w+1, \dots, i-1 \tag{25}$$

$$\widehat{m}_{k,i} = \frac{m_{k,i}}{1-\beta_1^i} \qquad k = i-w, i-w+1, \dots, i-1 \tag{26}$$

$$\hat{v}_{k,i} = \frac{v_{k,i}}{1-\beta_2^i} \qquad k = i-w, i-w+1, \dots, i-1 \tag{27}$$

Because Eq. (20) presents the $\log \left(\frac{f_{\hat{\theta}_{k,t}}(X_i)}{f_{\theta_0}(X_i)}\right)$, which equals the log-likelihood ratio, Eq. (28) is used to calculate the monitoring statistics $\log(\Lambda_{k,t})$.

$$\log(\Lambda_{k,t}) = \log(\Lambda_{k,t-1}) + \log\left(\frac{f_{\widehat{\theta}_{k,i-1}}(X_i)}{f_{\theta_0}(X_i)}\right)\ i \geq k \tag{28}$$

Finally, we use the following Adaptive Cumulative Likelihood Ratio (ACLR), where the threshold $CL$ is adjusted to reach a prespecified false alarm rate.

$$\max_{t-\omega\leq k\leq t}(\log \Lambda_{k,t}) > CL \tag{29}$$

## 5. Experiments

In this section, some experiments are carried out using synthetic and real-world datasets to assess the effectiveness of the proposed method for detecting change points in the stream of observed events in social network data.

*5.1 Simulated events in social network data*

Dynamic networks are prevalent across various domains, particularly relevant to social network interactions. These dynamic networks are typically represented by observed event data in the format of $(t, i, j)$, indicating that an event occurred between node $i$ and $j$ at time $t$. The interactions may be directed or undirected, including friendship relation, commenting on a post, exchanging messages, likes and resharing posts. A pivotal aspect inherent in social network events and encapsulated within this data type is reciprocity. In the context of this research, a meticulous examination is conducted on the temporal sequence of observed events. Particularly in certain change detection algorithms, the time interval between events plays a significant role in detecting changes. In this research, the generated network event data intervals are set to $t = 750$.

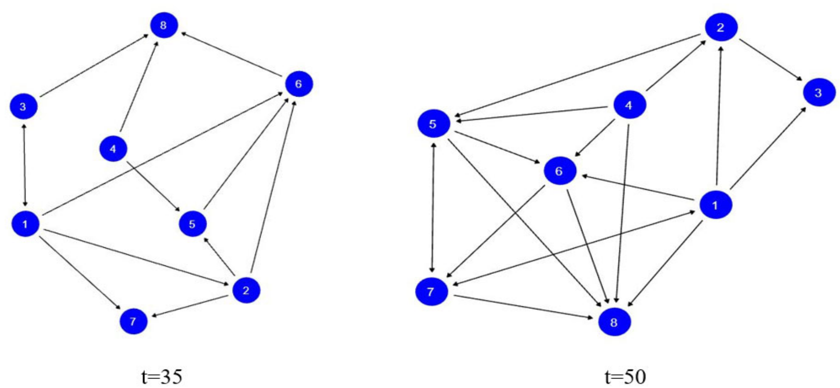


**Fig. 3.** Small networks simulated at two sequential time points

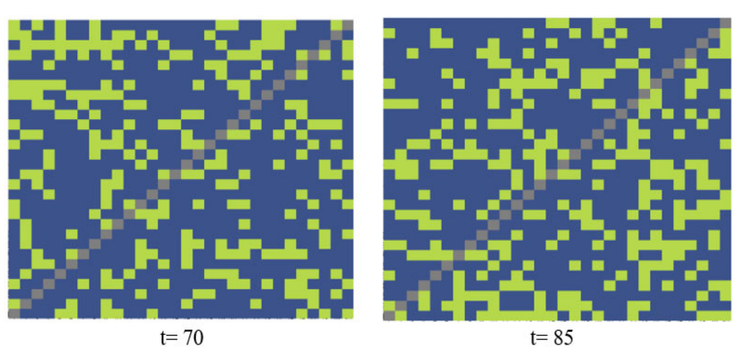


**Fig. 4.** Simulated medium size networks with 30 nodes at two consecutive time points

Fig. 3 and Fig. 4 show simulated networks at two distinct subsequent time points. The events in the social networks are simulated based on the Hawkes process with the parameters $\mu \sim Gamma(0.1,0.1)$ and $\alpha \sim Gamma(0.1,0.1)$. Table 1 to 3 show the performance of the Adaptive Cumulative Likelihood Ratio (ACLR) method and the commonly employed CUSUM, CAPA (Fisch et al., 2022), and EWMA methods for detecting change points of observed events within social networks. In each scenario, the Average Run Length (ARL) under a no-change situation is maintained at 500 by determining the control limit threshold through Monte Carlo simulation. The proposed approach uses the Block Bootstrap resampling procedure (Sroka, 2022) to establish the control limit CL. This method involves resampling the in-control dataset for each run by randomly selecting sequences of profiles from blocks until a shift signal is detected. Subsequently, we calculate an estimated Average Run Length without a shift based on the simulation runs. The control limit CL is then adjusted through a comparison with the nominal value (e.g., 500). For this study, we set the number of simulations at 10,000. Each scenario is simulated 100 times, and the results include the Expected Delay Detection (EDD) or the Average Run Length (ARL1) under the change scenario, as well as the standard deviation of run lengths (SDRL).

