Bayesian Analysis (2024) **19**, Number 2, pp. 649–676

Changepoint Detection on a Graph of Time Series[*]

Karl L. Hallgren[†], Nicholas A. Heard[‡], and Melissa J. M. Turcotte[§]

Abstract. When analysing multiple time series that may be subject to changepoints, it is sometimes possible to specify *a priori*, by means of a graph, which pairs of time series are likely to be impacted by simultaneous changepoints. This article proposes an informative prior for changepoints which encodes the information contained in the graph, inducing a changepoint model for multiple time series that borrows strength across clusters of connected time series to detect weak signals for synchronous changepoints. The graphical model for changepoints is further extended to allow dependence between nearby but not necessarily synchronous changepoints across neighbouring time series in the graph. A novel reversible jump Markov chain Monte Carlo (MCMC) algorithm making use of auxiliary variables is proposed to sample from the graphical changepoint model. The merit of the proposed approach is demonstrated through a changepoint analysis of computer network authentication logs from Los Alamos National Laboratory (LANL), demonstrating an improvement at detecting weak signals for network intrusions across users linked by network connectivity, whilst limiting the number of false alerts.

Keywords: changepoint detection, graphical model, informative prior, auxiliary variable MCMC, cyber-security.

1 Introduction

Consider $N > 1$ time series of random observations

$$\{x_{i,t} \,|\, 1 \leqslant i \leqslant N, t \geqslant 0\} \tag{1}$$

which are subject to changepoints. This article will suppose the existence of an underlying graph G on N nodes corresponding to each of the time series, such that changepoints are believed to occur simultaneously or closely together in time for time series connected by edges in G.

A motivating application for considering such dependencies is the task of changepoint detection in cyber-security. To identify the presence of a network intrusion, it is informative to monitor for changes in the authentication activity of each user in the network. However, cyber data often exhibit much variability and apparent changes are not guaranteed to correspond to an attack. As a result, to limit the number of false

[*] The authors acknowledge funding from EPSRC. Research presented in this article was supported by the Laboratory Directed Research and Development program of Los Alamos National Laboratory (New Mexico, USA) under project number 20180607ECR and Los Alamos National Laboratory.

[†] Department of Mathematics, Imperial College London, London, United Kingdom, klh16@ic.ac.uk

[‡] Department of Mathematics, Imperial College London, London, United Kingdom, n.heard@imperial.ac.uk

[§] Microsoft Corporation, Redmond, Washington, United States, melissa.turcotte@microsoft.com

https://doi.org/10.1214/23-BA1365

alerts and yet not overlook weak signals from genuine, small attack footprints, it is key to incorporate expert knowledge in the change detection procedure. A commonly held belief of security experts is that attacks are *a priori* likely to be identified through quasi-simultaneous changes in the behaviour of users that are linked by network connectivity (Sexton et al., 2015). Hence, it is of interest to encode a changepoint prior by means of a graph G representing the network of users, such that pairs of connected users in G are *a priori* more likely to be affected by quasi-simultaneous behavioural changes.

Limited attention has previously been given to encoding prior beliefs on graph-based dependence structure of discrete-time changepoints across multiple time series. Existing changepoint model for multiple time series, which admit changepoints may simultaneously affect a subset of the time series, typically assume *a priori* changepoint locations are exchangeable across time series (Jeng et al., 2012; Bardwell and Fearnhead, 2017; Bolton and Heard, 2018; Wang and Samworth, 2018; Bardwell et al., 2019; Grundy et al., 2020). Moreover, with the exception of Fisch et al. (2022), dependent changepoints across time series are often assumed to perfectly align, which is a limiting assumption in cyber-security monitoring where attacks may span a substantial period of time.

More generally, graphical models provide a useful framework for characterising joint distributions for random variables: the nodes of the graph identify the random variables and the edges characterise dependencies among these variables (Lauritzen, 1996). In particular, graphical models have been employed to encode prior beliefs, for example, in the context of Bayesian variable selection for regression models. Li and Zhang (2010) assumes that covariates lie on an undirected graph and formulates an Ising model prior on the covariate space to incorporate structural information.

This article proposes an informative, graphical model-based prior for changepoints that encodes beliefs on the dependence structure of changepoints across time series (1). For practical purposes, changepoints are represented in discrete time by a binary matrix $\boldsymbol{S} = (S_{i,t})$, such that $S_{i,t}$ indicates whether the time point t is a changepoint for the time series with index i. Then, extending the standard memoryless prior for changepoints (Fearnhead, 2006), independent and identical Markov random fields (Lauritzen, 1996) with respect to G are assumed *a priori* for the columns of $\boldsymbol{S}$. As a result, the model assumes that clusters of time series (according to G) are likely to be simultaneously affected by changepoints. Conditional on changepoints, the time series data are assumed to be independent of G and to follow a standard parametric changepoint model (Fearnhead, 2006). A key consequence of the graphical model is that stronger evidence from data is required to infer scattered synchronous changepoints than synchronous changepoints clustered according to G. Furthermore, a more general model is proposed that admits related changepoints not occurring at exactly the same time; the extended model supposes that changepoints may cluster according to G within some finite time windows of possibly unknown lengths, which are specific to each series.

A common approach to sampling changepoints for a single time series is that of Green (1995), using a reversible jump MCMC algorithm to explore the state space of changepoints: at each iteration of the algorithm, a new changepoint is proposed, or else an existing changepoint is either deleted or shifted to a new position. Specifying a joint model for dependent changepoints across multiple time series introduces additional computational challenges that are not present when changepoints are inferred for each time series independently. A simulation study will demonstrate that it can be impractical to simply propose updates to the changepoints of a randomly chosen time series via one of the moves of Green (1995). To efficiently explore the state space of dependent changepoints, it is necessary to consider joint proposals for changepoints across multiple time series.

We propose an MCMC algorithm making use of auxiliary variables (Besag and Green, 1993) to sample from the posterior distribution. Swendsen and Wang (1987) and Higdon (1998) provide notable examples of use of auxiliary variables in MCMC schemes that improve mixing and convergence for undirected graphical models. In brief, our sampling strategy is the following. The changepoint parameter space is augmented with auxiliary variables that induce clusters of time series according to the dependence graph G. Then, the MCMC algorithm of Green (1995) is extended to sample from the augmented parameter space, such that, at each iteration of the algorithm, a new cluster of changepoints may be proposed or an existing cluster of changepoints may be deleted or shifted.

Bayesian inference for changepoints quantifies uncertainty about the number and the positions of changepoints. However, in some applications such as cyber-security, it will also be necessary to report a point estimate for changepoint parameters. Yet no existing loss function in the literature seems suitable for taking into account both the number and the positions of changepoints. To address this gap, we propose using matchings in graphs (Bondy and Murty, 1976) to define a novel loss function for changepoints, which can be used to obtain a point estimate from a posterior sample of candidate changepoints.

The practical benefits of the proposed graphical model are demonstrated via a changepoint analysis of real computer network authentication data from Los Alamos National Laboratory (LANL), where a subset of the data relating to a ‘red team’ exercise provide a proxy for intruder behaviour (Kent, 2015). The challenge consists of monitoring for temporal changes in the authentication activity of network users to detect the presence of red team actors. The proposed changepoint prior is used to encode beliefs that signals for network intrusions are *a priori* likely to occur at nearby times for users historically linked by previous network connectivity. We show that, as a consequence, the proposed model can detect weak signals for red team activity in the network, whilst limiting the number of false alerts, in contrast with a standard model assuming independence of behavioural changes across users.

Finally, it should be noted that, in contrast with recent changepoint detection methods (Chen and Zhang, 2015; Chen, 2019a,b; Chu and Chen, 2019), the focus of this article is not the temporal evolution of a graph subject to changepoints. The graph G represents prior information that can be exploited to detect changepoints in time series.

The remainder of the article is organised as follows. Section 2 motivates our work with a cyber-security application. Section 3 presents Bayesian changepoint modelling for multiple time series. Section 4 introduces a novel, graph-based informative prior for changepoints. Section 5 proposes an auxiliary variable MCMC sampling strategy. Section 6 proposes a novel loss function for assessing changepoints. Section 7 presents results of a changepoint analysis of network authentication data, illustrating the practical benefits of the proposed model. The supplementary material (Hallgren et al., 2023) presents some technical material in Appendices A, B, C and a simulation study in Appendix D demonstrating the model introduced in Section 4.

2 Motivational application: changepoint detection in cyber-security

To motivate an informative graph-based prior for changepoints, we consider an application of changepoint detection in cyber-security. A cyber-attack typically changes the behaviour of connected endpoints on the target computer network (Sexton et al., 2015). Therefore, to detect the presence of a network intrusion, it is informative to monitor for synchronous, or quasi-synchronous, changes in the behaviour of entities that are *a priori* known to be linked by network connectivity.

2.1 Change detection in the authentication activity of users

Kent (2015) presents a comprehensive data set summarising 58 days of traffic on the enterprise computer network of Los Alamos National Laboratory (LANL), which is available online at `https://lanl.ma.ic.ac.uk/data/cyber1`. The network authentication data consist of records describing authentication activity of users connecting from one computer to another. The occurrence of a 'red team' penetration testing operation during the data collection period makes these data suitable for testing network intrusion detection methods. Further details on the data are given in Appendix A.1 in the supplementary material (Hallgren et al., 2023).

Let V denote the set of users in the enterprise. To detect occurrences of malicious activity in the network, the authentication activity of each user $i \in V$ is monitored via hourly counts of network logons per source computer. Let M denote the number of distinct source computers in the network. For each user $i \in V$, let

$$x_{i,t} = (x_{i,t,1}, \ldots, x_{i,t,M}), \tag{2}$$

where $x_{i,t,\ell}$ denotes the number of network logons initiated by user i from source computer ℓ during the t-th hour of the 58 day data collection period. For each user, it is of interest to detect temporal changes in the distribution of network logons across source computers as possible evidence for malicious activity. Figure 1 in the supplementary material (Hallgren et al., 2023) displays the authentication data for two users.

2.2 Motivation for an informative graph-based changepoint prior

The authentication data (2) exhibit much variability, and some observed changes can correspond to legitimate activity. Therefore, to limit the number of false alerts and yet not overlook weak signals from genuine attack footprints, it is key to incorporate prior knowledge in the change detection procedure.

When attackers penetrate a network, they rarely gain access to the target users directly; instead, they typically take control of a vulnerable user, for example via email phishing, and then they move laterally through the network, gaining access and compromising additional users, to achieve their objectives (Sexton et al., 2015). Attackers are typically constrained in the way they can navigate the network, and it will often be possible for cyber-security experts to specify a graph $G = (V, E)$, where an edge $(i, i') \in E \subseteq V \times V$ indicates it is believed *a priori* that attackers may switch credentials between user i and user i' at any time during the data collection period. Therefore, it is of interest to encode in the changepoint prior that cyber-attacks are *a priori* likely to result in quasi-synchronous changes in the authentication activity of multiple users that are connected in G. In this article, we consider the following specification of G for demonstration purposes: $(i, i') \in E$ if and only if both user i and user i' successfully initiated a network logon from the same source computer on the same day. This choice follows from the following considerations. In Windows operating systems, when a user logs on with their credentials (username and password hash) to a device on the domain, these credentials are cached locally on the device. Credential caching prevents users from continuously having to re-authenticate (single sign-on), and enables them to log on to the device even if the device is disconnected from the network. Attackers will exploit credentials which are cached on devices to upgrade their privileges and move laterally through the network. How long credentials may be cached on devices depends on the enterprise's network settings. In the absence of precise knowledge about the enterprise's network settings, it is reasonable to assume that if both user i and user i' have logged into a device on the same day then both those credentials may be cached on that device during the data collection period. As a result, if attackers had access to that device then they would have the ability to exploit cached credentials to switch credentials between user i and user i'.

