Taylor & Francis
Taylor & Francis Group

Journal of Computational and Graphical Statistics

ISSN: (Print) (Online) Journal homepage: https://www.tandfonline.com/loi/ucgs20

# Nonparametric Anomaly Detection on Time Series of Graphs

Dorcas Ofori-Boateng , Yulia R. Gel & Ivor Cribben

**To cite this article:** Dorcas Ofori-Boateng , Yulia R. Gel & Ivor Cribben (2021): Nonparametric Anomaly Detection on Time Series of Graphs, Journal of Computational and Graphical Statistics, DOI: 10.1080/10618600.2020.1844214

**To link to this article:** https://doi.org/10.1080/10618600.2020.1844214

JOURNAL OF COMPUTATIONAL AND GRAPHICAL STATISTICS
2020, VOL. 00, NO. 0, 1–12
https://doi.org/10.1080/10618600.2020.1844214



# Nonparametric Anomaly Detection on Time Series of Graphs

Dorcas Ofori-Boateng[a], Yulia R.Gel[a], and Ivor Cribben[b]

[a]Department of Mathematical Sciences, University of Texas at Dallas, TX; [b]Department of Accounting and Business Analytics, Alberta School of Business, Alberta, Canada

**ABSTRACT**

Identifying change points and/or anomalies in dynamic network structures has become increasingly popular across various domains, from neuroscience to telecommunication to finance. One particular objective of anomaly detection from a neuroscience perspective is the reconstruction of the dynamic manner of brain region interactions. However, most statistical methods for detecting anomalies have the following unrealistic limitation for brain studies and beyond: that is, network snapshots at different time points are assumed to be independent. To circumvent this limitation, we propose a distribution-free framework for anomaly detection in dynamic networks. First, we present each network snapshot of the data as a linear object and find its respective univariate characterization via local and global network topological summaries. Second, we adopt a change point detection method for (weakly) dependent time series based on efficient scores, and enhance the finite sample properties of change point method by approximating the asymptotic distribution of the test statistic using the sieve bootstrap. We apply our method to simulated and to real data, particularly, two functional magnetic resonance imaging (fMRI) datasets and the Enron communication graph. We find that our new method delivers impressively accurate and realistic results in terms of identifying locations of true change points compared to the results reported by competing approaches. The new method promises to offer a deeper insight into the large-scale characterizations and functional dynamics of the brain and, more generally, into the intrinsic structure of complex dynamic networks. Supplemental materials for this article are available online.

**ARTICLE HISTORY**
Received February 2019
Revised May 2020

**KEYWORDS**
Change point; Dynamic networks; Multivariate time series; Sieve bootstrap

## 1. Introduction

Identifying and analyzing change points and/or anomalies (which we use interchangeably) has become an increasingly active area of research in network sciences (Host-Madsen and Zhang 2018; Messer, Albert, and Schneider 2018). For example, in financial trading, a regime shift in the network of transactions is frequently linked to various (often upcoming) insolvencies such as bankruptcies, defaults, and recessions while a change in the network topology of cryptocurrency transactions may suggest a potential money laundering scheme (e.g., Elliott, Golub, and Jackson 2014; Vandermarliere et al. 2015; Abay et al. 2019; Akcora et al. 2020). Similar to other biological networks (Wig, Schlaggar, and Petersen 2011), the idea of studying the brain as a dynamic functional network is helpful in understanding the complex network organization of the brain and can lead to profound clinical breakthroughs (Bassett and Bullmore 2006; Rubinov and Sporns 2010).

Most currently available statistical methods for anomaly detection in dynamic networks have the following limitation: network snapshots at different time points are assumed to be independent (e.g., Akoglu and Faloutsos 2010; Peel and Clauset 2015; Harshaw et al. 2016). This assumption appears to be unrealistic in many applications. For example, cryptocurrency entities and their interactions in a network of transactions evolve in time but obviously daily snapshots of the same network cannot be assumed to be independent. This dependence or autocorrelation effect is well documented in the time series literature since the 1960s (Zellner 1962; Wolff, Gastwirth, and Rubin 1967). However, in many cases, this effect is often overlooked in practice, which leads to unreliable and false conclusions. Neglecting the dependence among the network snapshots at different time points leads to inflated false-positive rates of change points or anomalies, especially for small and moderate sample sizes.

To overcome this limitation, we propose a new distribution-free framework, named *network evolution detection method* (NEDM), for anomaly detection in dynamic networks and evolution network structures of high dimensional time series. The setup for the proposed methodology entails the following: With each network snapshot as a graph object, we find its unique univariate characterization, for example, mean degree, clustering coefficient, and clique number. As a result, a series of possibly very high-dimensional network snapshots is transformed into a time series of scalars. Given the temporal dependence of network snapshots, it is infeasible to assume that the resulting time series of linear characteristics is independent. Next, we adopt a change point detection method (Gombay 2008) for (weakly) dependent time series that is based on efficient scores. To enhance the finite sample properties

**CONTACT** Ivor Cribben *cribben@ualberta.ca* Department of Accounting and Business Analytics, 4-30G Business Building, 11203 Saskatchewan Drive NW, Edmonton, AB T6G 2R6, Canada.

Supplementary materials for this article are available online. Please go to *www.tandfonline.com/r/JCGS*.

of the detected change points, we approximate the asymptotic distribution of the test statistic with a sieve bootstrap procedure (Kreiss 1988, 1992; Bühlmann et al. 1997). We derive asymptotic properties for the NEDM statistic, validate its performance in respect to competing anomaly detection methods via synthetic and real data experiments. We illustrate the utility of the NEDM by applying it to two input data types: two multivariate functional magnetic resonance imaging (fMRI) time series datasets (Cribben et al. 2012; Cribben, Wager, and Lindquist 2013; Cribben and Yu 2017) and a dynamic network of the Enron e-mail communication (Diesner, Frantz, and Carley 2005; Priebe et al. 2005; Park, Priebe, and Youssef 2012). For the sake of brevity, we defer the analysis of one fMRI dataset to the Supplementary Material (appendix).

In application, we find that the NEDM has the potential to unveil the time-varying cognitive states of both controls and subjects with neuropsychiatric diseases such as Alzheimer's, dementia, autism and schizophrenia in order to develop new understandings of these diseases. By applying the NEDM, we can consider whole brain dynamics, which promises to offer deeper insight into the large scale characterizations of functional architecture of the whole brain.

To sum up, the proposed NEDM has the following unique and significant attributes:

1. It is, to the best of our knowledge, the first paper to consider estimating change points in any graph summary statistic for the time-evolving network structure in a multivariate time series context.
2. It can consider thousands of time series and, in particular, the case where $P$, the number of time series is much greater than $T$ ($P >> T$), the number of time points.
3. Unlike existing methods it is not limited by assuming that network snapshots at different time points are independent.
4. It enhances the finite sample properties of the change point method by approximating the asymptotic distribution of the test statistic using the sieve bootstrap.
5. Although it is inspired by and developed for brain connectivity studies, it pertains to a general setting and can also be used in a variety of situations where one wishes to study the evolution of a high dimensional network over time.

The remainder of the article is organized as follows. We introduce the NEDM in Section 3. A simulation study that examines the finite sample performance of our method via sieve bootstrapping (Kreiss 1988, 1992; Bühlmann et al. 1997) is also covered in Section 3. In Section 4, we describe the properties for building synthetic data and then provide background information on the fMRI and Enron dynamic network datasets in Section 5. The performance of the proposed algorithm, on both the synthetic and real world data, is detailed in Section 6. A conclusion and future work is discussed in Section 7. Finally, proofs, supplementary material and processes are deferred to the Supplementary Material.

## 2. Related Work

There exists a vast body of studies on dynamic network models across various disciplines (see, Barabási and Albert 1999 and references therein). One method based on the minimum description length (MDL) principle and compression techniques (Sun et al. 2007) flattens the adjacency matrices into binary strings and uses compression cost to derive data specific features. Another procedure (DELTACON) proposed by Koutra, Vogelstein, and Faloutsos (2013) relied on the similarity measures between a pair of equal node networks. However, anomalous points reported by this procedure tend to suffer from limited interpretability due to their lack of statistical quantifiers (such as critical numbers or $p$-values).

The first comprehensive treatment of high dimensional time series factor models with multiple change points in their second-order structure has been put forward by Barigozzi, Cho, and Fryzlewicz (2018). To detect changes in the covariance matrix of a multivariate time series, Aue et al. (2009) introduced a method using a nonparametric CUSUM type test, and Dette and Wied (2016) proposed a test where the dimension of the data is fixed. Furthermore, Cribben et al. (2012) and Cribben, Wager, and Lindquist (2013) put forward a method for detecting changes in the precision matrices (or undirected graph) from a multivariate time series. In turn, Cribben and Yu (2017) introduced a graph-based multiple change point method for changes in the community network structure between high-dimensional time series, called *network change point detection*, that uses an eigen-space based statistic for testing the community structures changes in stochastic block model sequences. In addition, Barnett and Onnela (2016) developed a method for detecting change points in correlation networks that, unlike previous change point detection methods designed for time series data, requires minimal distributional assumptions.

## 3. Methodology

In this section, we describe the proposed contribution which tracks structural changes within the network structure of datasets. We use the terms change point detection and anomaly detection as well as the terms graphs and networks interchangeably. Table 1 in the supplementary materials details the notation for the rest of the article.

### 3.1. Input Data: Graphs and Multivariate Time Series

Our proposed methodology is applicable to two types of input data: data that are originally in a form of a graph, and multivariate time series that are used to construct a graph, based on a certain similarity measure, for example, correlation. Since one of our primary motivating applications is multivariate fMRI time series, below we describe in detail how networks can be constructed from such datasets.