Three distinct set of scenarios are investigated, categorized by network size, including small (8 nodes), medium (30 nodes), and large networks (100 nodes). Additionally in each category, $\delta_\mu = \{0.2,0.4,0.6,0.8\}$ and $\delta_\alpha = \{0.2,0.4,0.6,0.8\}$ shift values for background rate and influence parameters applied. Fig. 5 illustrates the performance of the proposed ACLR method compared to the competing CUSUM, CAPA and EWMA methods. As depicted in the figure, the ACLR method emerges as a dependable tool for detecting changes in events of dynamic social networks. These results underscore the efficacy of the ACLR method and its potential superiority over traditional approaches in the context of dynamic social network monitoring.

**Table 1**
Comparison of the methods in terms of $ARL_1$ and SDRL in small networks

| sim | $\delta_\mu^*$ | $\delta_\alpha^*$ | CAPA | | CUSUM | | EWMA | | ACLR | |
|---|---|---|---|---|---|---|---|---|---|---|
| | | | $ARL_1$ | SDRL | $ARL_1$ | SDRL | $ARL_1$ | SDRL | $ARL_1$ | SDRL |
| 1 | $1.0\delta_\mu$ | $1.0\delta_\alpha$ | 500.79 | 48.55 | 500.40 | 35.20 | 500.61 | 50.87 | **500.57** | **38.97** |
| 2 | $1.2\delta_\mu$ | $1.2\delta_\alpha$ | 57.85 | 12.28 | 23.34 | 7.15 | 30.81 | 12.65 | **22.07** | **4.25** |
| 3 | | $1.4\delta_\alpha$ | 47.84 | 10.49 | 21.41 | 7.07 | 28.68 | 17.53 | **17.34** | **4.17** |
| 4 | | $1.6\delta_\alpha$ | 45.02 | 10.23 | 18.43 | 6.65 | 28.25 | 12.79 | **16.54** | **4.40** |
| 5 | | $1.8\delta_\alpha$ | 41.11 | 11.24 | 14.12 | 6.64 | 26.44 | 12.47 | **11.95** | **3.19** |
| 6 | $1.4\delta_\mu$ | $1.2\delta_\alpha$ | 54.52 | 10.54 | 39.02 | 6.90 | 25.34 | 15.28 | **16.98** | **4.40** |
| 7 | | $1.4\delta_\alpha$ | 30.21 | 6.22 | 23.78 | 6.95 | 24.65 | 9.21 | **14.17** | **4.52** |
| 8 | | $1.6\delta_\alpha$ | 27.57 | 4.81 | 22.20 | 6.55 | 24.49 | 15.30 | **13.12** | **4.20** |
| 9 | | $1.8\delta_\alpha$ | 26.82 | 4.95 | 21.36 | 7.63 | 22.68 | 5.46 | **10.47** | **4.23** |
| 10 | $1.6\delta_\mu$ | $1.2\delta_\alpha$ | 66.50 | 13.66 | 46.28 | 7.11 | 22.37 | 7.39 | **16.13** | **3.98** |
| 11 | | $1.4\delta_\alpha$ | 41.45 | 6.57 | 31.12 | 6.97 | 21.77 | 10.36 | **12.93** | **4.39** |
| 12 | | $1.6\delta_\alpha$ | 38.94 | 6.31 | 29.32 | 8.41 | 21.35 | 9.33 | **10.88** | **4.18** |
| 13 | | $1.8\delta_\alpha$ | 35.85 | 6.88 | 26.52 | 7.66 | 19.93 | 8.33 | **10.04** | **5.07** |
| 14 | $1.8\delta_\mu$ | $1.2\delta_\alpha$ | 50.23 | 10.72 | 42.86 | 9.62 | 16.12 | 15.00 | **9.77** | **5.09** |
| 15 | | $1.4\delta_\alpha$ | 49.37 | 9.85 | 41.60 | 9.40 | 9.53 | 7.43 | **9.40** | **3.18** |
| 16 | | $1.6\delta_\alpha$ | 40.30 | 9.03 | 40.40 | 8.48 | 8.54 | 5.72 | **8.26** | **4.81** |
| 17 | | $1.8\delta_\alpha$ | 33.97 | 8.36 | 31.74 | 8.86 | 7.73 | 4.89 | **6.97** | **5.08** |