In Figure 1, for the application of interest, each arrow corresponds to the authentication activity of a user on the network, and shaded rectangles indicate which pairs of

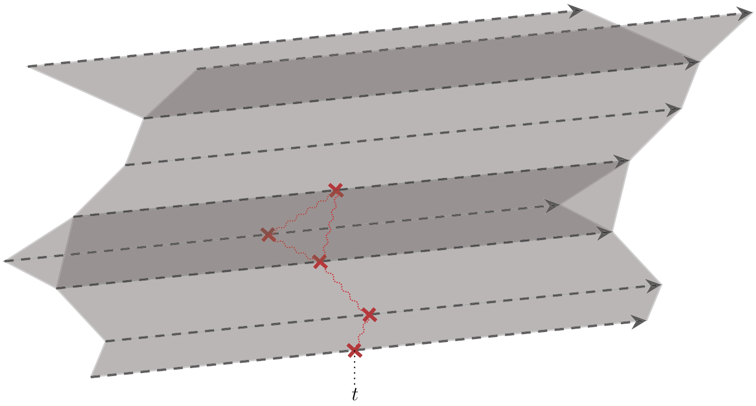

Figure 1: Cartoon representation of a cluster of synchronous changepoints (red crosses) on a graph of time series. Arrows represent time series, and shaded rectangles indicate which pairs of time series are likely to be impacted by simultaneous changepoints.

users are connected in G and therefore likely to be impacted by simultaneous changes during an attack. It is of interest to encode in the changepoint prior, by means of the graph G, that pairs of users $(i, i') \in E$ are likely to be simultaneously affected by malicious behavioural changes, thereby inducing a changepoint model for the authentication data that borrows strength across connected users in G to detect signals for clusters of synchronous changes, as sketched in Figure 1.

In contrast with recent intrusion detection methods (Chen and Zhang, 2015; Chen, 2019a; Metelli and Heard, 2019; Passino et al., 2021), the focus of this article is not the temporal evolution of a graph representing a network, and both V and E are constant in time. The graph G represents the best available static characterisation of the network that can be used to guide change detection in the authentication activity of users (2), and it is assumed to be readily available prior to running network intrusion detection methods; note that, in practice, the edge set could be derived from historic data. Section 8 discusses possible model extensions for settings where prior beliefs on which time series are likely to be impacted by simultaneous changepoints may be time-dependent.

3 Changepoint analysis for multiple time series

Let $G = (V, E)$ be a graph with node set $V = \{1, \ldots, N\}$ and edge set $E \subseteq V \times V$. For each node $i \in V$ we observe a time series $\boldsymbol{x}_i = (x_{i,0}, \ldots, x_{i,T})$ which may be subject to changepoints, and the edge set $E \subseteq V \times V$ indicates which pairs of time series are *a priori* likely to be impacted by quasi-simultaneous changepoints. Conditionally on changepoints, the data are assumed to be independent of G and follow a standard parametric changepoint model, presented in this section. Some limitations of the usual prior for independent changepoints are discussed, paving the way for the proposed informative prior for graph-dependent changepoints.

3.1 Model and notation

For each node $i \in V$, suppose there are $k_i \geqslant 0$ changepoints that partition the time series of observations for that node into $k_i + 1$ segments. The ordered locations of the changepoints, denoted by $\boldsymbol{\tau}_i = (\tau_{i,1}, \ldots, \tau_{i,k_i})$, belong to the set $\mathcal{T}_{k_i}$, where

$$\mathcal{T}_k = \left\{(\tau_1, \ldots, \tau_k) \in \mathbb{N}^k;\ 0 \equiv \tau_0 < \tau_1 < \cdots < \tau_k < \tau_{k+1} \equiv T + 1\right\}. \tag{3}$$

For each node i, the data $x_{\tau_{i,j-1}}, \ldots, x_{\tau_{i,j}-1}$ in each segment j are assumed to be drawn from a distribution from the same parametric family $L_i(\cdot|\theta_{i,j})$, with a segment specific parameter $\theta_{i,j}$ drawn independently from a prior density $\pi_i(\cdot)$.

The parameters of interest are the changepoint parameters $(\boldsymbol{k}, \boldsymbol{\tau})$, where $\boldsymbol{k} = (k_i)_{i \in V}$ and $\boldsymbol{\tau} = (\boldsymbol{\tau}_i)_{i \in V}$. Motivated by computational considerations, as in Fearnhead (2006) it is assumed in this article that segment parameters may be marginalised so that the likelihood of the data $\boldsymbol{x}$ conditional on changepoints,

$$\mathcal{L}(\boldsymbol{x}|\boldsymbol{k}, \boldsymbol{\tau}) = \prod_{i \in V} \prod_{j=1}^{k_i+1} \mathcal{L}_i(\tau_{i,j-1}, \tau_{i,j}), \tag{4}$$

where

$$\mathcal{L}_i(\tau_{i,j-1}, \tau_{i,j}) = \int L_i(x_{i,\tau_{i,j-1}}, \dots, x_{i,\tau_{i,j}-1} | \theta_{i,j}) \pi_i(\theta_{i,j}) d\theta_{i,j} \tag{5}$$

can be computed. Given a prior for the changepoint parameters, $\pi(\boldsymbol{k}, \boldsymbol{\tau})$, one can consequently compute the posterior density function for the changepoint parameters, up to a normalising constant.

Examples of changepoint models where segment parameters may be marginalised include models for independent and identically distributed data within segments (Fearnhead, 2006; Denison et al., 2002), changing linear regressions (Punskaya et al., 2002; Carlin et al., 1992), models for time-dependent data within segments, such as Markov models with time-varying transition matrices (Bolton and Heard, 2018), zero-mean and heteroscedastic processes with changing variance (Johnson et al., 2003), and changepoint models with segment parameters subject to seasonal effects (Turcotte, 2014). Moreover, some model extensions where segment parameters cannot be marginalised, and where segment parameters may be shared across segments, are discussed in Appendix C.2 in the supplementary material (Hallgren et al., 2023), indicating how the proposed sampling strategy could be adapted for these model extensions.

In particular, consider the class of changepoint models where, within each segment, the data are assumed to be independent and identically distributed such that

$$x_{i,t} \sim f_i(\,\cdot\,|\,\theta_{i,j}), \quad \tau_{i,j-1} \leqslant t < \tau_{i,j}, \tag{6}$$

for some parametric density $f_i(\,\cdot\,|\,\theta_{i,j})$ dependent on some segment parameter $\theta_{i,j} \sim \pi_i(\cdot)$. The integrals in (5) can be calculated analytically when π_i is chosen to be conjugate to f_i; and for non-conjugate cases, (5) may be calculated numerically for low-dimensional segment parameters. For the cyber-security application discussed in Section 2, the changepoint model (6) is suitable for the count data with, for all i, f_i denoting the density of the multinomial distribution with unknown probability parameter vectors $\theta_{i,j}$ with an uninformative, conjugate prior Dirichlet($\mathbf{1}^M$), where $\mathbf{1}^M$ denotes the M-dimensional vector of ones. As a result, each changepoint $\tau_{i,j}$ corresponds to a temporal change in the distribution of counts of logons initiated by the user $i \in V$ across M host computers in the network.

3.2 Limitations of the standard prior for independent changepoints

When changepoints are assumed to be independent across time series, the posterior distribution of changepoints can be estimated for each time series separately. In this setting, it is standard to assume *a priori* that, for all time series, discrete time changepoints follow a Bernoulli process (Fearnhead, 2006) such that

$$\pi(\boldsymbol{k}, \boldsymbol{\tau}|p) = \prod_{i \in V} p^{k_i}(1-p)^{T-k_i} \mathbb{1}_{\mathcal{T}_i}(\boldsymbol{\tau}_i) \tag{7}$$

for some Bernoulli parameter $0 < p < 1$, which encodes prior belief on the expected number of changepoints.

For the cyber-security application where G represents a network of users, the standard prior in (7) cannot fully encode prior beliefs on changepoints. Appendix A.2 in the supplementary material (Hallgren et al., 2023) exposes limitations resulting from the assumption of changepoint independence across time series through a comparative study. No choice of p seems satisfactory: choosing a small value for p will limit the number of false alerts due to noise in user-specific legitimate activity; yet it will also prevent the detection of weak signals for changes shared by different users which are linked in the network, that may be of great interest. It would be preferable to specify *a priori* that changepoints are more likely to occur simultaneously across time series that are linked in G, in order to require strong evidence from the data for changes impacting a single user, or possibly weak signals for changes that impact multiple users linked in the network.

4 Graphical models for dependent changepoints across multiple time series

This section proposes a novel graphical prior for dependent changepoints across multiple time series. Given the graph of time series $G = (V, E)$, where $V = \{1, \dots, N\}$, changepoints are modelled by means of an undirected graphical model encoding that pairs of time series $(i, i') \in E$ are *a priori* likely to be simultaneously affected by changepoints. The graphical model is further extended by relaxing the assumption that dependent changepoints across time series are synchronous; the extended model assumes dependent changepoints across time series correspond to nearby but not necessarily identical time points.

4.1 Synchronous dependent changepoints across time series

4.1.1 Model definition

In Section 3.1, changepoints were most simply defined in terms of their number and locations, $(\boldsymbol{k}, \boldsymbol{\tau})$. Subsequently, it will be useful to represent changepoints by means of a binary matrix. For changepoint parameters $(\boldsymbol{k}, \boldsymbol{\tau})$, let $\boldsymbol{S} = (S_{i,t})$ be the corresponding binary matrix such that, for all $i \in V$ and $t = 1, \dots, T$,

$$S_{i,t} = \begin{cases} 1 & \text{if } \exists j \in \{1, \dots, k_i\} \text{ s.t. } t = \tau_{i,j} \\ 0 & \text{otherwise,} \end{cases} \tag{8}$$

so that $(\boldsymbol{k}, \boldsymbol{\tau})$ and $\boldsymbol{S}$ are equivalent representations of the changepoints. Moreover, let $S_{i,0} = S_{i,T+1} = 1$ for all i.

To encode the dependence structure of synchronous changepoints across time series in $G = (V, E)$, let $\boldsymbol{\lambda} = (\lambda_{i,i'})$ be a symmetric matrix of non-negative edge weights for the graph satisfying $\lambda_{i,i'} > 0$ if and only if $(i, i') \in E$ for all $i, i' \in V$. Then, conditional on $\boldsymbol{\lambda}$, changepoints are assumed to have a prior distribution described by the weighted, undirected graph G such that, for all $(\boldsymbol{k}, \boldsymbol{\tau})$,

$$\pi(\boldsymbol{k}, \boldsymbol{\tau}|p, \boldsymbol{\lambda}) = \frac{1}{Z(p, \boldsymbol{\lambda})} \prod_{t=1}^{T} \exp\left\{\bar{p} \sum_{i \in V} S_{i,t} + \sum_{i<i'} \lambda_{i,i'} S_{i,t} S_{i',t}\right\}, \tag{9}$$

for some $0 < p < 1$, where $\bar{p} = \operatorname{logit}(p) = \log\{p/(1-p)\}$ and some normalising constant $Z(p, \boldsymbol{\lambda})$ that has no convenient closed form in general but will present no computational complications since the MCMC algorithm for changepoint parameters proposed in Section 5 only requires computation of ratios of the prior density (9).

