Nonparametric Anomaly Detection on Time Series of Graphs

Dorcas Ofori-Boateng, Yulia R.Gel*
Department of Mathematical Sciences, University of Texas at Dallas, U.S.A.
and
Ivor Cribben†
Department of Accounting, Operations, and Information Systems,
Alberta School of Business, Canada

November 3, 2020

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 of the particular objectives of the anomaly detection task from the 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) data sets 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 intrinsic structure of complex dynamic networks.

*The work of D.O.B and Y.R.G. was partially supported by NSF grant numbers IIS 1633331, DMS 1736368 and DMS 1925346.

†The work of I.C was partially supported by NSERC: Grant/Award Number: RGPIN-2018-06638 and the Xerox Faculty Fellowship, Alberta School of Business.

Keywords: Dynamic Networks, Multivariate Time Series, Change point, 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 et al., 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 et al., 2014; Vandermarliere et al., 2015; Abay et al., 2019; Akcora et al., 2020). Similar to other biological networks (Wig et al., 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., Peel and Clauset, 2015; Akoglu and Faloutsos, 2010; 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 et al., 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 coeffi-

cient, 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 of 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 data (Cribben et al., 2012, 2013; Cribben and Yu, 2017) and a dynamic network of the Enron email communication (Park et al., 2012; Priebe et al., 2005; Diesner et al., 2005). For the sake of brevity, we defer the analysis of one fMRI data set 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 paper 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 data sets in Section 5. The performance of the proposed algorithm, on both the synthetic and real world data, is detailed in Section 6. Conclusion and future work is provided 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 et al. (2013) relies 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 et al. (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, 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 no distributional assumptions.

3 Methodology

In this section, we describe the proposed contribution which tracks structural changes within the network structure of data sets. 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 paper.

3.1 Input Data: Graphs and Multivariate Time Series

Our proposed methodology is applicable to two types of the 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, e.g., correlation. Since one of our primary motivating applications is multivariate fMRI time series, below we describe in details how networks can be constructed from such data sets.

Networks from a multivariate time series In many applications such as, for instance, 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 end points (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 ω[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, e.g., 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.

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

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:

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

We assume that the appropriate model to fit to $\{Y_t\}_{t=1}^n$ is a strictly stationary and purely non-deterministic autoregressive $[AR(p)]$ model with Gaussian independent and identically distributed (i.i.d.) white noise ε_t. The assumption of Gaussianity for ε_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 α-level of 0.05, and an $AR(1)$ process with φ 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 net-

works, 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

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

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

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

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 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 utilizes the efficient score vector $\nabla_{\boldsymbol{\xi}} \ell_k(Y_1, \ldots, Y_n; \boldsymbol{\xi}) = \nabla_{\boldsymbol{\xi}} \ell_k(\boldsymbol{\xi})$ with ℓ_k as the log-likelihood function on $\{Y_t\}_{t=1}^n$. For 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

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

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

If ξ_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:

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

and

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

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

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

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

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

If we are interested in *two-sided* hypothesis testing, a change in ξ_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

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

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 lower power of the test. Under the premise that $\{Y_t\}_{t=1}^n$ follows an $AR(p)$ model with Gaussian i.i.d. 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(∞) 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.

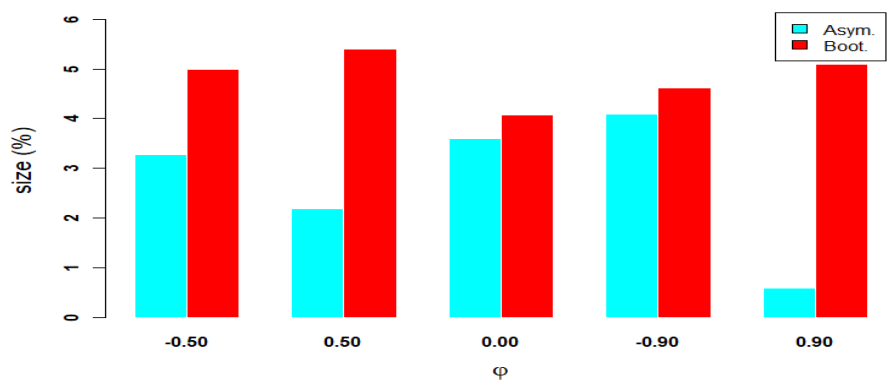

Figure 1: The size of the test plotted against φ for the asymptotic distribution and sieve bootstrap procedure when testing for a change in the mean of an $AR(1)$ model using $T = 100$.