**Table 2**
Comparison of the methods in terms of $ARL_1$ and SDRL in medium size networks

| sim | $\delta_\mu^*$ | $\delta_\alpha^*$ | CAPA | | CUSUM | | EWMA | | ACLR | |
|---|---|---|---|---|---|---|---|---|---|---|
| | | | $ARL_1$ | SDRL | $ARL_1$ | SDRL | $ARL_1$ | SDRL | $ARL_1$ | SDRL |
| 1 | $1.0\delta_\mu$ | $1.0\delta_\alpha$ | 500.42 | 54.38 | 499.85 | 37.46 | 500.26 | 45.30 | **500.59** | **32.18** |
| 2 | $1.2\delta_\mu$ | $1.2\delta_\alpha$ | 81.23 | 19.97 | 44.54 | 8.81 | 36.88 | 11.29 | **20.28** | **4.98** |
| 3 | | $1.4\delta_\alpha$ | 79.62 | 21.25 | 43.54 | 8.67 | 35.06 | 34.50 | **19.51** | **5.62** |
| 4 | | $1.6\delta_\alpha$ | 79.46 | 17.61 | 42.95 | 8.52 | 32.70 | 30.01 | **18.97** | **5.04** |
| 5 | | $1.8\delta_\alpha$ | 66.94 | 16.39 | 36.50 | 8.72 | 28.64 | 26.29 | **18.58** | **5.14** |
| 6 | $1.4\delta_\mu$ | $1.2\delta_\alpha$ | 78.38 | 18.85 | 45.12 | 6.79 | 26.11 | 27.27 | **17.84** | **5.06** |
| 7 | | $1.4\delta_\alpha$ | 72.22 | 17.42 | 42.56 | 6.62 | 24.61 | 23.35 | **17.60** | **5.06** |
| 8 | | $1.6\delta_\alpha$ | 68.16 | 17.05 | 40.92 | 6.62 | 22.84 | 20.26 | **17.53** | **5.57** |
| 9 | | $1.8\delta_\alpha$ | 61.10 | 14.50 | 38.60 | 6.48 | 20.58 | 21.90 | **17.34** | **5.59** |
| 10 | $1.6\delta_\mu$ | $1.2\delta_\alpha$ | 82.03 | 23.76 | 48.04 | 5.64 | 28.97 | 28.15 | **17.18** | **5.64** |
| 11 | | $1.4\delta_\alpha$ | 72.34 | 19.26 | 44.77 | 5.67 | 27.12 | 29.35 | **15.02** | **5.02** |
| 12 | | $1.6\delta_\alpha$ | 71.00 | 21.53 | 43.45 | 5.53 | 25.56 | 23.08 | **14.85** | **4.91** |
| 13 | | $1.8\delta_\alpha$ | 69.81 | 18.56 | 42.84 | 6.36 | 22.96 | 24.20 | **14.31** | **5.53** |
| 14 | $1.8\delta_\mu$ | $1.2\delta_\alpha$ | 79.19 | 23.02 | 57.78 | 6.87 | 14.88 | 14.58 | **12.95** | **5.59** |
| 15 | | $1.4\delta_\alpha$ | 76.94 | 24.83 | 56.08 | 6.83 | 13.03 | 13.38 | **12.48** | **6.04** |
| 16 | | $1.6\delta_\alpha$ | 73.40 | 20.75 | 54.43 | 6.66 | 12.43 | 12.05 | **11.23** | **5.57** |
| 17 | | $1.8\delta_\alpha$ | 65.82 | 18.27 | 49.28 | 6.58 | 10.64 | 11.90 | **10.38** | **4.82** |

Fig. 6 to Fig. 8 display boxplots representing results from 100 replications in each change scenario for small, medium, and large networks. Regarding the central tendency of results in each method, the figures consistently reveal that the median values in the boxplots for the ACLR method mostly show a lower central location for ARL1, indicating better performance in expected delay detection. Furthermore, in assessing the spread or variability of data within each method, as depicted by the interquartile range (IQR) in the boxplots, the ACLR method exhibits a narrower IQR, suggesting a more consistent and stable performance.

**Table 3**
Comparison of the methods in terms of $ARL_1$ and SDRL in large networks

| sim | $\delta_\mu^*$ | $\delta_\alpha^*$ | CAPA | | CUSUM | | EWMA | | ACLR | |
|---|---|---|---|---|---|---|---|---|---|---|
| | | | $ARL_1$ | SDRL | $ARL_1$ | SDRL | $ARL_1$ | SDRL | $ARL_1$ | SDRL |
| 1 | $1.0\delta_\mu$ | $1.0\delta_\alpha$ | 499.88 | 59.14 | 499.91 | 43.46 | 500.85 | 95.37 | **500.38** | **47.16** |
| 2 | | $1.2\delta_\alpha$ | 119.37 | 56.04 | 117.76 | 6.61 | 99.42 | 20.41 | **80.62** | **8.10** |
| 3 | $1.2\delta_\mu$ | $1.4\delta_\alpha$ | 110.61 | 58.21 | 114.36 | 5.67 | 98.33 | 20.45 | **79.29** | **8.07** |
| 4 | | $1.6\delta_\alpha$ | 109.92 | 55.15 | 112.47 | 5.50 | 96.96 | 20.32 | **76.60** | **8.11** |
| 5 | | $1.8\delta_\alpha$ | 98.63 | 55.68 | 112.23 | 5.75 | 92.94 | 18.69 | **76.38** | **8.03** |
| 6 | | $1.2\delta_\alpha$ | 117.01 | 51.58 | 122.93 | 5.63 | 85.27 | 20.06 | **76.33** | **8.08** |
| 7 | $1.4\delta_\mu$ | $1.4\delta_\alpha$ | 113.65 | 54.04 | 122.38 | 5.21 | 83.26 | 22.27 | **76.24** | **8.00** |
| 8 | | $1.6\delta_\alpha$ | 111.47 | 55.94 | 118.42 | 6.52 | 82.93 | 21.27 | **73.78** | **8.06** |
| 9 | | $1.8\delta_\alpha$ | 110.02 | 54.87 | 117.85 | 6.67 | 75.60 | 20.51 | **72.70** | **8.00** |
| 10 | | $1.2\delta_\alpha$ | 128.16 | 58.48 | 124.40 | 6.89 | 73.91 | 22.67 | **72.02** | **8.06** |
| 11 | $1.6\delta_\mu$ | $1.4\delta_\alpha$ | 115.29 | 58.72 | 122.66 | 5.57 | 72.14 | 21.38 | **69.42** | **8.01** |
| 12 | | $1.6\delta_\alpha$ | 109.87 | 56.55 | 119.04 | 6.73 | 70.04 | 18.53 | **68.36** | **8.03** |
| 13 | | $1.8\delta_\alpha$ | 102.97 | 55.54 | 111.52 | 6.61 | 69.03 | 21.64 | **66.32** | **8.15** |
| 14 | | $1.2\delta_\alpha$ | 108.38 | 58.51 | 105.96 | 4.52 | 67.45 | 18.87 | **61.82** | **7.98** |
| 15 | $1.8\delta_\mu$ | $1.4\delta_\alpha$ | 107.29 | 54.35 | 105.24 | 7.47 | 57.26 | 22.35 | **54.30** | **8.12** |
| 16 | | $1.6\delta_\alpha$ | 99.30 | 52.51 | 103.78 | 7.46 | 56.75 | 21.11 | **53.74** | **7.96** |
| 17 | | $1.8\delta_\alpha$ | 94.52 | 57.96 | 102.67 | 7.66 | 55.67 | 23.25 | **52.68** | **8.18** |