If the edge set E is the empty set, implying $\lambda_{i,i'} = 0$ for all i and i', then the prior distribution in (9) is equivalent to the standard prior for independent changepoints (7); for all changepoint parameters $(\boldsymbol{k}, \boldsymbol{\tau})$ and for all $0 < p < 1$,

$$\pi(\boldsymbol{k}, \boldsymbol{\tau}|p, \boldsymbol{0}) = \prod_{i \in V} p^{\sum_{t=1}^{T} S_{i,t}} (1-p)^{T - \sum_{t=1}^{T} S_{i,t}}, \tag{10}$$

where $\boldsymbol{0}$ is the null matrix. The memoryless property of the standard prior (Fearnhead, 2006) is maintained by the extended prior (9), conditional on fixed value of p. The latter assumes independent and identical Markov random fields (Lauritzen, 1996) for the columns of $\boldsymbol{S}$. The memoryless property would be lost if (9) were marginalised over a prior distribution for p.

The graphical prior distribution (9) takes into account both the number of changepoints across time series and their relative positions; the parameter p controls prior belief on the sparsity of changepoints, and the edge weight parameters $\boldsymbol{\lambda}$ control the synchronisation of changepoints between time series. For all pairs (i, i'), the larger the edge weight $\lambda_{i,i'} > 0$, the higher the probability for time series i and i' to be simultaneously affected by changepoints. Hence, the prior in (9) may specify changepoints are likely to occur simultaneously across clusters of time series according to G.

To understand how to set the changepoint prior parameters p and $\boldsymbol{\lambda}$ in practice, it is instructive to consider the conditional prior distribution of the components of the binary matrix $\boldsymbol{S}$. Under (9), the conditional distribution of $S_{i,t}$ given $\boldsymbol{S}_{-(i,t)} = \{S_{i',t'} : (i', t') \neq (i, t)\}$ is

$$\pi(S_{i,t}|\boldsymbol{S}_{-(i,t)}, p, \boldsymbol{\lambda}) \propto \exp\left\{S_{i,t}\left(\bar{p} + \sum_{i' : (i,i') \in E} \lambda_{i,i'} S_{i',t}\right)\right\}, \quad S_{i,t} \in \{0, 1\}. \tag{11}$$

Therefore, for all i and t, the hyperparameter p corresponds to the prior probability that t is a changepoint for the ith time series given that no changepoints occur at time t for the graph neighbour time series of i; and, for all i' such that $(i, i') \in E$, the interaction parameter $\lambda_{i,i'}$ governs how much the conditional prior probability increases if the neighbour time series i' is impacted by a changepoint at time t. Moreover, to perceive the influence of the changepoint prior parameters on the posterior distribution of changepoints, it is helpful to consider the full conditional distribution of $S_{i,t}$ given

$$\boldsymbol{S}_{-(i,t)} = \{S_{i',t'};\, (i',t') \neq (i,t)\},$$

$$\pi(S_{i,t}|\boldsymbol{S}_{-(i,t)}, p, \boldsymbol{\lambda}, \boldsymbol{x}) \propto \left(\frac{\mathcal{L}_i(\tau_t', t)\mathcal{L}_i(t, \tau_t'')}{\mathcal{L}_i(\tau_t', \tau_t'')}\right)^{S_{i,t}} \pi(S_{i,t}|\boldsymbol{S}_{-(i,t)}, p, \boldsymbol{\lambda})$$
$$\propto \exp\left\{S_{i,t}\left(\log\left\{\frac{\mathcal{L}_i(\tau_t', t)\mathcal{L}_i(t, \tau_t'')}{\mathcal{L}_i(\tau_t', \tau_t'')}\right\} + \bar{p} + \sum_{i':\,(i,i')\in E} \lambda_{i,i'} S_{i',t}\right)\right\}, \tag{12}$$

where $\mathcal{L}_i$ is defined in (5), $\tau_t' = \max\{t' : t' < t, S_{i,t'} = 1\}$ and $\tau_t'' = \min\{t' : t' > t, S_{i,t'} = 1\}$. In essence, p determines the level of evidence required from the data to suggest a changepoint, and the edge weight parameters control, relative to p, how weak signals for synchronous changepoints can be combined across time series.

4.1.2 A special case: identical edge weight parameters

In practice, it will often be natural to assume that, for all $(i, i') \in E$, $\lambda_{i,i'} = \lambda$ for some fixed value $\lambda > 0$. For all i and t, let

$$n_{i,t} = \sum_{i':(i,i')\in E} S_{i',t} \tag{13}$$

be the number of neighbour time series of i that are affected by a changepoint at time t. Then, under (9), the conditional prior distribution of $S_{i,t}$ given $\boldsymbol{S}_{-(i,t)} = \{S_{i',t'} : (i',t') \neq (i,t)\}$ is

$$\pi(S_{i,t}|\boldsymbol{S}_{-(i,t)}, p, \lambda) = \frac{\exp\{S_{i,t}(\bar{p} + \lambda n_{i,t})\}}{\exp\{\bar{p} + \lambda n_{i,t}\} + 1}, \quad S_{i,t} \in \{0, 1\}. \tag{14}$$

Moreover, λ will typically be chosen relative to $\bar{p}$ and the degree distribution of the nodes in G. For example, it can be convenient to assume $\lambda = \lambda_s|\bar{p}|/n$, where n denotes the maximum degree of the nodes in G, for some $\lambda_s > 0$.

4.2 Examples of graphical dependence structures for changepoints

The prior distribution (9) is suitable for a wide variety of settings. This section provides graph motifs that can be regarded as building blocks to encode the dependence structure of changepoints across multiple time series. For these examples, we assume identical non-zero edge weights as considered in Section 4.1.2 and provide some insight on how to choose the changepoint prior parameters p and λ. These exemplar dependence structures for changepoints are explored via a simulation study in Appendix D in the supplementary material (Hallgren et al., 2023).

4.2.1 Lattices

It might be natural to choose the edge set E to induce an $N_1 \times N_2$ lattice graph when the number of time series is $N = N_1 N_2$ for some $N_1, N_2 > 0$. For all i, let $0 \leqslant i_1 \leqslant N_1 - 1$ and $0 \leqslant i_2 \leqslant N_2 - 1$ be the unique natural numbers such that $i = i_2 N_1 + i_1 + 1$. Then

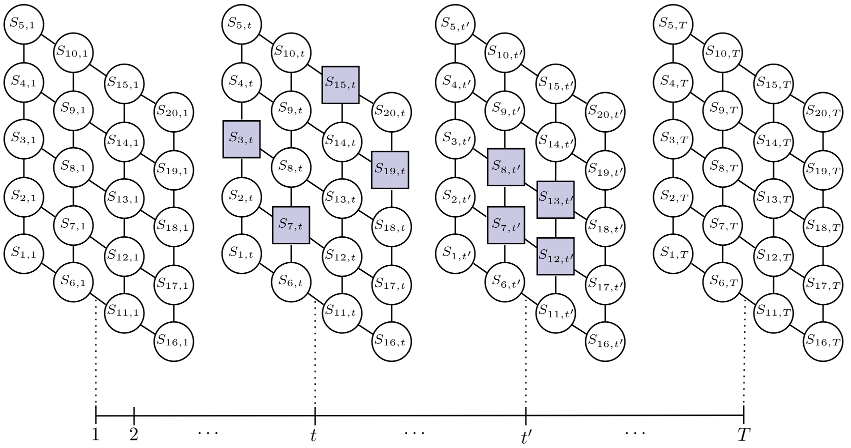

Figure 2: Cartoon representation of a changepoint matrix $\boldsymbol{S} = (S_{i,t})$, defined in (8), for 20 time series of length T which lie on a 5×4 lattice graph G. Edges indicate dependence between components of $\boldsymbol{S}$ according to the prior (9) given the lattice graph G. Blue squares indicate $S_{i,t} = 1$ and white circles indicate $S_{i,t} = 0$ for all i and t.

the $N_1 \times N_2$ lattice graph is such that $(i, i') \in E$ if and only if $|i_1 - i'_1| + |i_2 - i'_2| = 1$. For example, suppose the data $\boldsymbol{x}$ are recorded for the analysis of some spatio-temporal phenomenon such that $x_{i,t}$ denotes the observation at time t and at the coordinate i of some $N_1 \times N_2$ grid over a map of the region of interest, and it is of interest to detect the times and the coordinates at which the distribution of the data changes.

Figure 2 illustrates the dependence structure for changepoints induced by the graphical changepoint prior (9) given a lattice graph G on 20 time series of length T. The larger the edge weight $\lambda > 0$, the higher the probability for pairs of time series connected on the lattice graph G to be simultaneously impacted by changepoints. As a result, in Figure 2, changepoints at time t', which are connected by edges, are *a priori* more likely than isolated changepoints at time t. The prior (9) can therefore specify that changepoints are likely to occur as clusters of simultaneous changepoints on the lattice. The conditional probability (14) specifies that p is the prior probability that a changepoint occurs in isolation on the lattice, and is constrained such that $n_{i,t} \in \{0, \dots, 4\}$.

4.2.2 r-chains

Another dependence structure of interest arises when there is a natural ordering of the time series, which is encoded by the time series indices $1 < \cdots < N$, and changepoints are *a priori* likely to occur as chains of simultaneous changepoints across consecutive time series. For instance, suppose the data consist of multiple time series that are recorded to monitor various aspects of a system; and, it is of interest to detect some event which evolves through multiple phases, such that each phase is likely to manifest through the perturbation of one aspect of the system. In such a setting, it is appropriate to consider the following graph for the time series indices, which we call an r-chain graph: let $(i,i') \in E$ if and only if $1 \leqslant |i-i'| \leqslant r$, for some $r > 0$ chosen to allow gaps of length $r-1$ within chains of changepoints. For r-chain graphs, $n_{i,t} \in \{0,\dots,2r\}$.

4.2.3 Complete graphs

Suppose a complete graph for the time series indices, that is $(i,i') \in E$ for all $i \neq i'$, so that, according to Section 4.1.2, $\lambda_{i,i'} = \lambda > 0$ for all $i \neq i'$. In such a setting, the prior given in (9) assumes changepoint locations are exchangeable across time series, like the Multi-Variate Collective And Point Anomalies (MVCAPA) model proposed in Fisch et al. (2022), and therefore solely takes into account the number of time series impacted by a changepoint at time t, for all t.

4.2.4 Unknown graph

This article assumes that the graph G is known *a priori* and contains useful information concerning the dependence structure of changepoints across time series. Future work could reverse this idea and consider applications where estimating G is one of the inferential objectives. It might often be computationally unrealistic to specify an unconstrained prior for G admitting that $\lambda_{i,j} \geqslant 0$ for all $(i,i') \in V \times V$. However, in some settings it might be appropriate to consider a class of possible graphs G, such as those considered in the previous two subsections; for example, it may be assumed *a priori* that G is an r-chain with $r \geqslant 0$ unknown.