*Networks from a multivariate time series.* In many applications, such as neuroscience and finance, input data are multivariate time series, and the first step consists of constructing a graph structure based on a user-selected (dis)similarity measure. That is, suppose $\boldsymbol{D}$ is a $T \times P$ multivariate time series, where $T$ and $P$ are the number of time points and the number of time series, respectively. From the multivariate $\boldsymbol{D}$, we take a $q$-row sample, $\boldsymbol{F} \in \mathbb{R}^{q \times P}$, in a sequential "one-in, one-out" manner. This mode of data segmentation is known as the

*overlapping/sliding window* technique (Keogh et al. 2001) and is ideal for maintaining the time dependency structure within $\boldsymbol{D}$, while taking as many samples as possible from the data matrix, $\boldsymbol{D}$. Because each $q$-row sample contains information from previous or successive samples, this segmentation procedure has the advantage of capturing all (and any) network structure disturbance. Note that, depending on the window length $q$, some data information may not be adequately captured because not every row in the data matrix has an equal number of appearances in all the $\boldsymbol{F}$ folds created. For instance, the first row ($\boldsymbol{D}_{[1,\ ]}$) and the last row ($\boldsymbol{D}_{[T,\ ]}$) are least likely to be included in the collection of all $\boldsymbol{F}$s. However, since we hypothesize that the changes points occur within the data and not at the endpoints (at the beginning and end of the time series) our detection procedure and results do not suffer.

Armed with $\boldsymbol{F}$, we compute a correlation matrix $\boldsymbol{R} = (\rho_{rs}) \in \mathbb{R}^{P\times P}$ by quantifying the linear association between the $P$ vertices in $\boldsymbol{F}$. With $\boldsymbol{R}$ and a pre-defined threshold $\omega$[1], we define the finite graph-associated adjacency matrix $\boldsymbol{A} = (a_{rs}) \in \mathbb{R}^{P\times P}$ using

$$a_{rs} = \begin{cases} 1 & \text{if } \rho_{rs} \geqslant \omega \\ 0 & \text{otherwise.} \end{cases}$$

Given each adjacency matrix $\boldsymbol{A}$, we then construct a graph object $G_t = (V_t, E_t)$. (Here, to identify a minimum feasible number $P$, as a potential rule of thumb, a practitioner can employ approaches discussed, for example, by Ozkaya et al. 2017).

*Input data as a graph structure.* Alternatively, the original input data can take form of a graph $G_t$ observed at time point $t, t = 1, 2, \ldots$. Such examples include communication networks (see, for instance, the Enron study in Sections 5 and 6), power grid networks, and the emerging blockchain technology. Indeed, one of the salient blockchain features is that all transactions are permanently recorded on distributed ledgers and publicly available. As a result, a blockchain graph $G_t$ can be constructed directly on each transaction, bypassing application of correlation and other similarity measures.

Finally, armed with the sequence of the graph objects $G_t$, we then calculate various global and local graph summary statistics (Newman 2003; Barabási and Pósfai 2016). In particular, from each graph, we estimate the following graph summaries ($Y_t$): *average clustering coefficient* (ACC), *average path length* (APL), *maximum node betweenness centrality* (maxBETW), *clique number* (CLQNUM), *mean degree* (MD) and *minimum local clustering coefficient* (minLCC).

### 3.2. Detection Procedure and the Sieve Bootstrap

Following the dimension reduction and data structure simplification in Section 3.1, our next task is to identify the time(s) at which the regime shift(s) occur in the series of lower dimensional embeddings. In our case, we are interested in testing the following hypotheses:

$H_0$ : There is no change in the underlying structure of $\{Y_t\}_{t=1}^n$
$H_a$ : At an unknown time $t \in \{1, \ldots, n\}$ a change occurs in $\{Y_t\}_{t=1}^n$.

We assume that the appropriate model to fit to $\{Y_t\}_{t=1}^n$ is a strictly stationary and purely nondeterministic autoregressive [AR($p$)] model with Gaussian independent and identically distributed (iid) white noise $\varepsilon_t$. The assumption of Gaussianity for $\varepsilon_t$ can be relaxed and substituted by the appropriate moment conditions (Gombay 2008). (We run experiments on time series with non-Gaussian innovations and find that while the proposed change point detection is applicable to a non-Gaussian case, performance largely depends on deviations from the normality assumption and sample size. For instance, under the null hypothesis of no change, a nominal $\alpha$-level of 0.05, and an AR(1) process with $\varphi$ of 0.5, it takes approximately 100 observations to achieve an approximate size of the test of 0.05 for a case of $t$-distribution with 9 degrees of freedom; in turn, it requires about 200 observations from a uniform distribution to achieve a similar empirical size of the test.) In turn, an approximation of time series via AR($p$) models, including the case of $p \to \infty$, is widely used in theory and methodology of time series analysis (for overview see, for instance, Pourahmadi 2001; Shumway and Stoffer 2017, and references therein).

*Remark.* Note that under the considered problem of change point detection on networks, the network topological summary statistic $\{Y_t\}_{t=1}^n$ is estimated from the data. There currently exist no theoretical results on asymptotic properties of network statistics, especially in a conjunction with dynamic networks; that is, besides invoking a central limit theorem for network mean degree, nothing can be formally said on a linear process representation of $\{Y_t\}_{t=1}^n$ or its distributional properties. As a result, our approach is approximation-based and data-driven; that is, while we cannot provide theoretical bounds on the linear process approximation of $\{Y_t\}_{t=1}^n$, we can validate performance of the proposed NEDM against known ground truth change points.

In particular, we define $Y_t$ as

$$Y_t - \mu = \sum_{k=1}^{p} \varphi_k (Y_{t-k} - \mu) + \varepsilon_t \ , \quad t \geqslant p+1 \qquad (1.1)$$

with $\mathbf{E}[Y_t] = \mu$ and let $\boldsymbol{\xi} = (\mu, \sigma^2, \varphi_1, \ldots, \varphi_p)^T$. Given Equation (1.1), we formally test the null hypothesis of no change

$$\begin{aligned} H_0 &: \boldsymbol{\xi} = \boldsymbol{\xi}_0, \forall\ t \in \{1, \ldots, n\} \quad \text{against} \\ H_a &: \boldsymbol{\xi} = \boldsymbol{\xi}_0,\ \forall\ 1 \leqslant t \leqslant \tau - 1,\ at\ \tau = [\rho n],\ \rho \in (0,1)\ ; \\ &\quad\ \boldsymbol{\xi} = \boldsymbol{\xi}_a,\ \forall\ t \geqslant \tau. \end{aligned}$$

Most studies on change point detection calculate the pre-regime switch and post-regime switch parameter values of all possible change points $\tau \in (1, n)$, and then either use the strength of their differences to determine a regime switch or use these parameter values in the likelihood function (Picard 1985; Inoue 2001). However, we use a detection algorithm that involves a one-time parameter estimation and allows us to

[1] In the experimental analysis we select parameters $q$ and $\omega$ based on previous neuroscience studies. Alternatively, $q$ and $\omega$ can be selected via cross-validation.

test for change in the individual elements of $\boldsymbol{\xi}$. In line with allowing for one-sided tests and for flexibility, we adopt the change point statistic of Gombay (2008) which uses the efficient score vector $\nabla_{\boldsymbol{\xi}}\ell_k(Y_1,\ldots,Y_n;\boldsymbol{\xi}) = \nabla_{\boldsymbol{\xi}}\ell_k(\boldsymbol{\xi})$ with $\ell_k$ as the log-likelihood function on $\{Y_t\}_{t=1}^n$. For $1 \leqslant r \leqslant p+2$ denote $\hat{\mu}$, $\hat{\sigma}^2$, $\hat{\varphi}_1$, $\ldots$, $\hat{\varphi}_p$ as the simultaneous solutions of the $p+2$ equations $\partial/\partial\xi_r\ell_n(\boldsymbol{\xi}) = 0$, and let

$$\hat{\mathbf{B}}(u) = n^{-1/2}I^{-1/2}(\hat{\xi}_n)\begin{pmatrix} \frac{\partial}{\partial\mu}\ell_{[nu]}(\hat{\xi}_n) \\ \frac{\partial}{\partial\sigma^2}\ell_{[nu]}(\hat{\xi}_n) \\ \nabla_{\varphi}\ell_{[nu]}(\hat{\xi}_n) \end{pmatrix} \tag{1.2}$$

be a Gaussian process (a partial sums process approximation for the structure of $\nabla_{\boldsymbol{\xi}}\ell_k(\boldsymbol{\xi})$).