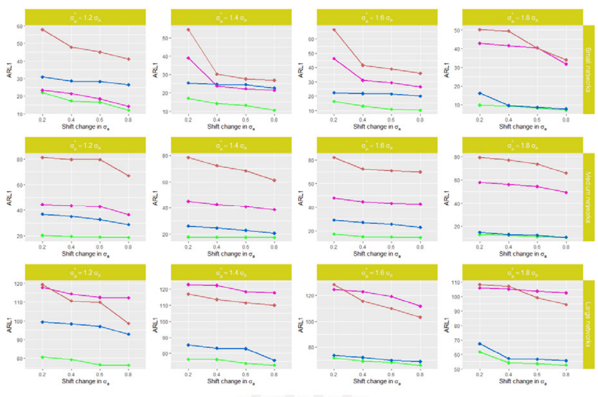

**Fig. 5.** EDD benchmark for small, medium, and large networks

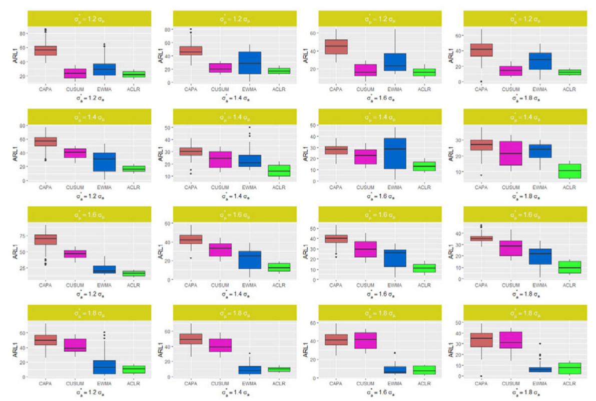

**Fig. 6.** Boxplot analysis of small network results across change scenarios

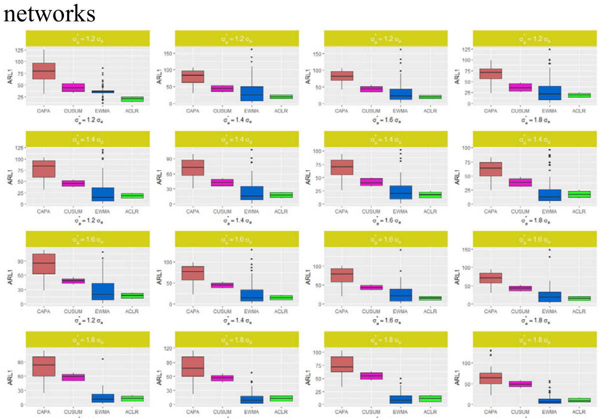

**Fig. 7.** Boxplot analysis of medium network results across change scenarios

**Fig. 8.** Boxplot analysis of large network results across change scenarios

Additionally, the presence of a non-trivial number of outliers in the boxplots provides evidence of the robustness of the ACLR method compared to CUSUM, CAPA and EWMA, indicating its resilience to extreme observations. The ACLR method offers transparent and interpretable results and demonstrates practical utility in sequential change point detection in dynamic social networks.

*5.2 Real-world data*

*5.2.1 Epidemic data*

This experiment employs ACLR method to analyze the dynamics of Salmonella enterica spread within two wild populations of Australian sleepy lizards, Tiliqua rugosa. Although network models are crucial for comprehending infectious disease transmission, face challenges due to limited knowledge of transmission rates and mechanisms. The proposed ACLR method serves as a network monitoring tool, aiding in monitoring transmission mechanisms, and detecting change points in events. The dataset captures a dynamic network of proximity interactions over 70 days between 43 lizards at site 1 and 44 lizards at site 2. Edges signify instances where pairs of individuals were in proximity, specifically within a 14-meter distance. Furthermore, the edge weights correspond proportionally to the frequency of physical interactions between the respective node pairs. Fig. 9 is a graphical representation of four sample networks, each encapsulating daily interactions.