4.3 Extension to asynchronous changepoint dependence

The model in Section 4.1 assumes changepoints are likely to simultaneously affect clusters of time series according to the dependence graph G. In this section, we relax this model to allow dependence between changepoints in different time series at nearby points in time. The extended model relies on representing changepoints as lagged realisations of simultaneous but unobserved latent changepoints. The latent changepoints are distributed according to the model introduced in Section 4.1 and, conditional on these latent changepoints, time series-specific lags are assumed to be uniformly distributed over some small time window.

Let $(\boldsymbol{k},\boldsymbol{\tau})$ be changepoint parameters for multiple time series as defined in Section 3.1, where it is assumed that the ith time series is subject to k_i changepoints whose positions are denoted $\boldsymbol{\tau}_i = (\tau_{i,1},\dots,\tau_{i,k_i}) \in \mathcal{T}_{k_i}$ as defined in (3). The asynchronous model further assumes that, for all time series $i = 1,\dots,N$, there exist latent changepoint positions $\tilde{\boldsymbol{\tau}} = (\tilde{\boldsymbol{\tau}}_1,\dots,\tilde{\boldsymbol{\tau}}_N)$, $\tilde{\boldsymbol{\tau}}_i = (\tilde{\tau}_{i,1},\dots,\tilde{\tau}_{i,k_i}) \in \mathcal{T}_{k_i}$, and lags $\boldsymbol{d} = (\boldsymbol{d}_1,\dots,\boldsymbol{d}_N)$, $\boldsymbol{d}_i = (d_{i,1},\dots,d_{i,k_i}) \in \{0,\dots,w_i\}^{k_i}$, for $\boldsymbol{w} = (w_1,\dots,w_N)$, where $w_i \geqslant 0$ is an upper bound for the lags, such that, for all $j = 1,\dots,k_i$, the jth changepoint for time series i is

$$\tau_{i,j} = \tilde{\tau}_{i,j} + d_{i,j}. \tag{15}$$

Let $\tilde{\tau}_{i,0} = \tau_{i,0} = 0$ and $\tilde{\tau}_{i,k_i+1} = \tau_{i,k_i+1} = T+1$. For all $(k_i, \boldsymbol{\tau}_i)$ and $w_i \geqslant 0$, if $\tilde{\boldsymbol{\tau}}_i = \boldsymbol{\tau}_i$ and $\boldsymbol{d}_i$ is the zero vector then (15) holds, and therefore the existence of a corresponding pair $(\tilde{\boldsymbol{\tau}}_i, \boldsymbol{d}_i)$ with $\boldsymbol{\tau}_i \in \mathcal{T}_{k_i}$ is guaranteed. If $w_i = 0$ then the latent changepoints and the changepoints must be identical; but, in general, given changepoints $(k_i, \boldsymbol{\tau}_i)$ and $w_i > 0$, there are multiple distinct pairs of latent changepoints $(k_i, \tilde{\boldsymbol{\tau}}_i)$ and lags $\boldsymbol{d}_i$ satisfying (15) and $\tilde{\boldsymbol{\tau}}_i \in \mathcal{T}_{k_i}$.

For some applications the upper bounds for the lags, $\boldsymbol{w}$, may be fixed. In particular, for some reference time series $i \in V$, it can be set that $w_i = 0$, implying that $\tau_{i,j} = \tilde{\tau}_{i,j}$ for all j, so that changepoints for time series i' with $w_{i'} \geqslant 0$ are lagged relative to changepoints for time series i. However, in general, upper bounds for the lags will not be known. For example, in the motivational application in cyber-security, no user $i \in V$ can be assumed to be the first user to be affected by an attack, making it awkward to pick a reference time series i, and the exact duration of attacks is not known *a priori*. It will be assumed that, independently for all time series i, $w_i \sim \text{Geometric}(\eta)$ for some value $0 < \eta < 1$ chosen to reflect the expected duration of an attack.

Suppose the latent changepoints $(\boldsymbol{k}, \tilde{\boldsymbol{\tau}})$ are distributed according to the prior distribution (9) given some $0 < p < 1$ and graph edge weight parameters $\boldsymbol{\lambda}$. Then, independently for all time series i, conditional on w_i and $(k_i, \tilde{\boldsymbol{\tau}}_i)$, the lags $\boldsymbol{d}_i = (d_{i,1}, \dots, d_{i,k_i})$ are assumed to be uniformly distributed on the set

$$\mathcal{D}(w_i, k_i, \tilde{\boldsymbol{\tau}}_i) = \left\{(d_{i,1}, \dots, d_{i,k_i}) \in \{0, \dots, w_i\}^{k_i} : \quad (\tilde{\tau}_{i,1} + d_{i,1}, \dots, \tilde{\tau}_{i,k_i} + d_{i,k_i}) \in \mathcal{T}_{k_i}\right\}, \tag{16}$$

such that, for all $\boldsymbol{d} = (\boldsymbol{d}_1, \dots, \boldsymbol{d}_N)$,

$$\pi(\boldsymbol{d} | \boldsymbol{k}, \tilde{\boldsymbol{\tau}}, \boldsymbol{w}) = \prod_{i=1}^{N} \frac{\mathbb{1}_{\mathcal{D}(w_i, k_i, \tilde{\boldsymbol{\tau}}_i)}(\boldsymbol{d}_i)}{\text{card}(\mathcal{D}(w_i, k_i, \tilde{\boldsymbol{\tau}}_i))}. \tag{17}$$

Proposition 1 gives a recursion to derive the cardinality of (16).

Proposition 1. (Cardinality of $\mathcal{D}$). *Let $w \geqslant 0$, $k \geqslant 0$, $\boldsymbol{\tau} = (\tau_1, \dots, \tau_k) \in \mathcal{T}_k$ with $\tau_0 = 1$ and $\tau_{k+1} = T+1$, and let $\mathcal{D}(w, k, \boldsymbol{\tau})$ be the set defined in (16).*

(i) For all $j \geqslant 1$ and $l \geqslant 0$, let $\rho(j, l) = \min\{w+1, T+1-\tau_j\} - (\tau_{j+l} - \tau_j)$ and

$$Q(j, l) = \frac{(\rho(j,l) + l)!}{(\rho(j,l) - 1)!(l+1)!} \mathbb{1}_{\{0, \dots, w\}}(\tau_{j+l} - \tau_j). \tag{18}$$

Additionally, let $Z(0) = 1$, $Z(1) = Q(1, 0)$ and, recursively for all $k > 1$,

$$Z(k) = \sum_{j=1}^{k} (-1)^{k-j} Z(j-1) Q(j, k-j). \tag{19}$$

Then

$$\text{card}(\mathcal{D}(w, k, \boldsymbol{\tau})) = Z(k). \tag{20}$$

In particular, if $k = 0$ then $\boldsymbol{\tau}$ is the empty sequence and $\mathcal{D}(w, k, \boldsymbol{\tau})$ contains a unique element, namely the empty sequence.

(ii) $\mathrm{card}(\mathcal{D}(w, k, \boldsymbol{\tau})) \leqslant (w+1)^k$ *and the equality holds if and only if* $\tau_{j+1} - \tau_j > w$ *for all* j*.*

Proof. See Appendix B in the supplementary material (Hallgren et al., 2023). □

Consequently, the joint prior density for $(\boldsymbol{k}, \tilde{\boldsymbol{\tau}}, \boldsymbol{d})$ is

$$\pi(\boldsymbol{k}, \tilde{\boldsymbol{\tau}}, \boldsymbol{d} | p, \boldsymbol{\lambda}, \boldsymbol{w}) = \frac{\pi(\boldsymbol{k}, \tilde{\boldsymbol{\tau}} | p, \boldsymbol{\lambda})}{\prod_{i=1}^{N} \mathrm{card}(\mathcal{D}(w_i, k_i, \boldsymbol{\tau}_i))} \tag{21}$$

and the induced changepoint prior distribution for $(\boldsymbol{k}, \boldsymbol{\tau})$ is

$$\pi(\boldsymbol{k}, \boldsymbol{\tau} | p, \boldsymbol{\lambda}, \boldsymbol{w}) = \sum_{(\tilde{\boldsymbol{\tau}}, \boldsymbol{d}) \in \Upsilon(\boldsymbol{k}, \boldsymbol{\tau}, \boldsymbol{w})} \pi(\boldsymbol{k}, \tilde{\boldsymbol{\tau}}, \boldsymbol{d} | p, \boldsymbol{\lambda}, \boldsymbol{w}), \tag{22}$$

where $\Upsilon(\boldsymbol{k}, \boldsymbol{\tau}, \boldsymbol{w})$ denotes the set of pairs of latent changepoints and lags, $(\boldsymbol{k}, \tilde{\boldsymbol{\tau}}, \boldsymbol{d})$, that identify the changepoints $(\boldsymbol{k}, \boldsymbol{\tau})$ according to (15).

5 Markov chain Monte Carlo inference

Joint sampling of changepoints across time series is required when the assumption of independence for changepoints is relaxed. In this section, we propose a reversible jump MCMC algorithm (Green, 1995) to sample changepoints in multiple time series, $(\boldsymbol{k}, \boldsymbol{\tau})$. The changepoint parameter space is augmented with auxiliary variables (Besag and Green, 1993; Higdon, 1998) that induce clusters of time series indices according to G. Then, the reversible jump MCMC algorithm of Denison et al. (2002) is extended to sample from the augmented parameter space, thereby providing a means to efficiently explore the changepoint parameter space. At each iteration of the algorithm, a new cluster of changepoints may be proposed or an existing cluster of changepoints may be deleted or shifted. The validity of the proposed MCMC algorithm follows immediately from the reversibility of the proposed moves. Appendix C in the supplementary material (Hallgren et al., 2023) gives some indications on the time complexity of the algorithm and discusses possible extensions for settings where segment parameters cannot be marginalised. For notational simplicity it is assumed there are no missing data, but even with data which are not independent and identically distributed within segments, any missing observations would present no methodological complication, since missing data can be sampled from their predictive distribution within the proposed MCMC scheme (Gelman et al., 2004).

5.1 Sampler for synchronous dependent changepoints

We begin by proposing an MCMC algorithm to sample from the posterior distribution of changepoints when changepoint parameters are *a priori* distributed according to the prior introduced in Section 4.1, $\pi(\boldsymbol{k}, \boldsymbol{\tau} | p, \boldsymbol{\lambda})$, given $0 < p < 1$ and some interaction parameters $\boldsymbol{\lambda}$.

To sample changepoints for multiple time series, consider the following adaptation of the standard MCMC algorithm to sample changepoints for a unique time series (Denison et al., 2002), which will be called the "single site updating" MCMC algorithm thereafter. At each iteration of the algorithm, with $(\boldsymbol{k}, \boldsymbol{\tau})$ denoting the latest particle of the sample chain, propose one of the following two moves: for a uniformly chosen index (i, t), propose $S_{i,t}$ to be updated to $1 - S_{i,t}$, thereby allowing birth or death of a changepoint; alternatively, the position of a randomly chosen changepoint, $\tau_{i,j}$, is sampled uniformly from $\{\tau_{i,j-1} + 1, \ldots, \tau_{i,j+1} - 1\}$.

With the graphical prior distribution (9), synchronous changepoints can be correlated across time series, and the single site updating MCMC algorithm can become impractical, as illustrated in Appendix D.4.3 in the supplementary material (Hallgren et al., 2023) through a simulation study. Instead, it will be necessary to propose moves that allow birth, death or shift of clusters of synchronous changepoints according to the graph induced by $\boldsymbol{\lambda}$.