If $\xi_r$, for $1 \leqslant r \leqslant p+2$, changes at $\tau = [\rho n]$ with $0 < \rho < 1$, then the estimator $\hat{\tau}$ is defined by:

$$\hat{\tau}_n = \min_n\left\{\frac{\partial}{\partial\xi_r}\ell_k(\hat{\boldsymbol{\xi}}_n) = \max_{1<m\leqslant n}\frac{\partial}{\partial\xi_r}\ell_m(\hat{\boldsymbol{\xi}}_n)\right\}, \tag{1.3}$$

and

$$\hat{\tau}_n = \min_n\left\{\frac{\partial}{\partial\xi_r}\ell_k(\hat{\boldsymbol{\xi}}_n) = \min_{1<m\leqslant n}\frac{\partial}{\partial\xi_r}\ell_m(\hat{\boldsymbol{\xi}}_n)\right\}, \tag{1.4}$$

for one-sided tests (both left and right, respectively); and for two-sided tests

$$\hat{\tau}_n = \min_n\left\{\left|\frac{\partial}{\partial\xi_r}\ell_k(\hat{\boldsymbol{\xi}}_n)\right| = \max_{1<m\leqslant n}\left|\frac{\partial}{\partial\xi_r}\ell_m(\hat{\boldsymbol{\xi}}_n)\right|\right\}. \tag{1.5}$$

Proof for the consistency of $\hat{\tau}$ is provided in Gombay (2008). The asymptotic independence of the components of $\hat{B}(u)$ allows us define the change point test statistic for each $\xi_r \in \boldsymbol{\xi}$ (see Gombay 2008 for the related discussion). Hence, we reject the null hypothesis, for *one-sided tests* along the sequence $\{Y_t\}_{t=1}^n$, if

$$\sup_{0\leqslant u\leqslant 1}\hat{B}^{(r)}(u) \geqslant C_1(\alpha), \tag{1.6}$$

where $C_1(\alpha)$ is calculated from

$$\left\{x : P\Big(\sup_{0\leqslant u\leqslant 1}B^{(1)}(u) \geqslant x\Big) = e^{-2x^2} = \alpha\right\}.$$

If we are interested in *two-sided* hypothesis testing, a change in $\xi_r$ (along the sequence $\{Y_t\}_{t=1}^n$) is acknowledged whenever

$$\sup_{0\leqslant u\leqslant 1}|\hat{B}^{(r)}(u)| \geqslant C_2(\alpha), \tag{1.7}$$

such that $C_2(\alpha)$ is calculated from

$$\left\{x : P\Big(\sup_{0\leqslant u\leqslant 1}|B^{(1)}(u)| > x\Big) = \sum_{k\neq 0}(-1)^{k+1}e^{-2(kx)^2} = \alpha\right\}.$$

Convergence of the test statistic $\hat{B}^{(r)}(u)$ to its asymptotic distribution can be relatively slow. The Type I error estimates tend to be conservative with low power of the test. Under the premise that $\{Y_t\}_{t=1}^n$ follows an AR($p$) model with Gaussian iid white noise, we propose to adopt a sieve bootstrap procedure for constructing the distribution of the change point test statistic $\hat{B}^{(j)}(u)$ for finite samples. The idea of such a bootstrap for time-dependent data—originally named AR($\infty$) bootstrap—goes back to the results of Kreiss (1988, 1992). The approach was later further investigated by Bühlmann et al. (1997) who coined the term *sieve* for this bootstrap method. The procedure is outlined in Algorithm 1, and its theoretical properties are stated in Theorem 1.

To compare the finite sample performance of the asymptotic distribution to the sieve bootstrap method, we simulate data (with a total of 5000 Monte Carlo iterations) from an AR(1) model in Equation (1.1) and evaluate the following: the size and power of the test (for precise details on the simulation procedure used, please see the Supplementary Materials). The choice of the model coefficient ($\varphi$) depends on ensuring the assumption of weak stationarity for the simulated AR(1) series. Indeed note that as $\varphi$ approaches 1, the time series gets closer to a random walk process, that is, we approach to a boundary case of violating the assumptions of weak stationarity, outlined for the change point statistic based on the efficient scores (Theorem 1 of Gombay 2008). As a result, the performance of the change point detection method deteriorates.

Figure 1 depicts the size of the test for the simulated data. It shows that the asymptotic distribution behaves more conservatively compared to our sieve bootstrap distribution. Additionally, we find that the size is more conservative for the asymptotic distribution when $\varphi$ is closer to +1 (i.e., the size is worse when $\varphi = 0.9$ compared to $\varphi = 0.5$). However, for negative $\varphi$ cases,

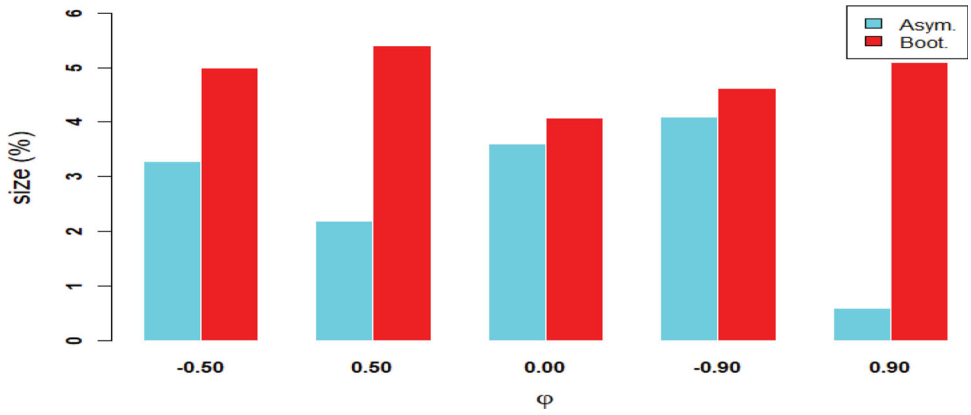


**Figure 1.** The size of the test plotted against $\varphi$ for the asymptotic distribution and sieve bootstrap procedure when testing for a change in the mean of an *AR*(1) model using $T = 100$.

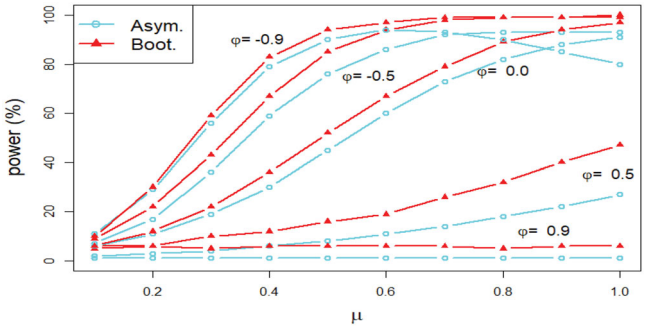


**Figure 2.** The power of the test plotted against $\varphi$ for the asymptotic distribution and sieve bootstrap procedure when testing for a change in the mean of an AR(1) model with ($T = 100$).

we find that the size of the test under the asymptotic distribution approaches the declared 5% level of significance as $\varphi$ approaches −1. Apart from the fact that the sieve bootstrap's size is closer to the nominal rate than the asymptotic distribution, we find that as $\varphi$ approaches +1 the size of its test steadily hovers around the declared 5% level of significance. This supports our assertion that the asymptotic distribution of the change point test statistic has relatively conservative Type I error, and validates our sieve bootstrap procedure for finite samples.

The results for power is displayed in Figure 2 and from this we find that as $\mu$ increases, the power of the test also increases under both the asymptotic and the bootstrap distributions. In addition, as $\varphi$ approaches +1, there is a drop in the power values for both the asymptotic and sieve bootstrap (with larger power values reported by our bootstrap procedure). Next to this, we find that as $\varphi$ approaches −1 the power of the test improves for both the bootstrap and the asymptotic distribution; with the bootstrap again outperforming the asymptotic distribution.[2]

*Theorem 1* (Sieve bootstrap). Let $Y_t$ be an autoregressive process as defined in equation (1.1) [with $\mathbf{y} = \{Y_t\}_{t=1}^n$], $\mathbf{E}|\varepsilon_t|^6 < \infty$, $\sum_{j=0}^{\infty} j|\varphi_j| < \infty$, and $p(n) = o((n/\log(n))^{1/4})$. With $\hat{\mathbf{B}}^*(u)$ as the bootstrap estimate for the $p+2$-dimensional Gaussian process $\mathbf{B}(u)$ , and as $n \to \infty$, we have

$$\sup_{0\le u\le 1} |\mathbb{P}^*[n^{1/2}(\hat{B}^{(r)^*}(u) - B^{(r)^*}(u)) \le \mathbf{y}] - \mathbb{P}[n^{1/2}(\hat{B}^{(r)}(u) - B^{(r)}(u)) \le \mathbf{y}]| = o_p(1).$$

Proof of Theorem 1 is in the Supplementary Materials.

Suppose $\tilde{\varepsilon}_t = \hat{\varepsilon}_t - \bar{\varepsilon}_n$, with $\bar{\varepsilon}_n = 1/n\sum_{t=1}^n \hat{\varepsilon}_t$, then the nonparametric sieve bootstrap estimate $\hat{B}^{*(r)}(u)$ can be replaced with a hybrid parametric bootstrap estimate $\hat{B}^{\diamond(r)}(u)$ by generating a Gaussian sample $\{\tilde{\varepsilon}_t^{\diamond}\}_{t=1}^n \sim N(0, s_n^2)$, with $s_n^2 = \sum_{t=1}^n (\tilde{\varepsilon}_t - \bar{\tilde{\varepsilon}}_t)^2/(n-1)$ as the sample variance of $\{\tilde{\varepsilon}_t\}_{t=1}^n$. This yields a finite sample performance similar to that of $\{\tilde{\varepsilon}_t^*\}_{t=1}^n$, with reduced computing time.

[2]From Gombay (2008), a specific value of $\varphi$ (and not $|\varphi|$ or $\varphi^2$) generates a specific change point statistic ($\hat{\mathbf{B}}(u)$) and the change point estimator ($\hat{\tau}$).

**Algorithm 1:** The nonparametric sieve bootstrap procedure for the change point statistic under an $AR(p)$ process

**Input** : Time series $\{Y_t\}_{t=1}^n$; Change point test statistic $T_n(.)$; Level of significance $\alpha$; Number of bootstrap replications $\Gamma$.