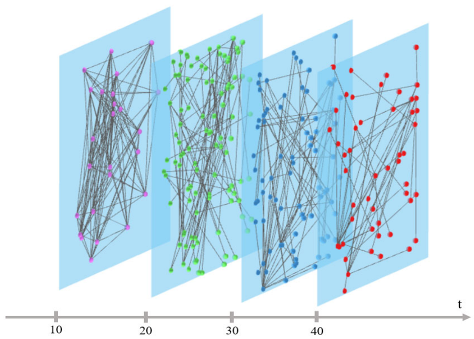


**Fig. 9.** Identifying transmission mechanisms of disease spread from contact networks

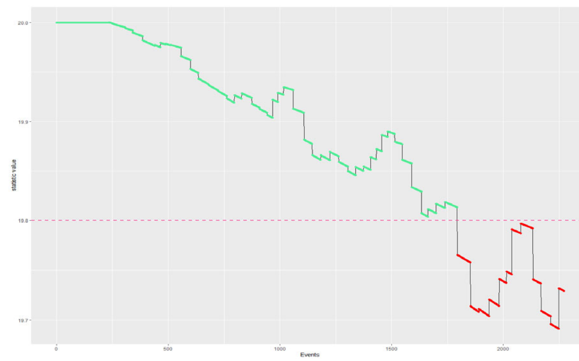

**Fig. 10.** Control Chart for monitoring contact networks in site 2

To construct the control chart, data from site 1 was utilized to determine the control limit, and subsequently, it was tested against data from site 2. Notably, the control chart for site 2, as depicted in Fig. 10, reveals distinctive change points, particularly in event number 1794, which occurred on day 58 of the observed interactions. This subtle alteration provides a crucial insight into the complexities of Salmonella transmission within lizard populations, particularly in the subsequent weeks following the recognized change point.

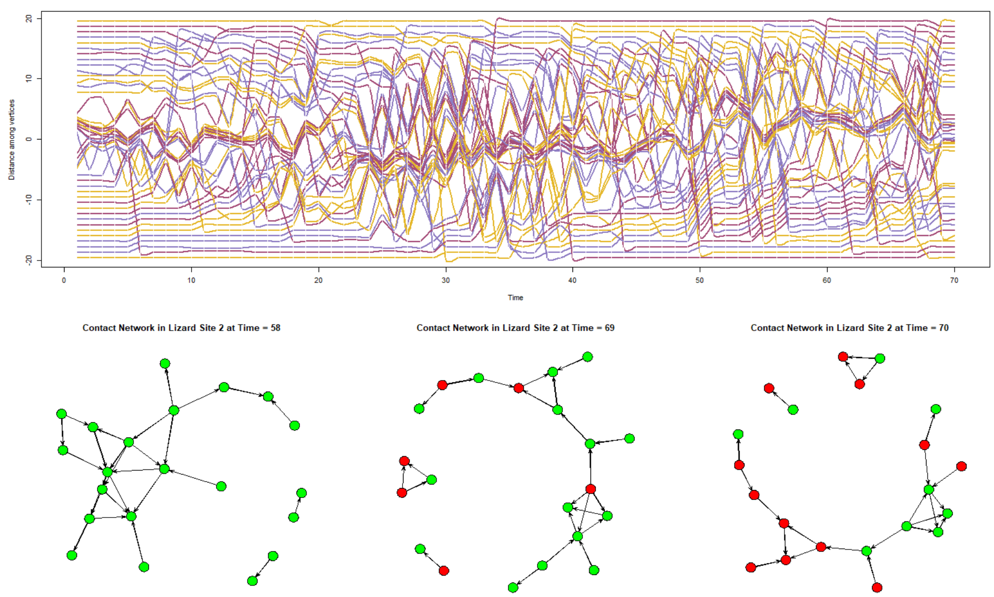


**Fig. 11.** Trace of contact networks in site 2

The diagrams in Fig. 11 depict the contact networks within site 2 at three different time points: 58, 69, and 70. It is evident from these illustrations that following a recognized change in the timeline, there has been a noticeable surge in the number of confirmed infection cases. Moreover, the graphs depict the varying proximity between nodes throughout the duration of the study, with the vertical axis indicating the relative closeness or distance between vertices. Each vertex's movement across time steps is connected by a spline, resulting in a horizontal trajectory resembling a timeline. Furthermore, the histogram plot in Fig. 12 depicting the number of infections corroborates the observed change point in the fortnight containing day 58, revealing a substantial shift in infection rates in the subsequent two weeks. The meticulous application of ACLR, complemented by detailed network analysis, not only enriches our comprehension of infectious disease dynamics but also underscores the critical role of change analysis in contact networks in shaping transmission patterns. This approach not only deepens our understanding of infectious disease dynamics but also highlights the crucial role of changes in contacts in shaping transmission patterns.

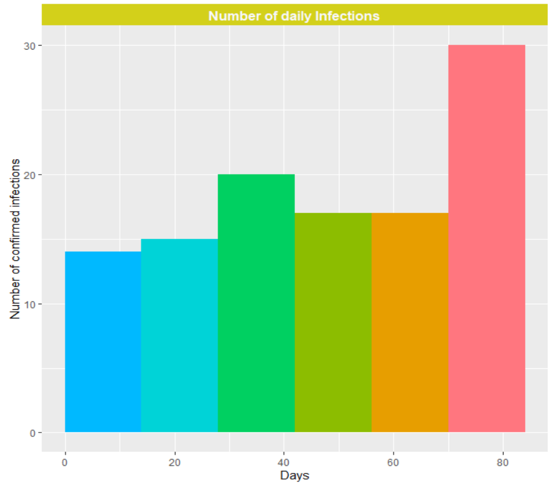


**Fig. 12.** Number of fortnightly infections in site 2

### *5.2.2 Ask Ubuntu*

Ask Ubuntu is a collaborative Q&A platform where the Ubuntu community comes together to ask and answer questions about the Ubuntu operating system. It operates on the Stack Exchange Network. It stands as the flagship site within the extensive Stack Exchange network, and facilitates knowledge-sharing, enabling individuals to enhance their expertise and propel their professional journeys. The website operates as a dynamic hub where users can pose questions and provide answers. Moreover, active participation involves members voting on the merit of both questions and answers. This platform fosters a collaborative environment, where users engage in constructive discussions through comments on various posts.