5.1.1 Augmenting the parameter space with auxiliary variables

To provide a means of moving efficiently through the state space of the changepoint parameters, the parameter space is augmented with binary auxiliary variables $\boldsymbol{u} = (\boldsymbol{u}_1, \ldots, \boldsymbol{u}_T)$ such that, for all t, $\boldsymbol{u}_t$ is an $N \times N$ symmetric binary graph adjacency matrix with (i, i') element $u_t(i, i')$. For all t, the prior density of $\boldsymbol{u}_t$ is assumed to take the conditionally independent form

$$\pi(\boldsymbol{u}_t|\boldsymbol{\lambda}, \delta, \boldsymbol{k}, \boldsymbol{\tau}) = \prod_{i<i'} q_t(i, i')^{u_t(i,i')}\{1 - q_t(i, i')\}^{1-u_t(i,i')}, \tag{23}$$

where

$$q_t(i, i') = 1 - \exp\{-\delta\lambda_{i,i'}(1 - |S_{i,t} - S_{i',t}|)\} \tag{24}$$

is the conditional probability that $u_t(i, i') = 1$, given a partial decoupling parameter $\delta \geqslant 0$ (Higdon, 1998) whose role will be discussed in Section 5.1.2. After observing data $\boldsymbol{x}$ distributed according to (6), the joint posterior density of the augmented parameters $(\boldsymbol{k}, \boldsymbol{\tau}, \boldsymbol{u})$ is

$$\pi(\boldsymbol{k}, \boldsymbol{\tau}, \boldsymbol{u}|p, \boldsymbol{\lambda}, \boldsymbol{x}, \delta) = \pi(\boldsymbol{u}|\boldsymbol{\lambda}, \delta, \boldsymbol{k}, \boldsymbol{\tau})\pi(\boldsymbol{k}, \boldsymbol{\tau}|p, \boldsymbol{\lambda}, \boldsymbol{x}), \tag{25}$$

where

$$\pi(\boldsymbol{u}|\boldsymbol{\lambda}, \delta, \boldsymbol{k}, \boldsymbol{\tau}) = \prod_{t=1}^{T} \pi(\boldsymbol{u}_t|\boldsymbol{\lambda}, \delta, \boldsymbol{k}, \boldsymbol{\tau}). \tag{26}$$

For all t, consider the graph $H_t = (V, E_t)$ with vertex set $V = \{1, \ldots, N\}$ and edge set E_t such that $(i, i') \in E_t$ if and only if $u_t(i, i') = 1$ for all $i, i' \in V$. According to (24), if $u_t(i, i') = 1$ then $q_t(i, i') > 0$ and, consequently, $S_{i,t} = S_{i',t}$. As a result, with $\mathcal{C}_t$ denoting the set of connected components of H_t, for all clusters of time series $\gamma \in \mathcal{C}_t$, $S_{i,t} = S_{i',t}$ for all $i, i' \in \gamma$. In other words, the auxiliary variables $\boldsymbol{u}_t$ induce a partition of the time series, $\mathcal{C}_t$, such that, for each cluster $\gamma \in \mathcal{C}_t$, either all time series or no time series in γ are affected by a changepoint at time t. Moreover, according to (24), if $S_{i,t} = S_{i',t}$ then the conditional probability that $u_t(i, i') = 1$ increases with $\lambda_{i,i'} > 0$, so that clusters induced by $\boldsymbol{u}_t$ will tend to be clusters on the graph induced by the edge weight parameters.

5.1.2 MCMC algorithm

To generate realisations from the posterior distribution of the changepoints, we consider a "cluster updating" MCMC algorithm that samples from the extended joint posterior density (25). By inducing clusters of time series determined by the edge weight parameters for each time point t, the auxiliary variables $\boldsymbol{u}$ provide a means to efficiently explore the state space of $(\boldsymbol{k}, \boldsymbol{\tau})$.

The parameter δ is a tuning parameter for the cluster updating MCMC algorithm. The size of clusters will tend to increase with δ; in particular, if $\delta = 0$ then, for all t, each cluster corresponds to a unique time series index, even if the edge weights of the dependence graph are large, so that the "cluster updating" MCMC algorithm reduces to the "single site updating" algorithm. Typically δ is fixed (Higdon, 1998) to control the probabilities in (23) and therefore the expected size of clusters. However, we propose to treat $\delta \geqslant 0$ as an unknown parameter with prior distribution $\pi(\delta)$, so that expected sizes of cluster may vary in the sample; specifically, we assume that $\delta = 0$ with probability $0 \leqslant \delta_0 \leqslant 1$ and otherwise δ is drawn from $\text{Beta}(\delta_1, \delta_2)$ for $\delta_1, \delta_2 > 0$.

For the cluster updating MCMC algorithm, at each iteration of the algorithm, with $(\boldsymbol{k}, \boldsymbol{\tau}, \boldsymbol{u}, \delta)$ denoting the latest particle of the sample chain, one of the following moves is proposed.

Birth/death move

Conditional on the auxiliary variables, the birth/death move proposes the birth or death of a cluster of synchronous changepoints. Sample t' uniformly from $\{1, \ldots, T\}$. A cluster γ of time series indices is randomly chosen from $\mathcal{C}_{t'}$, the set of clusters of time series induced by the auxiliary variables $\boldsymbol{u}_{t'}$. Then, leaving the auxiliary variables unchanged, propose changepoint parameters $(\boldsymbol{k}', \boldsymbol{\tau}')$ such that, for all $i = 1, \ldots, N$ and $t = 1, \ldots, T$,

$$S'_{i,t} = \begin{cases} 1 - S_{i,t} & \text{if } i \in \gamma \text{ and } t = t' \\ S_{i,t} & \text{otherwise,} \end{cases} \tag{27}$$

where $\boldsymbol{S}$ and $\boldsymbol{S}'$ are binary matrix representations of $(\boldsymbol{k}, \boldsymbol{\tau})$ and $(\boldsymbol{k}', \boldsymbol{\tau}')$ according to (8), respectively.

Shift move

The shift move proposes to shift the position of a cluster of synchronous changepoints. First, a time unit t is uniformly chosen from $\{t : \sum_{i=1}^{N} S_{i,t} > 0\}$. Let $\mathcal{C}_t^* \subseteq \mathcal{C}_t$ denote the set of clusters of time series indices γ induced by $\boldsymbol{u}_t$ such that, for all $i \in \gamma$, $S_{i,t} = 1$. A cluster γ is uniformly chosen from $\mathcal{C}_t^*$. For all $i \in \gamma$, let j_i be the index of the changepoint with position t for the ith time series, that is $\tau_{i,j_i} = t$. Then, sample uniformly t' from $\bigcap_{i\in\gamma}\{\tau_{i,j_i-1}+1,\ldots,\tau_{i,j_i+1}-1\}$ and propose changepoint parameters $(\boldsymbol{k},\boldsymbol{\tau}')$ that are identical to $(\boldsymbol{k},\boldsymbol{\tau})$ but with $\tau'_{i,j_i} = t'$, for all $i \in \gamma$.

In parallel, it is required to propose auxiliary variables $\boldsymbol{u}'$ which are adapted to $(\boldsymbol{k},\boldsymbol{\tau}')$. The updated auxiliary variables differ from $\boldsymbol{u}$ as follows. For all $i, i' \in \gamma$, $u'_{t'}(i,i') = u_t(i,i')$ and $u'_t(i,i') = u_{t'}(i,i')$; for all $i \in \gamma$ and $i' \notin \gamma$ such that $t' \notin \boldsymbol{\tau}_{i'}$, $u'_{t'}(i,i') = 0$; and, ensuring reversibility of the move, for all $i \in \gamma$ and $i' \notin \gamma$ such that $t \notin \boldsymbol{\tau}_{i'}$, $u'_t(i,i')$ is sampled conditionally on $(\boldsymbol{k},\boldsymbol{\tau}')$ according to the Bernoulli $(1-\exp\{-\delta\lambda_{i,i'}(1-|S_{i,t}-S_{i',t}|)\})$ target distribution implied by (23).

Update of auxiliary variables

Changepoints are left unchanged, δ is sampled from its prior distribution, and auxiliary variables are sampled from the full conditional distribution given in (23), thereby proposing an updated clustering of time series indices for all t.

5.2 Sampler for asynchronous dependent changepoints

According to the changepoint model (22) introduced in Section 4.3, changepoints do not need to occur at the same time to be related. Consequently, to explore the changepoint parameter space it will be required to propose the birth, death or shift of clusters of asynchronous changepoints. This section extends the MCMC algorithm from Section 5.1 to sample from the posterior distribution of changepoints when changepoint parameters are *a priori* distributed according to $\pi(\boldsymbol{k},\boldsymbol{\tau}|p,\boldsymbol{\lambda},\boldsymbol{w})$ from (22).

Recall that under the asynchronous model, changepoints $(\boldsymbol{k},\boldsymbol{\tau})$ are deterministically specified by latent changepoints $(\boldsymbol{k},\tilde{\boldsymbol{\tau}})$ and unknown lags $\boldsymbol{d}$ according to (15). Therefore, a sample from $(\boldsymbol{k},\boldsymbol{\tau})$ can be obtained from a sample from $(\boldsymbol{k},\tilde{\boldsymbol{\tau}},\boldsymbol{d})$. Next, we propose a sampler from the joint posterior distribution of $(\boldsymbol{k},\tilde{\boldsymbol{\tau}},\boldsymbol{d})$, updated from the prior density (21) by the observed data $\boldsymbol{x}$, thereby providing a means to obtain a sample from the posterior distribution of $(\boldsymbol{k},\boldsymbol{\tau})$.

As in Section 5.1, the parameter space is augmented with auxiliary variables $\boldsymbol{u} = (\boldsymbol{u}_1,\ldots,\boldsymbol{u}_T)$ to facilitate the exploration of the state space of the parameters of interest $(\boldsymbol{k},\tilde{\boldsymbol{\tau}},\boldsymbol{d})$. Conditionally on latent changepoints $(\boldsymbol{k},\tilde{\boldsymbol{\tau}})$ and independently of the lags and the data, for all t and $i < i'$, $u_t(i,i')$ is assumed to be distributed according to (23), such that $u_t(i,i') \sim \text{Bernoulli}\left(1-\exp\{-\delta\lambda_{i,i'}(1-|\tilde{S}_{i,t}-\tilde{S}_{i',t}|)\}\right)$, where $\tilde{\boldsymbol{S}}$ is the binary matrix representation of $(\boldsymbol{k},\tilde{\boldsymbol{\tau}})$ according to (8). As described in Section 5.1.1, it follows that the auxiliary variables $\boldsymbol{u}_t$ induce a partition of the time series, $\mathcal{C}_t$, such that, for each cluster $\gamma \in \mathcal{C}_t$, either all time series or no time series in γ are affected by a latent changepoint at time t.