For $\{X_t\}_{t=1}^n = \{Y_t - \hat{\mu}\}_{t=1}^n$, fit an $AR(p)$ process and extract the coefficients $\{\hat{\boldsymbol{\varphi}}_r\}_{r=1}^p$.

Calculate $T_n(X_1, X_2..., X_n)$.

Obtain $\tilde{\varepsilon}_t$ from the residuals $\hat{\varepsilon}_t = X_t - \sum_{k=1}^p \hat{\varphi}_k X_{t-k}$. (**Note**: $\tilde{\varepsilon}_t = \hat{\varepsilon}_t - \bar{\varepsilon}_n$, where $\bar{\varepsilon}_n = \frac{1}{n}\sum_{t=1}^n \hat{\varepsilon}_t$).

**for** $i \leftarrow 1 : \Gamma$ **do**

  Sample, with replacement, $\{\tilde{\varepsilon}_t^*\}_{t=1}^n$ from $\{\tilde{\varepsilon}_t\}_{t=1}^n$.

  Simulate $\{X_t^*\}_{t=1}^n$, with innovation $\{\tilde{\varepsilon}_t^*\}_{t=1}^n$ and coefficients $\{\hat{\boldsymbol{\varphi}}_r\}_{r=1}^p$.

  Calculate the change point test statistic $T_n^*(X_1^*, X_2^*, ..., X_n^*)$.

  Define $Z_i$ for each Monte-Carlo iteration as:

  $$Z_i = \begin{cases} 1 & |T_n^*| \geqslant |T_n| \\ 0 & \text{otherwise.} \end{cases}$$

**end**

The bootstrap $p$-value for testing $H_0$ is given by $\sum_{i=1}^{\Gamma} \frac{Z_i}{\Gamma}$.

*Corollary 1.* If $Y_t$ satisfies Equation (1.1), then under $H_0$

$$\sup_{0\le u\le 1} |\mathbb{P}^*[n^{1/2}(\hat{B}^{\diamond(r)}(u) - B^{\diamond(r)}(u)) \le \mathbf{y}] - \mathbb{P}[n^{1/2}(\hat{B}^{(r)}(u) - B^{(r)}(u)) \le \mathbf{y}]| = o_p(1). \tag{1.9}$$

Proof of Corollary 1 is in the Supplementary Materials.

## 4. Simulations

To validate our NEDM, we compare it to the BOM (Barnett and Onnela 2016) using simulated data. In particular, we provide a sensitivity analysis for the NEDM under various $q$, $\omega$ and ROI choices. We do not compare the NEDM to the KVFM (Koutra, Vogelstein, and Faloutsos 2013) because it does not provide a quantifier for the statistical significance (critical numbers or $p$-values) for the change point detected. Such a quantifier is necessary for deducing the power and size of the test for the methods under study. The simulation study covers two scenarios: a no-change point and a one-change point. As evaluation metrics, we use the true detection rates (i.e., power of the test) and the false alarm levels under a pre-defined significance level (i.e., size of the test). A total of 100 Monte Carlo simulations is carried out in each scenario under various window lengths $q = \{5, 10, 15, 20, 25, 30\}$, threshold parameters $\omega = \{0.05, 0.1, 0.15\}$, two different time series lengths ($T = 200, 300$) and for two graph node sizes ($P = 5, 10$). In addition, we provide similar analyses with $T = 300, P = 50$ in Appendix B of the Supplementary Materials.

The first simulation illustrates the baseline (no change point) scenario using a vector autoregression (VAR; Zellner 1962; Hamilton 1995) model; the VAR model is a generalization of

the univariate AR process with more than one time-evolving component. Given the $(p\times 1)$ vector of time series variables $\boldsymbol{F}_t = (f_{1t}, f_{2t}, \ldots, f_{pt})^T$, the $w$-lag vector autoregressive (VAR(w)) process is defined as

$$\boldsymbol{F}_t = \boldsymbol{a} + \boldsymbol{\Pi}_1 \boldsymbol{F}_{t-1} + \boldsymbol{\Pi}_2 \boldsymbol{F}_{t-2} + \cdots + \boldsymbol{\Pi}_w \boldsymbol{F}_{t-w} + \boldsymbol{\epsilon}_t, \quad t = 1, \ldots, T. \tag{1.10}$$

where $\boldsymbol{\Pi}_i$ is an $(p \times p)$ coefficient matrix and $\boldsymbol{\epsilon}_t$ is an $(p \times 1)$ unobservable mean white noise vector process with time invariant covariance matrix $\boldsymbol{\Sigma}$ (Zivot and Wang 2007). The VAR model is used to reconstruct the linear inter-dependency element prevalent among multivariate time series applications such as fMRI data.

The second simulation, which depicts a one-change point scenario, is created by concatenating two data streams from distinct multivariate Gaussian distributions: $(\boldsymbol{D}_1, \boldsymbol{D}_2)$ where $\boldsymbol{D_1} \sim N(\boldsymbol{\mu} = \boldsymbol{0}, \boldsymbol{\Sigma_1} = (\Sigma_{1ij}))$ and $\boldsymbol{D_2} \sim N(\boldsymbol{\mu} = \boldsymbol{0}, \boldsymbol{\Sigma}_2 = (\Sigma_{2ij}))$ with

$$\Sigma_{1ij} = \begin{cases} 1 & \text{if } i = j \\ 0 & \text{otherwise} \end{cases} \qquad \Sigma_{2ij} = \begin{cases} 1 & \text{if } i = j \\ 0.9 & \text{otherwise.} \end{cases}$$

The results for the simulations are presented in Figures 3 and 4, and a discussion of the results is provided in Section 6.

## 5. Experimental Data

Next, we demonstrate the NEDM's application to two input data types: multivariate fMRI time series and a portion of the Enron e-mails network. In the fMRI case study, we compare the performance of the NEDM against two other techniques for anomaly (change point) detection in the multivariate setting: the KVFM (Koutra, Vogelstein, and Faloutsos 2013) and the BOM (Barnett and Onnela 2016). With the Enron networks, we only implement the NEDM with the APL network summary (because from our experiments the APL-based analysis produced the best outcome), and compare the change points we find with various events that characterized the timeframe of the Enron scandal. In addition, we compare the performance of the NEDM to results obtained by the KVFM. We do not compare our results with the BOM because the BOM is only applicable to multivariate time series data.

### 5.1. Case Study: Anxiety fMRI data

The data was taken from an anxiety-inducing experiment (Cribben et al. 2012; Cribben, Wager, and Lindquist 2013). The task was a variant of a well-studied laboratory paradigm for eliciting social threat, in which participants must give a speech under evaluative pressure. The design was an off-on-off design, with an anxiety-provoking speech preparation task

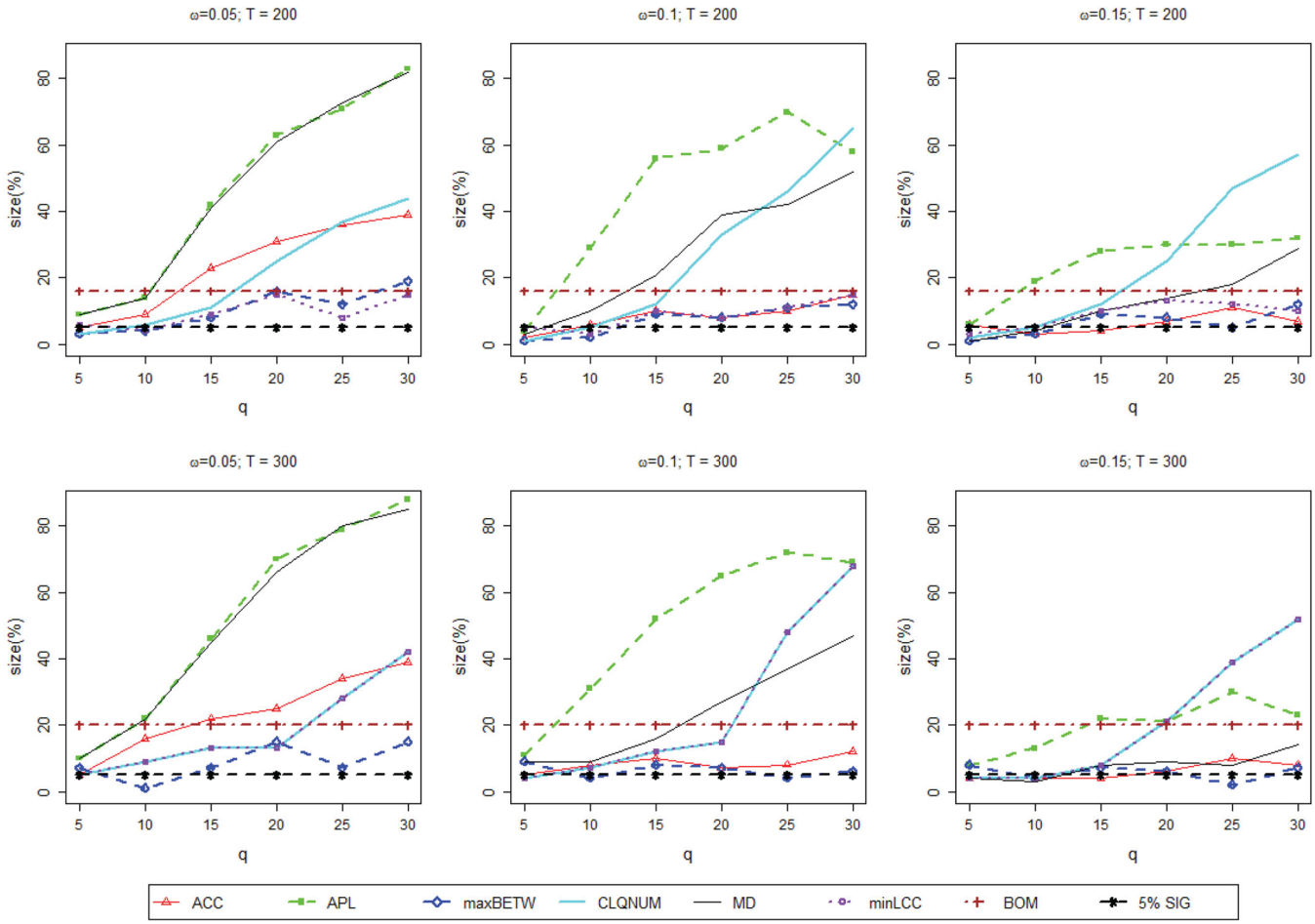


**Figure 3.** The size of the test for the no-change (change in mean) scenario with $T = \{200, 300\}$, $\omega = \{0.05, 0.1, 0.15\}$, $q = \{5, 10, 15, 20, 25, 30\}$ and $P = 5$.