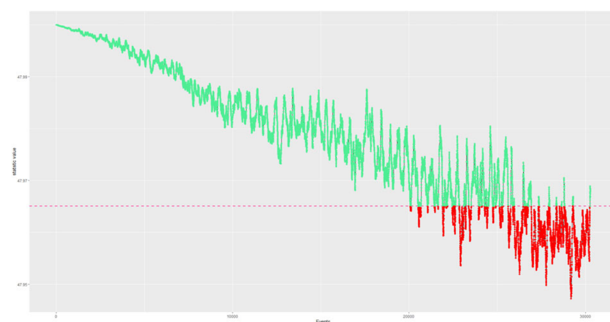

**Fig. 13.** Obtained control chart for Ask Ubuntu events monitoring

In this section, we consider the event datasets on Ask Ubuntu. The edges denoted as $(u, v, t)$ signify the occurrence of an event at a specific time $t$. These events encompass user $u$: 1) contributing answers to questions posed by user $v$, 2) offering comments on queries posted by user $v$, or 3) participating in discussions through comments on answers provided by user $v$. Fig. 13 illustrates the obtained control chart for the observed events in Ask Ubuntu website. Upon conducting a detailed investigation of the identified change points, it was revealed that they correlate with the changes corresponding to new releases of Ubuntu in April, July, September, and October of 2015 (Miller & Mokryn, 2020). The proposed chart based on ACLR method demonstrates its efficacy in facilitating change point detection analysis within social network platforms.

Fig. 14 and Fig. 15 illustrate the correlation of network closeness centrality measures among events from January to May. A user with high closeness centrality may be someone who frequently provides valuable answers, engages in discussions, or actively participates in the community. Their presence contributes to the efficient flow of information and interactions within the community. The figure indicates a noticeable increase in the average closeness centrality values for all nodes in the network in March, followed by a further increase in April. This period is highlighted as a significant point in time, representing a change in observed events on Ask Ubuntu. The substantial rise in user engagement during this time could potentially be attributed to issues related to a new version's release. Understanding change points on the Ask Ubuntu website offers numerous advantages. Firstly, it empowers the platform to optimize performance by aligning server resources with shifts in user behavior, ensuring a seamless experience. Moreover, recognizing change points enhances user engagement through tailored content, features, and interface elements that align with evolving user preferences, consistently improving the overall user experience. Change points also provide valuable insights for strategic decision-making and are pivotal for supporting the business and revenue strategy of Ask Ubuntu.

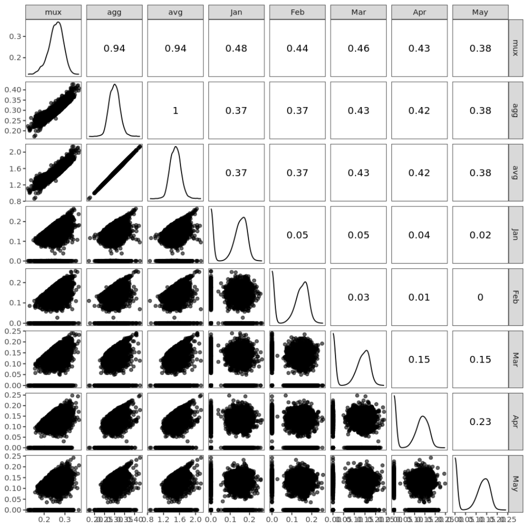


Fig. 14. Comparison of network closeness centrality from Jan to May

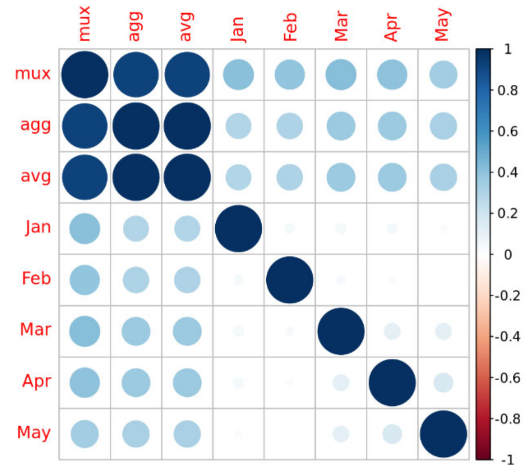


**Fig. 15.** Correlation of network closeness centrality from Jan to May

*5.2.3 Facebook data*

In this experiment, the proposed ACLR method is applied to the Facebook wall posts dataset. The events in this dataset are wall posts among users on the social network Facebook located in the New Orleans region. Fig. 16 illustrates the control chart depicting the change point in Facebook wall posts. Upon closer analysis, the identified change point is associated with the mid-2008 period, during which the rate of user activities decreased. Subsequently, Facebook introduced a new design for wall posts, leading to a subsequent increase in user engagement (Wang & Resnick, 2021).The effectiveness of the suggested chart becomes apparent as it aids in the analysis of change points within social network platforms. This proposed chart proves valuable for detecting shifts, highlighting its utility in examining variations in the dynamics of Facebook wall posts.