The joint posterior density of the augmented parameters $(\boldsymbol{k},\tilde{\boldsymbol{\tau}},\boldsymbol{d},\boldsymbol{u})$ is

$$\pi(\boldsymbol{k},\tilde{\boldsymbol{\tau}},\boldsymbol{d},\boldsymbol{u}|p,\boldsymbol{\lambda},\boldsymbol{w},\boldsymbol{x},\delta) = \pi(\boldsymbol{u}|\boldsymbol{\lambda},\delta,\boldsymbol{k},\tilde{\boldsymbol{\tau}})\pi(\boldsymbol{k},\tilde{\boldsymbol{\tau}},\boldsymbol{d}|p,\boldsymbol{\lambda},\boldsymbol{w},\boldsymbol{x}). \tag{28}$$

To sample from the posterior distribution of $(\boldsymbol{k},\tilde{\boldsymbol{\tau}},\boldsymbol{d},\boldsymbol{u})$, or $(\boldsymbol{k},\tilde{\boldsymbol{\tau}},\boldsymbol{d},\boldsymbol{u},\boldsymbol{w})$ if upper bounds for the lags are *a priori* unknown, the MCMC algorithm discussed in Section 5.1 is extended as follows: the birth/death and shift moves are adapted to pairs of latent changepoints and lags; and additional moves are introduced for updating the lags. For the lags, note that according to (16), for all $i = 1, \dots, N$ and $j = 1, \dots, k_i$, to maintain monotonicity in the changepoints, the lag associated to the jth changepoint of the ith time series must satisfy

$$d_{i,j} \in \mathcal{D}_j(w_i, k_i, \tilde{\boldsymbol{\tau}}_i) = \{\ell \in \mathbb{N};\, d_- \leqslant \ell \leqslant d^+\}, \tag{29}$$

where $d_- = \max(0, d_{i,j-1} + \tau_{i,j-i} - \tau_{i,j} + 1)$ and $d^+ = \min(w_i, d_{i,j+1} + \tau_{i,j+1} - \tau_{i,j} - 1)$; and the full conditional probability distribution of $d_{i,j}$ is such that, for all $d_{i,j} \in \mathcal{D}_j(w_i, k_i, \tilde{\boldsymbol{\tau}}_i)$,

$$\pi\left(d_{i,j} \,|\, \boldsymbol{d}_{-(i,j)}, k_i, \tilde{\boldsymbol{\tau}}_i, w_i, \boldsymbol{x}_i\right) \propto \mathcal{L}_i\left(\tilde{\tau}_{i,j-1} + d_{i,j-1}, \tilde{\tau}_{i,j} + d_{i,j}\right) \mathcal{L}_i\left(\tilde{\tau}_{i,j} + d_{i,j}, \tilde{\tau}_{i,j+1} + d_{i,j+1}\right) \tag{30}$$

where $\boldsymbol{d}_{-(i,j)} = \{d_{i',j'};\, (i',j') \neq (i,j)\}$ and $\mathcal{L}_i$ is defined in (5).

For the extended cluster updating MCMC algorithm, at each iteration of the algorithm, with $(\boldsymbol{k}, \tilde{\boldsymbol{\tau}}, \boldsymbol{d}, \boldsymbol{u}, \boldsymbol{w})$ denoting the latest particle of the sample chain, one of the following moves is proposed.

Extended birth/death move

The extended birth/death move proposes the birth or death of a cluster of asynchronous changepoints. First, conditionally on the auxiliary variables, latent changepoints $(\boldsymbol{k}', \tilde{\boldsymbol{\tau}}')$ are proposed according to the birth/death move detailed in Section 5.1: for all time series $i \in \gamma \subseteq \{1, \dots, N\}$, the birth or death of latent changepoint with position t is proposed. Then, updated lags are proposed conditional on $(\boldsymbol{k}', \tilde{\boldsymbol{\tau}}')$: If the birth of changepoints is proposed, then, for all time series $i \in \gamma$, there is j_i such that $\tilde{\tau}'_{i,j_i} = t$, and the lags $\boldsymbol{d}'_i = (d_{i,1}, \dots, d_{i,j_i-1}, d'_{i,j_i}, d_{i,j_i} \dots, d_{i,k_i})$ are proposed for the ith time series, where d'_{i,j_i} is sampled from the full conditional distribution (30); otherwise, if the death of changepoints is proposed, then, for all $i \in \gamma$, there is j_i such that $\tilde{\tau}_{i,j_i} = t$, and the lags $\boldsymbol{d}'_i = (d_{i,1}, \dots, d_{i,j_i-1}, d_{i,j_i+1} \dots, d_{i,k_i})$ are proposed for the ith time series.

Extended shift move

The extended shift move proposes to shift the positions of a cluster of asynchronous changepoints. First, latent changepoints $(\boldsymbol{k}', \tilde{\boldsymbol{\tau}}')$ and auxiliary variables $\boldsymbol{u}'$ are proposed according to the shift move discussed in Section 5.1: for all time series with index $i \in \gamma$, the position t' is proposed for latent changepoint with position t. Then, for all $i \in \gamma$, letting j_i denote the index such that $\tilde{\tau}'_{i,j_i} = t'$, propose d'_{i,j_i} from the full conditional distribution (30).

Update of auxiliary variables

δ is sampled from its prior distribution and, conditional on latent changepoints $(\boldsymbol{k}, \tilde{\boldsymbol{\tau}})$, auxiliary variables are sampled from their full conditional distribution (23).

Update of lags

A pair (i, j) is uniformly chosen from $\{(i, j); i = 1, \dots, N \text{ and } j = 1, \dots, k_i\}$, and the lag $d_{i,j}$ is sampled from the full conditional distribution given in (30).

Update of upper bounds for lags

If the maximal lags $\boldsymbol{w}$ are *a priori* unknown, for a randomly chosen time series with index i it is proposed to update w_i to $w_i' = w_i + \sigma$ with probability $1/2$, and to update w_i to $w_i' = |w_i - \sigma|$ otherwise, where σ is drawn from Geometric(ρ) for some $0 < \rho < 1$. Proposing to update w_i requires proposing updated lags $\boldsymbol{d}_i' = (d_{i,1}', \dots, d_{i,k_i}') \in \mathcal{D}(w_i', k_i, \tilde{\boldsymbol{\tau}}_i)$ for the ith time series. For $j = 1, \dots, k_i$, given w_i' and $(d_{i,1}', \dots, d_{i,j-1}', d_{i,j+1} \dots d_{i,k_i})$, the lag $d_{i,j}'$ is sampled from the full conditional distribution (30).

6 Estimating changepoint parameters

To summarise the posterior distribution of changepoint parameters for multiple time series, for each time series i, following Green (1995), one may consider the posterior marginal distribution of the number of changepoints k_i, and the posterior distribution of the changepoint positions $\boldsymbol{\tau}_i$ conditional on k_i. However, in practice, for each time series i, it may be necessary to report a point estimate $(\hat{k}_i, \hat{\boldsymbol{\tau}}_i)$ for the changepoint parameters $(k_i, \boldsymbol{\tau}_i)$. Following normative Bayesian theory, to define an optimal Bayes estimate for changepoints, we propose a loss function that evaluates the quality of estimated changepoints. When assessing the cost associated with the estimate $(\hat{k}_i, \hat{\boldsymbol{\tau}}_i)$ of $(k_i, \boldsymbol{\tau}_i)$, both the number and the positions of changepoints must be taken into account. To address this challenge, we use matchings in graphs, as defined in Definition 1 and Definition 2, to define a loss function L for changepoint estimates in Definition 3.

Definition 1. (Maximum matching in a graph). *Let $B = (V, E)$ be a graph where V is a vertex set and $E \subseteq V \times V$ is an edge set. A matching M in B is a subset of E such that no two edges in M share a common vertex. A maximum matching in B is a matching that is not a subset of a larger matching in B.*

Definition 2. (Minimum weight maximum matching in a graph). *Let $B = (V, E)$ be a graph with weights $w_{i,j} \geqslant 0$ for all $(i, j) \in E$. A minimum weight maximum matching in B is a maximum matching in B for which the sum of weights of the edges is minimised.*

When B is a weighted bipartite graph, the Kuhn–Munkres algorithm, also known as the Hungarian algorithm, (Bondy and Murty, 1976) finds a minimum weight maximum matching in B; the time complexity of the algorithm is $\mathcal{O}(\text{card}(E)\text{card}(V) + \text{card}(V)^2 \log\log \text{card}(V))$, where card$(V)$ and card(E) denote the cardinality of the vertex set and the cardinality of the edge set of B, respectively.

Definition 3. (Loss function L for changepoint estimates). *Let $\gamma \geqslant 0$. For all $k_i, \hat{k}_i \geqslant 0$, $\boldsymbol{\tau}_i = (\tau_{i,1}, \dots, \tau_{i,k_i}) \in \mathcal{T}_{k_i}$ and $\hat{\boldsymbol{\tau}}_i = (\hat{\tau}_{i,1}, \dots, \hat{\tau}_{i,\hat{k}_i}) \in \mathcal{T}_{\hat{k}_i}$ (3), let B_i be the weighted complete bipartite graph with vertex sets $V_i = \{0, \dots, k_i\}$ and $\hat{V}_i = \{0, \dots, \hat{k}_i\}$, and weights*

$$w_{i,j,j'} = \min\{\gamma, |\tau_{i,j} - \hat{\tau}_{i,j'}|\} \tag{31}$$

for all $j \in V_i$ *and* $j' \in \hat{V}_i$*. Given a minimum weight maximum matching* M_i *in* B_i*, for all* $j \in V_i$ *and* $j' \in \hat{V}_i$*, let* $m_{i,j,j'} = 1$ *if* j *and* j' *are matched, that is* $(j, j') \in M_i$*, and* $m_{i,j,j'} = 0$ *otherwise. Then, define the loss to be*

$$L\left[(\hat{k}_i, \hat{\boldsymbol{\tau}}_{\boldsymbol{i}}), (k_i, \boldsymbol{\tau}_i)\right] = \gamma|\hat{k}_i - k_i| + \frac{1}{2}\sum_{j \in V_i}\sum_{j' \in \hat{V}_i} m_{i,j,j'} w_{i,j,j'}. \tag{32}$$

Consider the complete bipartite graph B_i with independent vertex sets $V_i = \{0, \dots, k_i\}$, $\hat{V}_i = \{0, \dots, \hat{k}_i\}$ and weights (31). A minimum weight maximum matching M_i in B_i gives a matching of the elements of $\hat{\boldsymbol{\tau}}_i$ and $\boldsymbol{\tau}_i$ that minimises the sum of distances (31) between matched changepoints. Given M_i, according to the loss function (32), the cost associated with the estimate $(\hat{k}_i, \hat{\boldsymbol{\tau}}_i)$ of $(k_i, \boldsymbol{\tau}_i)$ is then obtained by adding the cost γ for each unmatched changepoint and the total distance between matched changepoints. Note that according to (31), the cost of matching two changepoint positions, that are separated by more than γ time units, is equal to the cost of an unmatched changepoint, namely γ. Therefore, the loss function L takes into account both the number and the positions of changepoints, and the cost γ is chosen to be the maximum acceptable distance between a changepoint position and its estimated position. The optimal Bayes estimate $(\hat{k}_i, \hat{\boldsymbol{\tau}}_i)$ is the changepoint parameters that minimise the expected posterior loss with respect to the posterior marginal distribution of the changepoints $(k_i, \boldsymbol{\tau}_i)$.

Given an approximate sample from the posterior distribution (Section 5), an approximate Bayes estimate $(\hat{k}_i, \hat{\boldsymbol{\tau}}_i)$ for each series can be identified numerically by finding within the sample the changepoint parameters that minimise the estimated posterior expected loss.

Appendix D in the supplementary material (Hallgren et al., 2023) presents a simulation study that demonstrates the model introduced in Section 4 and the MCMC sampling strategy discussed in Section 5, using the loss function introduced in this section. In particular, various graphs and changepoint parameters are considered to illustrate the flexibility of the proposed model, and the convergence of the sampler is demonstrated under a wide range of settings.