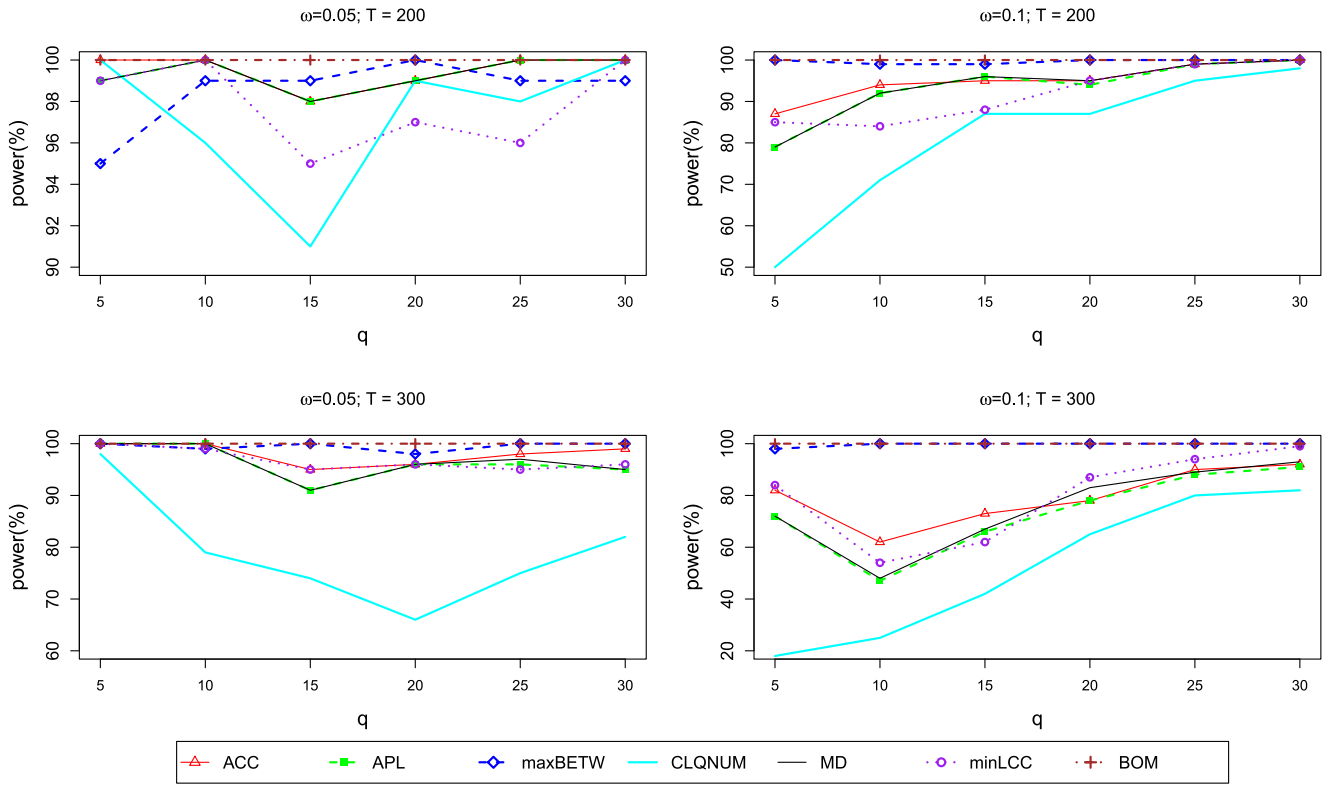


**Figure 4.** The power of the test for the one-change (change in variance) scenario with $\omega = \{0.05, 0.1\}$, $q = \{5, 10, 15, 20, 25, 30\}$, $T = \{200, 300\}$, $\tau = \{88, 156\}$ and $P = 10$.

occurring between lower anxiety resting periods. Participants were informed that they were to be given 2 min to prepare a 7 min speech, and that the topic would be revealed to them during scanning. They were told that after the scanning session they would deliver the speech to a panel of expert judges, though there was "a small chance" they would be randomly selected not to give the speech. After the start of fMRI acquisition, participants viewed a fixation cross for 2 min (resting baseline). At the end of this period, participants viewed an instruction slide for 15 s that described the speech topic, which was "why you are a good friend." The slide instructed participants to be sure to prepare enough for the entire 7 min period. After 2 min of silent preparation, another instruction screen appeared (a relief instruction, 15 s duration) that informed participants that they would not have to give the speech. An additional 2 min period of resting baseline completed the functional run. Data were acquired and preprocessed as described in previous work (Wager et al. 2009). During the course of the experiment a series of 215 images were acquired ($TR = 2$ s). In order to create ROIs, time series were averaged across the entire region. The data consists of 4 ROIs and heart rate for $n = 23$ subjects. The regions in the data were chosen because they showed a significant relationship to heart rate in an independent dataset. The temporal resolution of the heart rate was 1 s compared to 2 s for the fMRI data. Hence, the heart rate was down-sampled by taking every other measurement.

### 5.2. Case Study: Enron e-mail Networks

The Enron emails dataset is a benchmark dataset applied in numerous instances of anomaly detection (e.g., Peel and Clauset 2015; Park, Priebe, and Youssef 2012; Priebe et al. 2005; Diesner, Frantz, and Carley 2005). More information on this dataset can be found online (*http://www.cs.cmu.edu/~enron/*). We used the cleaned version of the employee-to-employee e-mail (sent and received) network over the period November 1998 to July 2001. We initialize each employee as a single node and aggregate the data by month. This implies that if there is at least one e-mail between two employees within the month under study, an edge is connected to the respective nodes. Figure 5 displays the cumulative nature of the Enron network between November 1998 to July 2001, and the state of the network after two specific month/year periods. In total, we obtain 33 networks with 102 nodes in each network.

## 6. Results

The NEDM uses the following summary statistics: Average clustering coefficient (ACC), *average path length* (APL), *maximum node betweenness centrality* (maxBETW), *clique number* (CLQNUM), *mean degree* (MD) and *minimum local clustering coefficient* (minLCC). As we already mentioned in the previous section, we compare our method to Barnett's method (BOM) in the simulation study. However, we include results from KVFM for the Enron e-mail dataset.

### 6.1. Simulation Study

Figure 3 presents the results for the no-change point scenario. We find that as $q$ increases the size of the test reported by all network summaries under the NEDM increase, leading to more liberal results. Moreover, we find that as $\omega$ increases, the size