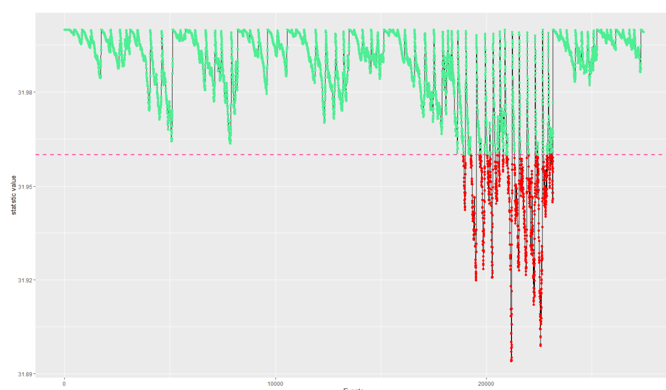


**Fig. 16.** Obtained control chart for Facebook wall posts monitoring

Comprehending the fluid nature of Facebook wall posts is paramount for the platform's operational efficacy and user satisfaction. Primarily, this awareness enables Facebook to finely tune user engagement by tailoring algorithms to individual preferences, ensuring the delivery of more pertinent content. Recognizing patterns in posting behavior serves to keep users actively engaged and invested in the platform, contributing to a positive user experience. Furthermore, monitoring of wall post dynamics allows Facebook to enforce content moderation policies with precision. The platform can promptly identify and address inappropriate content, fostering a secure and respectful online environment. This vigilance also enables Facebook to adapt swiftly to emerging trends, nurturing community building and social connections. Such adaptability not only bolsters user satisfaction and retention but positions Facebook as a dynamic and innovative platform, continuously evolving to meet the dynamic needs of its diverse user base.

## 6. Conclusion

This research addressed the challenging problem of adaptive sequential anomaly detection in observed events in social networks, specifically focusing on scenarios where the post-change parameters of the data distribution are unknown. This problem is paramount in various real-world system monitoring tasks, particularly in anomaly detection within social networks, where the underlying statistical properties can undergo significant changes over time. This study introduced an innovative approach that uses online convex optimization methods, notably the adaptive moment estimation (ADAM) algorithm. By integrating this estimation method into the sequential anomaly detection procedure, this study offers a practical solution for identifying change points in the stream of social network data. This approach addresses the limitations of recursive maximum likelihood estimators, particularly in scenarios where they might be impractical or unavailable. The experimental results demonstrate the efficiency of the proposed algorithm across a spectrum of scenarios, including both synthetic and real-world datasets. The results indicate the practical applicability of the Adaptive Cumulative Likelihood Ratio (ACLR) algorithm in social network monitoring, highlighting its potential to improve the accuracy and reliability of anomaly detection in dynamic social network environments. The proposed algorithm is a significant step forward in addressing the challenges of monitoring social networks where post-change parameter estimation is a non-trivial task. By combining online convex optimization methods with sequential likelihood ratios, this research provides a valuable contribution to the field of statistical process control, with potential applications in various domains beyond social network monitoring, including the analysis of earthquake patterns through spatiotemporal data, price analysis in financial markets, and the popularity of tweets.

Future research could be on enhancing algorithm robustness, addressing the challenges of multimodal data sources, improving interpretability, and fostering interdisciplinary collaboration. An important direction for theoretical investigations would be replacing the full likelihood with an appropriate pseudo-likelihood. These avenues will advance the state-of-the-art in anomaly detection, making it more effective and adaptable in dynamic and complex real-world data.

## References


Ali, F., El-Sappagh, S., Islam, S. M. R., Ali, A., Attique, M., Imran, M., & Kwak, K.-S. (2021). An intelligent healthcare monitoring framework using wearable sensors and social networking data. *Future Generation Computer Systems*, *114*, 23–43.

Aminikhanghahi, S., & Cook, D. J. (2017). A survey of methods for time series change point detection. *Knowledge and Information Systems*, *51*(2), 339–367.

Azoury, K. S., & Warmuth, M. K. (2001). Relative loss bounds for on-line density estimation with the exponential family of distributions. *Machine Learning*, *43*, 211–246.

Cao, Y., Xie, L., Xie, Y., & Xu, H. (2018). Sequential change-point detection via online convex optimization. *Entropy*, *20*(2), 108.

Cesa-Bianchi, N., & Lugosi, G. (2006). *Prediction, learning, and games*. Cambridge university press.

Chen, X., Liu, S., Sun, R., & Hong, M. (2019). On the convergence of a class of Adam-type algorithms for non-convex optimization. *7th International Conference on Learning Representations, ICLR 2019*.

Drury, B., Drury, S. M., Rahman, M. A., & Ullah, I. (2022). A social network of crime: A review of the use of social networks for crime and the detection of crime. *Online Social Networks and Media*, *30*, 100211.

Fisch, A. T. M., Eckley, I. A., & Fearnhead, P. (2022). A linear time method for the detection of collective and point anomalies. *Statistical Analysis and Data Mining: The ASA Data Science Journal*, *15*(4), 494–508.