7 Red team detection in network authentication data from LANL

This section presents results of an analysis of the LANL network authentication data presented in Section 2 that demonstrates the utility of the graphical changepoint model proposed in Section 4.

7.1 Presence of a red team

The occurrence of a red team exercise during the first month of the data collection provides surrogate intruder behaviour in the authentication data (Kent, 2015). In particular, 103 user IDs are known to have been used by the red team. We show the

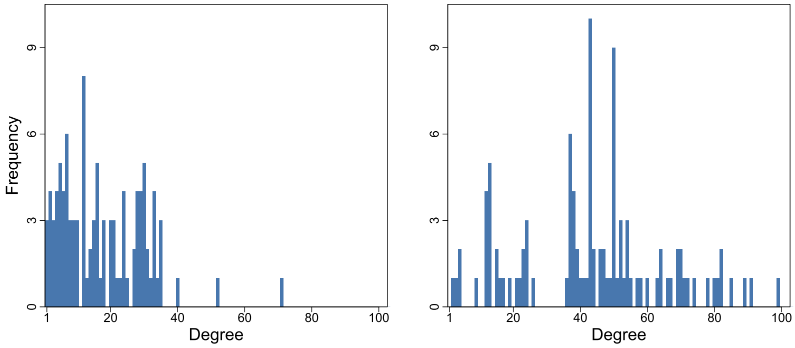

Figure 3: Degree distribution of users in the LANL network represented by the graph G'. Left panel: counts for legitimate users. Right panel: counts for red team users.

graphical model for dependent changepoints can combine evidence from multiple users which are linked in the network, to detect chains of quasi-synchronous weak signals for changes in the authentication activity of red team users, whilst limiting the number of false alerts.

For our demonstration purposes, it suffices to examine a subset of the full LANL network of users, which is represented by the graph $G = (V, E)$ defined in Section 2.2. Let R denote the set of red team users and let $B \subseteq \{u \in V \setminus R : \exists r \in R \text{ s.t. } (u, r) \in E\}$ denote card(R) randomly selected users that are not labelled as red team users in the data but are linked to red team users on the network. The focus is on the network corresponding to the subgraph G' induced in G by the set of users $V' = R \cup B$. Figure 3 shows the degree distribution of the 206 users in G'. Red team users tend to have a greater degree in G' than legitimate users; to traverse the network towards high value targets, intruders tend to take control of users that are highly linked on the network.

7.2 Changepoint modelling

Recall from Section 2 that, for each user $i \in V'$, the data $x_{i,0}, \ldots, x_{i,T}$ consist of hourly counts of network logons per source computer for the first month of data collection as defined in (2), which are now assumed to follow the model specified in (6) for multinomial data. Different graphical changepoint priors are considered to demonstrate the benefits of encoding prior beliefs about cyber-attacks. To encode prior belief that signals for changes resulting from an attack are likely to occur at similar times across users that are linked in the network G', the graphical changepoint prior specified in (22) is considered with an identical edge weight parameter $\lambda > 0$, as defined in Section 4.1.2, for all pairs of time series corresponding to users that are linked in G'. Moreover, for comparison purposes, the graphical changepoint prior (22) is also considered assuming the complete graph defined in Section 4.2.3 such that $\lambda_{i,j} = \lambda > 0$ for all pairs of users $(i, j) \in V' \times V'$. With $\lambda = 0$, the two graphical changepoint priors of interest correspond to the standard changepoint model assuming independence of changepoints across time series (10).

For comparison purposes and to illustrate the flexibility of the proposed model, a collection of changepoint prior parameters are considered: $\bar{p} \in \{-30, -50, -70\}$ and $\lambda = \lambda_s |\bar{p}|/n$ with $\lambda_s \in \{0, 0.5, 0.6, 0.7, 0.8\}$, where n denotes the average node degree in the graph. Moreover, different assumptions for the upper bounds $\boldsymbol{w}$ for the lags are compared: the *zero window* assumption with $w_i = 0$ for all i, implying signals for attacks are assumed to be synchronous across users; and, the *variable window* assumption with $w_i \sim \text{Geometric}(0.9)$ for all i, admitting signals for attacks may be asynchronous across users.

For each simulation, a sample of size 1 000 000 was obtained from the posterior distribution of the changepoints via the MCMC algorithm proposed in Section 5.2, with a burn-in of 300 000 iterations; the Bayes estimate for changepoints corresponding to the loss function (32) with $\gamma = 48$ was then derived from the sample.

7.3 Results

For each user in the network, each estimated changepoint represents a piece of evidence for possible malicious behaviour that might require further investigation by cyber analysts. Identifying inferred changepoints for time series corresponding to legitimate users $i \in B$ as false alerts, it is meaningful to compare models in terms of the estimated number of changepoints per time series $i \in V'$,

$$\bar{k} = \frac{1}{\text{card}(V')} \sum_{i \in V'} k_i, \tag{33}$$

and the proportion of estimated changepoints that impact redteam users,

$$k^{R/V'} = \frac{\sum_{i \in R} k_i}{\sum_{i \in V'} k_i}. \tag{34}$$

Moreover, since cyber-attacks tend to be identified through clusters of behavioural changes across machines that are linked on the network, it is of interest to prioritise for investigation the estimated changepoints that belong to clusters of quasi-synchronous changepoints on the network. Given some time window $\varpi \geqslant 0$, let the weight of $\tau_{i,j}$ be

$$c_{i,j}^{\varpi} = \frac{n_{i,j}^{\varpi} + 1}{n_i + 1}, \tag{35}$$

where

$$n_{i,j}^{\varpi} = \text{card}(\{(i, i') \in E : \exists j' \in \{1, \ldots, k_{i'}\} \text{ s.t. } |\tau_{i,j} - \tau_{i',j'}| \leqslant \varpi\}) \tag{36}$$

denote the number of users linked to user i in G' that are impacted by a changepoint within ϖ hours of $\tau_{i,j}$, and where $n_i = \text{card}(\{(i, i') \in E\})$ is the degree of node i in G', such that $(n_i + 1)^{-1} \leqslant c_{i,j} \leqslant 1$. The larger the weight $c_{i,j}$, the more connected $\tau_{i,j}$ to other changepoints across the network. To take into account both the number of changepoints and their connectedness, for each user i, changepoint estimates are also compared via the sum of weights

$$m_i \equiv m_i^{\varpi} = \sum_{j=1}^{k_i} c_{i,j}^{\varpi}, \tag{37}$$

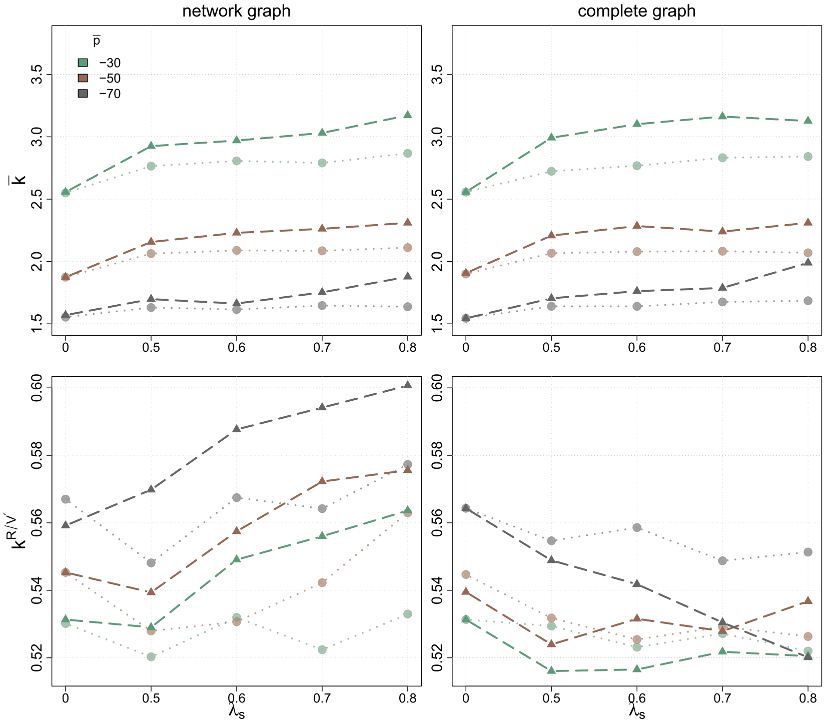

Figure 4: Average number of changepoints per time series $\bar{k}$ (top row), and proportion of changepoints impacting redteam users $k^{R/V'}$ (bottom row), assuming the network graph (left column) and the complete graph (right column), for different assumptions for the upper bounds for the lags - *zero window* (circles) and *variable window* (triangles), and for a collection of prior parameters $\bar{p}$ (identified by distinct colours) and λ_s.

such that $0 \leqslant m_i \leqslant k_i$. Note that for each user i, m_i increases with both the number of changepoints and their weights. Let

$$\bar{m} = \frac{1}{\text{card}(V')} \sum_{i \in V'} m_i, \quad m^{R/V'} = \frac{\sum_{i \in R} m_i}{\sum_{i \in V'} m_i} \tag{38}$$

be the average sum of changepoint weights per user and the proportion of changepoint weights associated to redteam users, respectively.

For each choice of graph and changepoint prior parameters, Figure 4 displays the estimated values of $\bar{k}$ and $k^{R/V'}$, and Figure 5 displays the estimated values of $\bar{m}$ and $m^{R/V'}$ assuming $\varpi = 48$. As $\bar{p}$ increases, weaker evidence is required to infer changepoints, and therefore, for each graph, $\bar{k}$ and $\bar{m}$ increase. As λ increases, the estimates for $\bar{k}$ and $\bar{m}$ tend to increase for each graph, but the estimates for $k^{R/V'}$ and $m^{R/V'}$ tend to increase only when assuming the network graph. This follows because the graphical changepoint model assuming the network graph successfully encodes prior knowledge that cyber attacks tend to correspond to coordinated activity across multiple

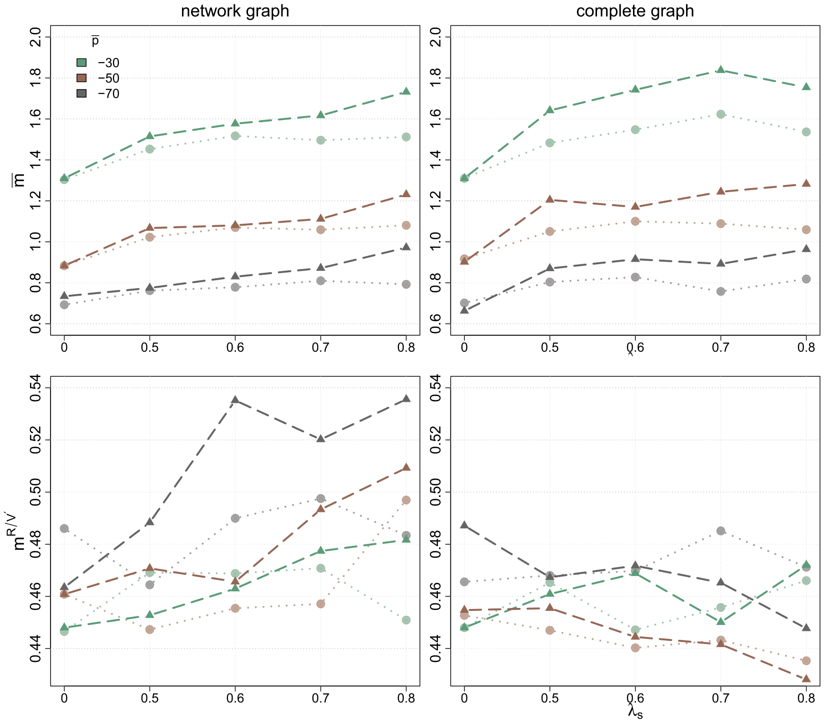

Figure 5: Estimated $\bar{m}$ and $m^{R/V'}$ (38), assuming the network graph (left column) and the complete graph (right column), for different assumptions for the upper bounds for the lags - *zero window* (circles) and *variable window* (triangles), and for a collection of prior parameters $\bar{p}$ (identified by distinct colours) and λ_s.

users linked by network connectivity, and consequently, as λ increases, it detects weaker signals for behavioural changes that correspond to red team activity, whilst crucially limiting the number of false alerts. For the proposed model with the complete graph, all time series are connected, and therefore, as λ increases, weaker signals for changes are detected for red team activity but also for legitimate activity, which would impede fast identification of the attack.