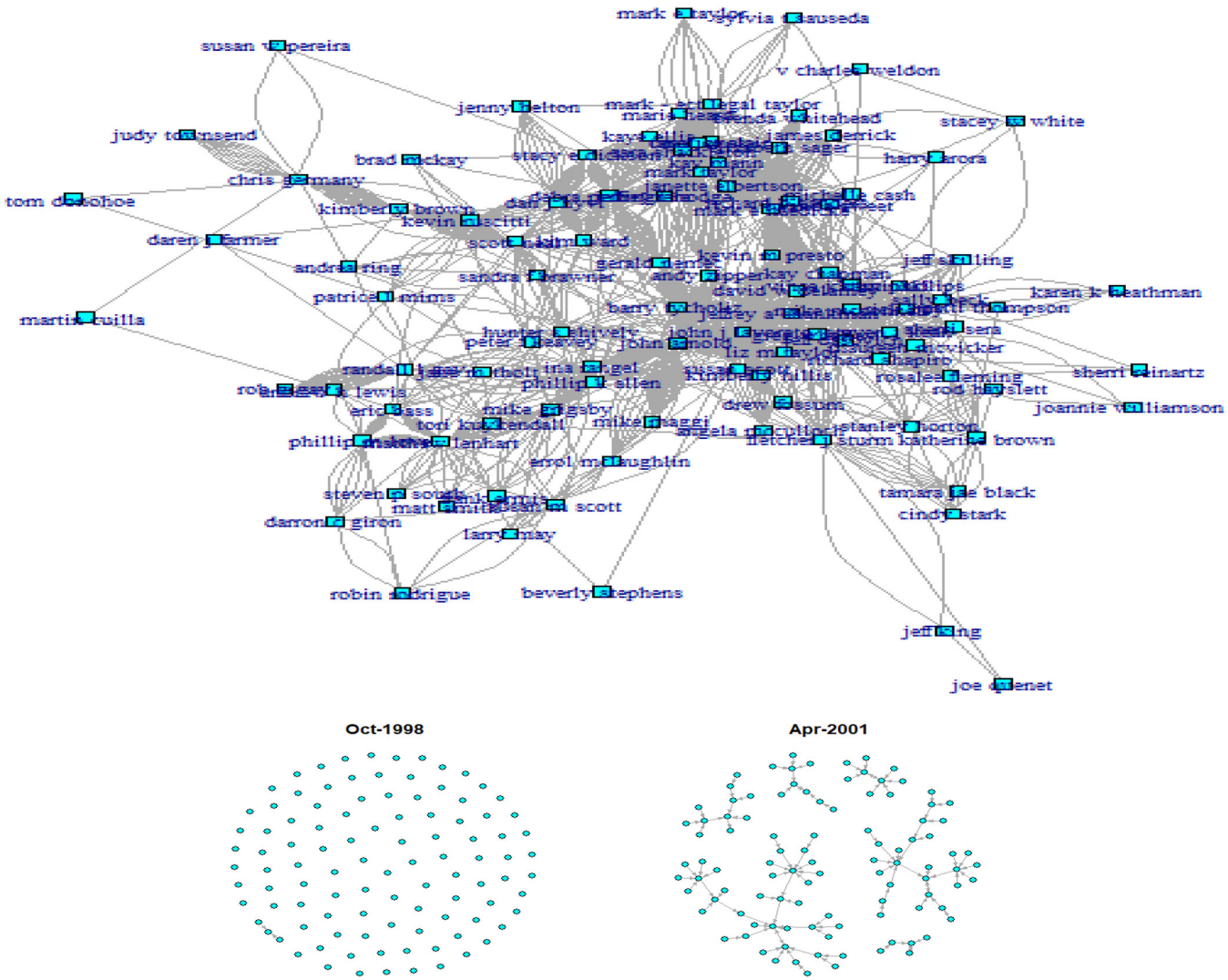


**Figure 5.** Enron employee-to-employee e-mail network (November 1998–July 2001). Top: Cumulative network from November 1998 to July 2001; Bottom (Left): e-mail network at November 1998, (Right): e-mail network at July 2001.

reported by all network summaries under the NEDM improve (and become closer to the 5% level of significance). In particular, we find that the size of the test reported by the NEDM under the maxBETW network summary almost always outperforms the size reported by the BOM (except for one situation when $w = 0.05$, $T = 200$ and $q = 30$). Furthermore, we notice that two other network summaries (ACC and MD) are highly sensitive to increasing $\omega$, and that their size values improve substantially as $\omega$ increases from 0.05 to 0.15. For the BOM results, we find that the size of the test tends to be more liberal as $P$ increases (the size increases from 18% to 20%), and that its performance is inferior to the NEDM with the maxBETW network summary. In conclusion, many of the graph summary statistics appear superior to the BOM in terms of the size of the test.

The results for the one-change point simulation is presented in Figure 4. Overall, the BOM appears to attain the highest power (i.e., it correctly flags the one-change scenario every time). However, this performance is equally matched by the NEDM with the maxBETW network summary. In addition, for other network summaries used in the NEDM, we found that power drops as $\omega$ and $T$ increase, and that power increases as $q$ increases. Furthermore, we notice that as $T$ increases from 200 to 300, the power reported by our NEDM with the CLQNUM summary statistic experiences a substantial drop, but this also improves as $q$ rises.

In summary, we found that both the power and the size of the test (detection ability) are sensitive to the choice of parameters ($q$, $\omega$ and $T$). In addition, while the BOM has excellent power, it suffers from liberal Type I errors. Overall, we found that the NEDM in combination with the maxBETW has the best performance in terms of maintaining the size of the test while also yielding excellent power.

### 6.2. *Case Study: Anxiety fMRI Data*

From our analyses, we obtained the results displayed in Figure 6 for the NEDM (with ACC, APL, maxBETW, and minLCC as summary statistics), the BOM and the KVFM. Results generated by the NEDM with CLQNUM and MD as the network topological summary statistic are deferred to the Supplementary Materials. Previous analyses of these data found change points at the times of the speech instruction slides, primarily time points 60 and 130 (Cribben et al. 2012; Cribben, Wager, and Lindquist

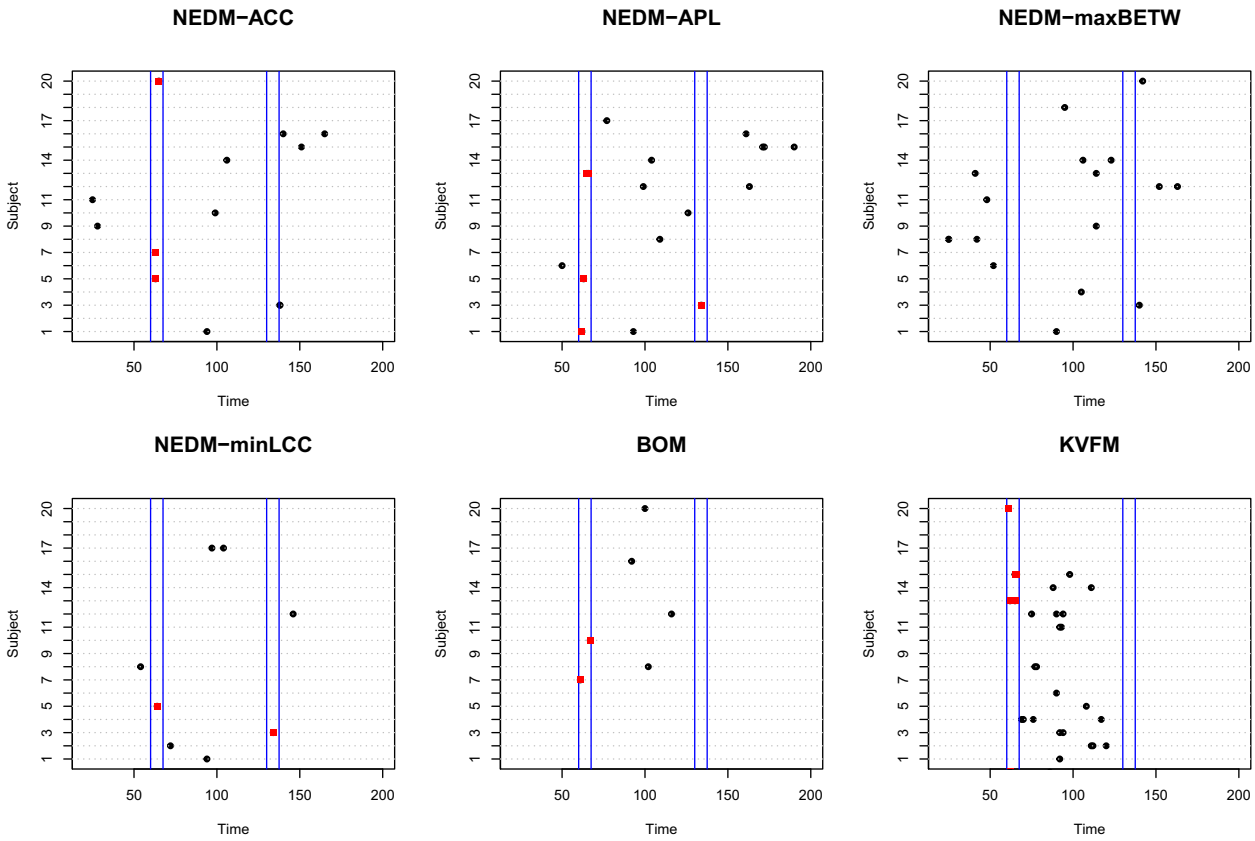


**Figure 6.** Change point detection from the NEDM using graph summary statistics ACC, APL, maxBET, W and minLCC, and the BOM and the KVFM.

2013). Additionally, all results obtained by the NEDM (plus the plot in the Supplementary material) found many of the expected change points at (and between) the time points of the speech instruction slides, and also during the speech preparation phase (time points 60–130).

While the KVFM found more changes points (as expected given it does not provide a quanitifier for statistical significance) at the time points of the speech instruction slides and also during the speech preparation phase, many of the change points for the subjects were very close to one another which makes them unrealistic for fMRI data. While the BOM also found some significant changes points at the speech instruction slides time points, the NEDM found significantly more. In addition, having change points so close to one another (as is the KVFM cases) makes it impossible to estimate the network structure between each pair of change points and also breaks down the ability to interpret and apportion the change phenomenon to a biological process (as in the case of the fMRI data used here). In turn, the framework of the NEDM allows for an estimation which depends on a visual display of the underlying dynamic brain networks with the advantage of noticing structural changes within these embeddings.

### 6.3. *Case study: Enron e-mail networks*

We now turn to the anomaly detection problem for data in the form of a graph, that is, the Enron e-mail network. From our list of network summary statistics, we use the APL because APL appears to be fairly sensitive to intrinsic properties of data in the form of dynamic networks. We link the results (Figure 7) obtained in this analysis to various events in the Enron scandal timeline. With the NEDM, the times for significant detection points occur at months 2, 7, 12, 15, 18 and 24. The time period for month 2 (November 1998 to December 1998) is linked to the hiring of Andrew Fastow as the finance chief. The time period of month 7 was from April 1999 to June 1999. This is around the time when Enron's CFO was exempted by the Board of Directors from the company's code of ethics so that he could run the private equity fund LJM1, and also around the time when the head of Enron's West Coast Trading Desk in Portland Oregon, Timothy Biden, began his first experiment to exploit the new rules of California's deregulated energy market.