He, F., Mao, T., Hu, T., & Shu, L. (2018). A new type of change-detection scheme based on the window-limited weighted likelihood ratios. *Expert Systems with Applications*, *94*, 149–163.

Heusel, M., Ramsauer, H., Unterthiner, T., Nessler, B., & Hochreiter, S. (2017). Gans trained by a two time-scale update rule converge to a local nash equilibrium. *Advances in Neural Information Processing Systems*, *30*.

Karami, A., & Niaki, S. T. A. (2024). An Online Support Vector Machine Algorithm for Dynamic Social Network Monitoring. *Neural Networks*, *171*, 497-511.

Karimi, M., & Sadjadi, S. J. (2024). Optimization of an economic ordering quantity model for non-instantaneous deteriorating items with ordering time constraint using dynamic programming. *Journal of Industrial and Management Optimization*, *20*(3), 1114–1141.

Kingma, D. P., & Ba, J. (2014). Adam: A method for stochastic optimization. *ArXiv Preprint ArXiv:1412.6980*.

Kotłowski, W., & Grünwald, P. (2011). Maximum likelihood vs. sequential normalized maximum likelihood in on-line density estimation. *Proceedings of the 24th Annual Conference on Learning Theory*, 457–476.

Kovács, S., Bühlmann, P., Li, H., & Munk, A. (2023). Seeded binary segmentation: a general methodology for fast and optimal changepoint detection. *Biometrika*, *110*(1), 249–256.

Lai, T. L., & Xing, H. (2010). Sequential change-point detection when the pre-and post-change parameters are unknown. *Sequential Analysis*, *29*(2), 162–175.

Li, S., Xiao, S., Zhu, S., Du, N., Xie, Y., & Song, L. (2018). Learning temporal point processes via reinforcement learning. *Advances in Neural Information Processing Systems*, *31*.

Lorden, G. (1971). Procedures for reacting to a change in distribution. *The Annals of Mathematical Statistics*, 1897–1908.

Mariani, S., & Cawley, P. (2021). Change detection using the generalized likelihood ratio method to improve the sensitivity of guided wave structural health monitoring systems. *Structural Health Monitoring*, *20*(6), 3201–3226.

Mei, Y. (2006). *Sequential change-point detection when unknown parameters are present in the pre-change distribution*.

Miller, H., & Mokryn, O. (2020). Size agnostic change point detection framework for evolving networks. *Plos One*, *15*(4), e0231035.

Pollak, M. (1985). Optimal detection of a change in distribution. *The Annals of Statistics*, 206–227.

Pollak, M. (1987). Average run lengths of an optimal method of detecting a change in distribution. *The Annals of Statistics*, 749–779.

Raginsky, M., Marcia, R. F., Silva, J., & Willett, R. M. (2009). Sequential probability assignment via online convex programming using exponential families. *2009 IEEE International Symposium on Information Theory*, 1338–1342.

Raginsky, M., Willett, R. M., Horn, C., Silva, J., & Marcia, R. F. (2012). Sequential anomaly detection in the presence of noise and limited feedback. *IEEE Transactions on Information Theory*, *58*(8), 5544–5562.

Romano, G., Eckley, I. A., Fearnhead, P., & Rigaill, G. (2023). Fast online changepoint detection via functional pruning CUSUM statistics. *Journal of Machine Learning Research*, *24*, 1–36.

Ruggieri, E., & Antonellis, M. (2016). An exact approach to Bayesian sequential change point detection. *Computational Statistics & Data Analysis*, *97*, 71–86.

Shiryaev, A. N. (1963). On optimum methods in quickest detection problems. *Theory of Probability & Its Applications*, *8*(1), 22–46.

Shiryaev, A. N. (2010). Quickest detection problems: Fifty years later. *Sequential Analysis*, *29*(4), 345–385.

Siegmund, D. O., & Yakir, B. (2008). Minimax optimality of the Shiryayev–Roberts change-point detection rule. *Journal of Statistical Planning and Inference*, *138*(9), 2815–2825.

Song, G., Li, Y., Chen, X., He, X., & Tang, J. (2016). Influential node tracking on dynamic social network: An interchange greedy approach. *IEEE Transactions on Knowledge and Data Engineering*, *29*(2), 359–372.

Tartakovsky, A. G., Rozovskii, B. L., Blazek, R. B., & Kim, H. (2006). A novel approach to detection of intrusions in computer networks via adaptive sequential and batch-sequential change-point detection methods. *IEEE Transactions on Signal Processing*, *54*(9), 3372–3382.

Wang, H., Xie, L., Xie, Y., Cuozzo, A., & Mak, S. (2023). Sequential Change-Point Detection for Mutually Exciting Point Processes. *Technometrics*, *65*(1), 44–56.

Wang, T., & Resnick, S. I. (2021). Common Growth Patterns for Regional Social Networks: A Point Process Approach. *Journal of Data Science*, 1–24.

Xie, L., Moustakides, G. V, & Xie, Y. (2023). Window-limited CUSUM for sequential change detection. *IEEE Transactions on Information Theory*.

Zarepour, M., & Habibi, R. (2023). A quasi-Bayesian change point detection with exchangeable weights. *Journal of Statistical Planning and Inference*, *222*, 226–240.

Zhu, X., Li, Y., Liang, C., Chen, J., & Wu, D. (2013). Copula based change point detection for financial contagion in chinese banking. *Procedia Computer Science*, *17*, 619–626.