Moreover, results show the benefits of the model extension which relaxes the assumption that signals for attacks are synchronous across users. As λ increases, when assuming the network graph, the increase of the estimates for $k^{R/V'}$ and $m^{R/V'}$ tend to be greater for the *variable window* scenario than for the *zero window* scenario. In contrast with the *zero window* scenario, the *variable window* scenario admits attacks may result in quasi-synchronous behavioural changes across the network, and consequently clusters of nearby but not necessarily synchronous weak signals for changepoints are detected across red team users.

The results show that, in comparison with the standard model for independent changepoints across time series, the proposed graphical changepoint model provides a flexible tool for cyber-analysts to incorporate expert knowledge in changepoint analysis for network monitoring, thereby facilitating network intrusion detection.

8 Discussion

This article considers a setting with N time series (1) subject to changepoints, where it is desirable to encode in the changepoint prior, by means of a graph $G = (V, E)$ on N nodes corresponding to each of the time series, that pairs of time series $(i, i') \in E$ are *a priori* more likely to be impacted by simultaneous changepoints. This setting is adapted to the application in cyber-security where each node in V corresponds to a time series representing the authentication activity of a network user, and an edge $(i, i') \in E$ indicates that it is believed *a priori* that attackers may switch credentials between user i and user i' at any time of the data collection period, so that users i and i' are *a priori* more likely to be impacted by quasi-simultaneous behavioural changes.

However, for some applications, it might be restrictive to assume that prior beliefs on which time series are likely to be impacted by simultaneous changepoints do not vary over time. For example, consider the following application in cyber security. Using system log data, it can be of interest to monitor the process activity of computers, which may be subject to changes when attackers perform malicious activity such as the installation or the execution of malware. Moreover, attackers will typically need to communicate with compromised computers to simultaneously execute malicious commands on these computers. As a result, the process activity of computers i and i' are more likely to be subject to simultaneous changes when some source computer simultaneously communicates to both i and i'. For such a setting, it would be more suitable to specify a time series of graphs $\{G_t = (V, E_t) \mid E_t \subseteq V \times V, t \geqslant 0\}$, such that pairs of time series $(i, i') \in E_t$ are *a priori* more likely to be impacted by simultaneous changepoints at time t. Each node in V would correspond to a time series representing the process activity of a computer in the network, and an edge $(i, i') \in E_t$ would indicate that communication events occurred at time t from some source computer to both i and i', so that computers i and i' are *a priori* more likely to be impacted by simultaneous behavioural changes at time t. For networks where many computers may leave or enter during the data collection period, a further model extension could consider relaxing the assumption that V is fixed, specifying a time series of graphs $\{G_t = (V_t, E_t) \mid E_t \subseteq V_t \times V_t, V_t \subset \mathbb{N}, t \geqslant 0\}$ such that V_t is the node set of computers active in the network at time t. With the introduction of time-dependent edge weight parameters $\boldsymbol{\lambda} = (\lambda_{i,i',t})$ such that $\lambda_{i,i',t} > 0$ if and only if $(i, i') \in E_t$, these model extensions would present no theoretical complication, with a straightforward adaption of the graphical changepoint prior and the proposed sampling strategy.

Supplementary Material

Supplementary Material for “Changepoint detection on a graph of time series” (DOI: 10.1214/23-BA1365SUPP; .pdf). The *Python* code and the data are available at https://github.com/karl-hallgren/cp_on_graph_of_timeseries/

References

Bardwell, L. and Fearnhead, P. (2017). "Bayesian detection of abnormal segments in multiple time series." *Bayesian Analysis*, 12(1): 193–218. MR3597572. doi: https://doi.org/10.1214/16-BA998. 650

Bardwell, L., Fearnhead, P., Eckley, I. A., Smith, S., and Spott, M. (2019). "Most recent changepoint detection in panel data." *Technometrics*, 61(1): 88–98. MR3933661. doi: https://doi.org/10.1080/00401706.2018.1438926. 650

Besag, J. and Green, P. J. (1993). "Spatial statistics and Bayesian computation." *Journal of the Royal Statistical Society. Series B (Methodological)*, 55(1): 25–37. MR1210422. 651, 662

Bolton, A. D. and Heard, N. A. (2018). "Malware family discovery using reversible jump MCMC sampling of regimes." *Journal of the American Statistical Association*, 113(524): 1490–1502. MR3902224. doi: https://doi.org/10.1080/01621459.2018.1423984. 650, 655

Bondy, J. A. and Murty, U. S. R. (1976). *Graph Theory with Applications*. New York: Elsevier. MR0411988. 651, 667

Carlin, B. P., Gelfand, A. E., and Smith, A. F. M. (1992). "Hierarchical Bayesian analysis of changepoint problems." *Journal of the Royal Statistical Society: Series C (Applied Statistics)*, 41(2): 389–405. MR1278223. 655

Chen, H. (2019a). "Change-point detection for multivariate and non-Euclidean data with local dependency." *arXiv:1903.01598*. 651, 654

Chen, H. (2019b). "Sequential change-point detection based on nearest neighbors." *The Annals of Statistics*, 47(3): 1381–1407. MR3911116. doi: https://doi.org/10.1214/18-AOS1718. 651

Chen, H. and Zhang, N. (2015). "Graph-based change-point detection." *The Annals of Statistics*, 43(1): 139–176. MR3285603. doi: https://doi.org/10.1214/14-AOS1269. 651, 654

Chu, L. and Chen, H. (2019). "Asymptotic distribution-free change-point detection for multivariate and non-Euclidean data." *The Annals of Statistics*, 47(1): 382–414. MR3910545. doi: https://doi.org/10.1214/18-AOS1691. 651

Denison, D., Holmes, C., Bani, M., and Smith, A. (2002). *Bayesian Methods for Nonlinear Classification and Regression*. Chichester: Wiley Series in Probability and Statistics. MR1962778. 655, 662, 663

Fearnhead, P. (2006). "Exact and efficient Bayesian inference for multiple changepoint." *Statistics and Computing*, 16(2): 203–213. MR2227396. doi: https://doi.org/10.1007/s11222-006-8450-8. 650, 654, 655, 657

Fisch, A. T. M., Eckley, I. A., and Fearnhead, P. (2022). "Subset Multivariate Collective and Point Anomaly Detection." *Journal of Computational and Graphical Statistics*, 31(2): 574–585. MR4425087. doi: https://doi.org/10.1080/10618600.2021.1987257. 650, 660

Gelman, A., Carlin, J. B., Stern, H. S., and Rubin, D. B. (2004). *Bayesian Data Analysis*. Chapman and Hall/CRC, 2nd ed. edition. MR2027492. 662

Green, P. J. (1995). "Reversible jump Markov Chain Monte Carlo computation and Bayesian model determination." *Biometrika*, 82(4): 711–732. MR1380810. doi: `https://doi.org/10.1093/biomet/82.4.711`. 650, 651, 662, 667

Grundy, T. J., Killick, R., and Mihaylov, G. (2020). "High-dimensional change-point detection via a geometrically inspired mapping." *Statistics and Computing*, 30(99): 1155–1166. MR4108696. doi: `https://doi.org/10.1007/s11222-020-09940-y`. 650

Hallgren, K. L., Heard, N. A., and Turcotte, M. J. (2023). "Supplementary material – Changepoint detection on a graph of time series." *Bayesian Analysis*. doi: `https://doi.org/10.1214/23-BA1365SUPP`. 652, 655, 656, 658, 662, 663, 668

Higdon, D. M. (1998). "Auxiliary variable methods for Markov Chain Monte Carlo with applications." *Journal of the American Statistical Association*, 93(442): 585–595. 651, 662, 663, 664

Jeng, X. J., Cai, T. T., and Li, H. (2012). "Simultaneous discovery of rare and common segment variants." *Biometrika*, 100(1): 157–172. MR3034330. doi: `https://doi.org/10.1093/biomet/ass059`. 650

Johnson, T., Elashoff, R., and Harkema, S. (2003). "A Bayesian change-point analysis of electromyographic data: Detecting muscle activation patterns and associated applications." *Biostatistics*, 4(1): 143–64. 655

Kent, A. D. (2015). "Cybersecurity Data Sources for Dynamic Network Research." In *Dynamic Networks in Cybersecurity*. Imperial College Press, London. 651, 652, 668

Lauritzen, S. L. (1996). *Graphical Models*. Oxford University Press, Oxford. MR1419991. 650, 657

Li, F. and Zhang, N. R. (2010). "Bayesian variable selection in structured high-dimensional covariate spaces with applications in genomics." *Journal of the American Statistical Association*, 105(491): 1202–1214. MR2752615. doi: `https://doi.org/10.1198/jasa.2010.tm08177`. 650

Metelli, S. and Heard, N. (2019). "On Bayesian new edge prediction and anomaly detection in computer networks." *The Annals of Applied Statistics*, 13(4): 2586–2610. MR4037442. doi: `https://doi.org/10.1214/19-aoas1286`. 654

Passino, F. S., Turcotte, M. J. M., and Heard, N. A. (2021). "Graph link prediction in computer networks using Poisson matrix factorisation." *The Annals of Applied Statistics*, to appear. MR4455882. doi: `https://doi.org/10.1214/21-aoas1540`. 654

Punskaya, E., Andrieu, C., Doucet, A., and Fitzgerald, W. (2002). "Bayesian curve fitting using MCMC with applications to signal segmentation." *IEEE Transactions on Signal Processing*, 50(3): 747–758. 655

Sexton, J. O., Storlie, C., and Neil, J. (2015). "Attack chain detection." *Statistical Anal-*

ysis and Data Mining, 8(5): 353–363. MR3418417. doi: https://doi.org/10.1002/sam.11296. 650, 652, 653

Swendsen, R. H. and Wang, J.-S. (1987). "Nonuniversal critical dynamics in Monte Carlo simulations." *Phys. Rev. Lett.*, 58: 86–88. 651

Turcotte, M. (2014). "Anomaly Detection in Dynamic Networks." *PhD thesis, Imperial College London*. 655

Wang, T. and Samworth, R. J. (2018). "High dimensional change point estimation via sparse projection." *Journal of the Royal Statistical Society: Series B (Statistical Methodology)*, 80(1): 57–83. MR3744712. doi: https://doi.org/10.1111/rssb.12243. 650

Acknowledgments

The authors thank Niall Adams for stimulating discussions about this work.