Next to this, month 12 is linked to the period August 1999 to October 1999 which was around the time when Enron's CFO started to raise money for two LJM funds (LJM1 and LJM2), which was later used to buy Enron's poorly performing assets in order to make its financial statement look better. The time period for months 15 and 18 was from December 1999 to March 2000. This was close in time that the energy prices in California rose significantly and the power reserves became low, which was followed by the blackouts in metropolitan areas. Many believed that one of the reasons for California's energy crisis was Enron's trading, which led to the investigation of Federal Energy Regulatory Commission (FERC). This investigation was connected with our next detection point at month 24, which was from August 2000 to December 2000, since it was the time the FERC investigation exonerated Enron from wrongdoings in the

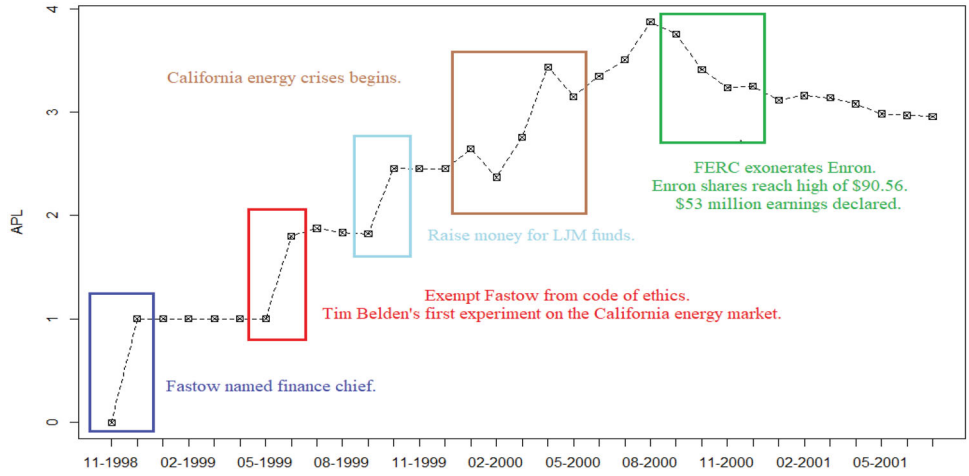


**Figure 7.** Change point detection within the Enron e-mail networks from November 1998 to July 2001 using the NEDM.

California energy crisis. The time period in the vicinity month 24 is a very interesting one not only because it is connected with FERC investigation, but also because Enron's share price hit an all-time high of $90.56 and then Enron used "aggressive" accounting to declare $53 million in earnings on a collapsing deal that hadn't earned anything at all in profit.

Figure 4 in the supplementary materials displays the outcome of the KVFM analysis applied to the Enron data. Note that, because the KVFM is based on graph similarity scores, we only have results from December 1998 to June 2001 (time point 0 is November 1998, and there are no more networks for comparison after July 2001). With the KVFM, anomalies are reported at times 2, 6 and 23. These times coincide with time periods [December 1998–January 1999], [April 1999–June 1999] and [September 2000–November 2000]. In Figure 2 (in the *supplementary materials*), we notice that the KVFM is only able to flag 3 out 5 anomalous time points that the NEDM reported.

## 7. Conclusion

In this article, we develop a new approach, the NEDM, for analyzing and modeling the network structure between (possibly) high dimensional multivariate time series from an fMRI study which consists of realizations of complex and dynamic brain processes. The method adds to the literature by improving understanding of the brain processes measured using fMRI. The NEDM is, to the best of our knowledge, the first paper to consider estimating change points for time evolving graph summary statistics in a multivariate time series context. Although this article is inspired by and developed for brain connectivity studies, our proposed method is applicable to more general settings and can also be used in a variety of situations where one wishes to study the evolution of a high-dimensional graph over time, that is, in conjunction with telecommunication, financial, and blockchain networks (Cribben 2019; Keshavarz, Michailidis, and Atchade 2013; Chen et al. 2020).

There are several novel aspects of the NEDM. First, it allows for estimation of graph summary statistics in a (possible) very high-dimensional multivariate time series setting, in particular, in situations where the number of time series is much greater than the number of time points ($P >> T$). Hence, in a biomedical neuroimaging setting, it can consider the dynamics of the whole brain or a very large number of brain time series, thereby providing deeper insights into the large-scale functional architecture of the brain and the complex processes within. Second, the NEDM is, to the best of our knowledge, the first piece of work to consider estimating change points of time evolving graph summary structure in a multivariate time series context. We introduced a novel statistical test for the candidate change points using the sieve bootstrap and showed that it outperformed the asymptotic distribution. However, as the NEDM is based on binary segmentation it is restricted by the minimum distance between change points.

It has been shown that neurological disorders disrupt the connectivity pattern or structural properties of the brain. Future work entails applying the NEDM to resting state fMRI data from subjects with brain disorders such as depression, Alzheimer's disease and schizophrenia and to control subjects who have been matched using behavioral data. By comparing change points and partition specific networks, the NEDM may lead to the robust identification of cognitive states at rest for both controls and subjects with these disorders. It is hoped that the large-scale temporal features resulting from the accurate description of brain connectivity from our novel method, which might lead to better diagnostic and prognostic indicators of the brain disorders. More specifically, by comparing the change points of healthy controls to patients with these disorders, we may be able understand the key differences in functional brain processes that may eventually lead to the identification of biomarkers for the disease.

As an extension to monitoring change points in graph objects, we intend to incorporate higher order structures, such as tensors, in the network snapshot characterization procedure of the NEDM. Moreover, we intend to extend our analysis to include the estimation of network summaries that are based on the *local* topology and geometry of the graph. In particular, we intend to incorporate a motif-based analysis (Milo et al. 2002; Dey, Gel, and Poor 2019; Sarkar, Guo, and Shakarian 2019) and the concepts of topological data analysis (TDA), particularly, persistent homology, in the derivation of graph

summary statistics (Carlsson 2009; Patania, Vaccarino, and Petri 2017; Carlsson 2019). Indeed, tracking local network topological summaries based on graph persistent homology offers multifold benefits. First, this approach enables us to consider edge-weighted networks. Second, it allows for enhancing analysis of the underlying network organization at multi-resolution levels. Third, simultaneously considering multiple local network topological statistics based on graph persistent homology minimizes the loss of network information that currently occurs due to reducing a high dimensional structure to a univariate time series representation of a single network summary.

Another interesting theoretical direction is to explore various types of regularized approximation models (Gel and Barabanov 2007; Bickel and Gel 2011; Politis 2015) for dynamics of local and global network topological summaries, and the associated error bounds, which as a result, can also yield an insight on theoretical guarantees of resampling and subsampling procedures (Kreiss et al. 2011; Fragkeskou and Paparoditis 2018) in application to (non)linear processes of network topological descriptors and related uncertainty quantification in network anomaly detection.

## Supplementary Material

**R Code and Data :** The supplemental files for this article include files containing R code and data for reproducing all the simulated and empirical studies in the paper.

**Appendix :** The supplemental files include an Appendix which contains the following: (i) Theorem 1 from Gombay (2008) and a Notation table, (ii) Simulation analysis for the *NEDM* size and power with $P = 50$, (iii) Additional results for the *NEDM* applied to the Anxiety fMRI data under the remaining network summaries, (iv) Plot for the *KVFM* Enron e-mail network analysis, (v) Analysis and discussion of case study (resting state fMRI data), (vi) Proofs for Theorem 1 and Corollary 1, (vii) Procedure for comparing asymptotic and sieve bootstrap distributions.

## Acknowledgements

The authors would like to thank Gagan S. Wig for stimulating discussions on brain networks.

## Funding

The work of D.O.B and Y.R.G. was partially supported by the National Science Foundation (NSF) of the United States (grant numbers IIS 1633331, DMS 1925346 and DMS 1736368). The work of I.C was partially supported by Natural Sciences and Engineering Research Council of Canada (NSERC: Grant/Award Number: RGPIN-2018-06638) and the Xerox Faculty Fellowship, Alberta School of Business.

## References

Abay, N. C., Akcora, C. G., Gel, Y. R., Kantarcioglu, M., Islambekov, U. D., Tian, Y., and Thuraisingham, B. (2019), "Chainnet: Learning on Blockchain Graphs With Topological Features," in *2019 IEEE International Conference on Data Mining (ICDM)*, Beijing, China: IEEE. [1]

Akcora, C. G., Li, Y., Gel, Y. R., and Kantarcioglu, M. (2020), "BitcoinHeist: Topological Data Analysis for Ransomware Detection on the Bitcoin Blockchain," in *International Joint Conference on Artificial Intelligence (IJCAI)*, Japan. [1]

Akoglu, L. and Faloutsos, C. (2010), "Event Detection in Time Series of Mobile Communication Graphs," Army Science Conference. [1]

Aue, A., Hörmann, S., Horváth, L., and Reimherr, M. (2009), "Break Detection in the Covariance Structure of Multivariate Time Series Models," *The Annals of Statistics*, 37, 4046–4087. [2]

Barabási, A.-L. and Albert, R. (1999), "Emergence of Scaling in Random Networks," *Science*, 286, 509–512. [2]

Barabási, A.-L. and Pósfai, M. (2016), *Network Science*, Cambridge University Press. [3]

Barigozzi, M., Cho, H., and Fryzlewicz, P. (2018), "Simultaneous Multiple Change-point and Factor Analysis for High-dimensional Time Series," *Journal of Econometrics*, 206, 187–225. [2]

Barnett, I. and Onnela, J.-P. (2016), "Change Point Detection in Correlation Networks," *Scientific reports*, 6, 18893. [2,5,6]

Bassett, D. S. and Bullmore, E. (2006), "Small-World Brain Networks," *The Neuroscientist*, 12, 512–523. [1]

Bickel, P. J. and Gel, Y. R. (2011), "Banded Regularization of Autocovariance Matrices in Application to Parameter Estimation and Forecasting of Time Series," *Journal of the Royal Statistical Society*, Series B, 73, 711–728. [11]

Bühlmann, P. (1997), "Sieve Bootstrap for Time Series," *Bernoulli*, 3, 123–148. [2,4]

Carlsson, G. (2009), "Topology and Data," *Bulletin of the American Mathematical Society*, 46. [11]

——— (2019), "Persistent homology and Applied Homotopy Theory," in *Handbook of Homotopy Theory*. [11]

Chen, F., Wan, H., Cai, H., and Cheng, G. (2020), "Machine Learning in/for Blockchain: Future and Challenges", arXiv preprint arXiv:1909.06189v2. [10]

Cribben, I. (2019), "Change Points in Heavy-tailed Multivariate Time Series: Methods Using Precision Matrices," *Applied Stochastic Models in Business and Industry*, 35, 299–320. [10]

Cribben, I., Haraldsdottir, R., Atlas, L. Y., Wager, T. D., and Lindquist, M. A. (2012), "Dynamic Connectivity Regression: Determining State-related Changes in Brain Connectivity," *Neuroimage*, 61, 907–920. [2,6,8]

Cribben, I., Wager, T., and Lindquist, M. (2013), "Detecting Functional Connectivity Change Points for Single-subject fMRI Data," *Frontiers in Computational Neuroscience*, 7, 143. [2,6,9]

Cribben, I., and Yu, Y. (2017), "Estimating Whole-brain Dynamics by Using Spectral Clustering," *Journal of the Royal Statistical Society*, Series C, 66, 607–627. [2]

Dette, H., and Wied, D. (2016), "Detecting Relevant Changes in Time Series Models," *Journal of the Royal Statistical Society*, Series B, 78, 371–394. [2]

Dey, A. K., Gel, Y. R., and Poor, H. V. (2019), "What Network Motifs Tell Us about Robustness and Reliability of Complex Networks," *Proceedings of the National Academy of Sciences*, available at www.pnas.org/cgi/doi/10.1073/pnas.1819529116. [10]

Diesner, J., Frantz, T. L., and Carley, K. M. (2005), "Communication Networks from the Enron Email Corpus "It's Always About the People. Enron is No Different"," *Computational & Mathematical Organization Theory*, 11, 201–228. [2,7]

Elliott, M., Golub, B., and Jackson, M. O. (2014), "Financial Networks and Contagion," *American Economic Review*, 104, 3115–53. [1]

Fragkeskou, M., and Paparoditis, E. (2018), "Extending the Range of Validity of the Autoregressive (sieve) Bootstrap," *Journal of Time Series Analysis*, 39, 356–379. [11]

Gel, Y. R., and Barabanov, A. (2007), "Strong Consistency of the Regularized Least-squares Estimates of Infinite Autoregressive Models," *Journal of Statistical Planning and Inference*, 137, 1260–1277. [11]

Gombay, E. (2008), "Change Detection in Autoregressive Time Series," *Journal of Multivariate Analysis*, 99, 451–464. [1,3,4,5,11]

Hamilton, J. D. (1995), "Time Series Analysis," in *Economic Theory. II*, Princeton, NJ: Princeton University Press, pp. 625–630. [5]

Harshaw, C. R., Bridges, R. A., Iannacone, M. D., Reed, J. W., and Goodall, J. R. (2016), "Graphprints: Towards a Graph Analytic Method for Network Anomaly Detection," in *Proceedings of the 11th Annual Cyber and Information Security Research Conference*, Oak Ridge, TN: ACM. [1]

Host-Madsen, A. and Zhang, J. (2018), "Coding of Graphs with Application to Graph Anomaly Detection," arXiv preprint arXiv:1804.02469. [1]

Inoue, A. (2001), "Testing for Distributional Change in Time Series," *Econometric Theory*, 17, 156–187. [3]

Keogh, E., Chu, S., Hart, D., and Pazzani, M. (2001), "An Online Algorithm for Segmenting Time Series," in *Data Mining, 2001. ICDM 2001, Proceedings IEEE International Conference on*, San Jose, CA: IEEE. [3]

Keshavarz, K., Michailidis, G., and Atchade, Y. (2020), "Sequential Change-Point Detection in High-Dimensional Gaussian Graphical Models," *Journal of Machine Learning Research*, 21, 1–57. [10]

Koutra, D., Vogelstein, J. T., and Faloutsos, C. (2013), "Deltacon: A Principled Massive-graph Similarity Function," in *Proceedings of the 2013 SIAM International Conference on Data Mining*, Austin, TX: SIAM. [2,5,6]

Kreiss, J.-P. (1988), *Asymptotic Statistical Inference for a Class of Stochastic Processes*, Ph.D. thesis, Verlag nicht ermittelbar. [2,4]

——— (1992), "Bootstrap procedures for AR (∞)—processes," in K.-H. Jöckel, G. Rothe, and W. Sendler, eds., *Bootstrapping and Related Techniques*, Vol. 376, Berlin, Heidelberg: Springer-Verlag, 107–113. [2,4]

Kreiss, J.-P., Paparoditis, E., and Politis, D. N. (2011), "On the Range of Validity of the Autoregressive Sieve Bootstrap," *The Annals of Statistics*, 39, 2103–2130. [11]

Messer, M., Albert, S., and Schneider, G. (2018), "The Multiple Filter Test for Change Point Detection in Time Series," *Metrika*, 81, 589–607. [1]

Milo, R., Shen-Orr, S., Itzkovitz, S., Kashtan, N., Chklovskii, D., and Alon, U. (2002), "Network Motifs: Simple Building Blocks of Complex Networks," *Science*, 298, 824–827. [10]

Newman, M. E. (2003), "The Structure and Function of Complex Networks," *SIAM Review*, 45, 167–256. [3]

Ozkaya, Y., Sariyuce, A. E., Catalyurek, U. V., and Pinar, A. (2017), "Active Betweenness Cardinality: Algorithms and Applications," arXiv preprint arXiv:1711.10634. [3]

Park, Y., Priebe, C. E., and Youssef, A. (2012), "Anomaly Detection in Time Series of Graphs Using Fusion of Graph Invariants," *IEEE Journal of Selected Topics in Signal Processing*, 7, 67–75. [2,7]

Patania, A., Vaccarino, F., and Petri, G. (2017), "Topological Analysis of Data," *EPJ Data Science*, 6. [11]

Peel, L. and Clauset, A. (2015), "Detecting Change Points in the Large-Scale Structure of Evolving Networks." in *AAAI*, volume 15. [1,7]

Picard, D. (1985), "Testing and Estimating Change-Points in Time Series," *Advances in Applied Probability*, 17, 841–867. [3]

Politis, D. N. (2015), "Model-Free Prediction in Regression," in S. Datta, ed., *Model-Free Prediction and Regression*, Cham, Switzerland: Springer, pp. 57–80. [11]

Pourahmadi, M. (2001), *Foundations of Time Series Analysis and Prediction Theory*, Vol. 379, New York, NY: Wiley. [3]

Priebe, C. E., Conroy, J. M., Marchette, D. J., and Park, Y. (2005), "Scan Statistics on Enron Graphs," *Computational & Mathematical Organization Theory*, 11, 229–247. [2,7]

Rubinov, M. and Sporns, O. (2010), "Complex Network Measures of Brain Connectivity: Uses and Interpretations," *Neuroimage*, 52, 1059–1069. [1]

Sarkar, S., Guo, R., and Shakarian, P. (2019), "Using Network Motifs to Characterize Temporal Network Evolution Leading to Diffusion Inhibition," *Social Network Analysis and Mining*, 9, 14. [10]

Shumway, R. H. and Stoffer, D. S. (2017), *Time Series Analysis and Its Applications: With R Examples*, Cham, Switzerland: Springer. [3]

Sun, J., Faloutsos, C., Papadimitriou, S., and Yu, P. S. (2007), "Graphscope: Parameter-free Mining of Large Time-evolving Graphs," in *Proceedings of the 13th ACM SIGKDD International Conference on Knowledge Discovery and Data Mining*, New York, NY: ACM, pp. 687–696. [2]

Vandermarliere, B., Karas, A., Ryckebusch, J., and Schoors, K. (2015), "Beyond the Power Law: Uncovering Stylized facts in interbank networks," *Physica A: Statistical Mechanics and its Applications*, 428, 443–457. [1]

Wager, T. D., Waugh, C. E., Lindquist, M., Noll, D. C., Fredrickson, B. L., and Taylor, S. F. (2009), "Brain Mediators of Cardiovascular Responses to Social Threat: Part I: Reciprocal Dorsal and Ventral Sub-regions of the Medial Prefrontal Cortex and Heart-rate Reactivity," *Neuroimage*, 47, 821–835. [7]

Wig, G. S., Schlaggar, B. L., and Petersen, S. E. (2011), "Concepts and Principles in the Analysis of Brain Networks," *Annals of the New York Academy of Sciences*, 1224, 126–146. [1]

Wolff, S., Gastwirth, J., and Rubin, H. (1967), "The Effect of Autoregressive Dependence on a Nonparametric Test (corresp.)," *IEEE Transactions on Information Theory*, 13, 311–313. [1]

Zellner, A. (1962), "An Efficient Method of Estimating Seemingly Unrelated Regressions and Tests for Aggregation Bias," *Journal of the American statistical Association*, 57, 348–368. [1,5]

Zivot, E. and Wang, J. (2007), *Modeling Financial Time Series with S-Plus®*, volume 191, Springer Science & Business Media. [6]