To compare 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 Appendix]. The choice of the model coefficient (φ) depends on ensuring the assumption of weak stationarity for the simulated AR(1) series. Indeed note that as φ 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.

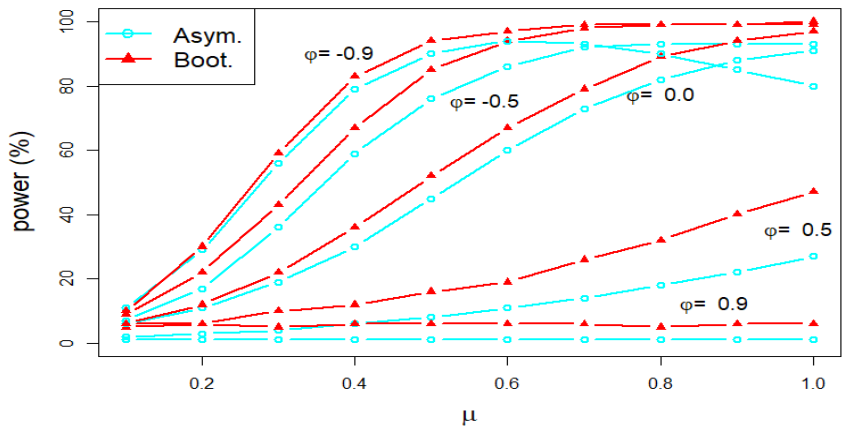

Figure 2: Power of the test plotted against φ for the asymptotic distribution and sieve bootstrap procedure when testing for a change in the mean of an $AR(1)$ model with ($T = 100$).

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 φ is closer to $+1$ (*i.e.,* the size is worse when $\varphi = 0.9$ compared to $\varphi = 0.5$). However for negative φ cases, we find that the size of the test under the asymptotic distribution approaches the declared 5% level of significance as φ 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 φ approaches $+1$ the size of the 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 notice that as μ increases, the power of the test also increases under both the asymptotic and the bootstrap distributions. In addition, as φ 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 φ 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]

Algorithm 1: 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 α; Number of bootstrap replications Γ.

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

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

Obtain $\widetilde{\varepsilon_t}$ from the residuals $\hat{\varepsilon}_t = X_t - \sum_{k=1}^p \hat{\varphi}_k X_{t-k}$.

(**Note**: $\widetilde{\varepsilon_t} = \hat{\varepsilon}_t - \bar{\varepsilon}_n$, where $\bar{\varepsilon}_n = \dfrac{1}{n}\sum_{t=1}^n \hat{\varepsilon}_t$).

for $i \leftarrow$ *1 :* Γ **do**

  Sample, with replacement, $\{\widetilde{\varepsilon_t^*}\}_{t=1}^n$ from $\{\widetilde{\varepsilon_t}\}_{t=1}^n$.

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

  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 $\sum_{i=1}^{\Gamma} \dfrac{Z_i}{\Gamma}$.

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*

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

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

Proof of Theorem 1 is in the Appendix.

Suppose $\widetilde{\varepsilon}_t = \hat{\varepsilon}_t - \bar{\varepsilon}_n$, 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 $\{\widetilde{\varepsilon}_t^{\diamond}\}_{t=1}^{n} \sim N(0, s_n^2)$, with $s_n^2 = \sum_{t=1}^{n}(\widetilde{\varepsilon}_t - \bar{\widetilde{\varepsilon}}_t)^2/(n-1)$ as the sample variance of $\{\widetilde{\varepsilon}_t\}_{t=1}^{n}$. This yields a finite sample performance similar to that of $\{\widetilde{\varepsilon}_t^{*}\}_{t=1}^{n}$, with reduced computing time.

Corollary 1 If Y_t satisfies equation (1.1), then under H_0

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

Proof of Corollary 1 is in the Appendix.

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, ω and ROI choices. We do not compare the $NEDM$ to the $KVFM$ (Koutra et al., 2013) because it does not provide a quantifier for the statistical significance (ac$p$-value) 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 length $q = \{5, 10, 15, 20, 25, 30\}$, threshold parameter $\omega = \{0.05, 0.1, 0.15\}$, 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 material.

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 4 and 5, 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 emails 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 et al., 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, 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 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 data set. 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 email networks

The Enron emails data set is a benchmark data set applied in numerous instances of anomaly detection (e.g., Peel and Clauset, 2015; Park et al., 2012; Priebe et al., 2005; Diesner et al., 2005). More information on this data set can be found online (`http://www.cs.cmu.edu/~enron/`). We used the cleaned version of the employee-to-employee

email (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 email between two employees within the month under study, an edge is connected to the respective nodes. Figure 3 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.

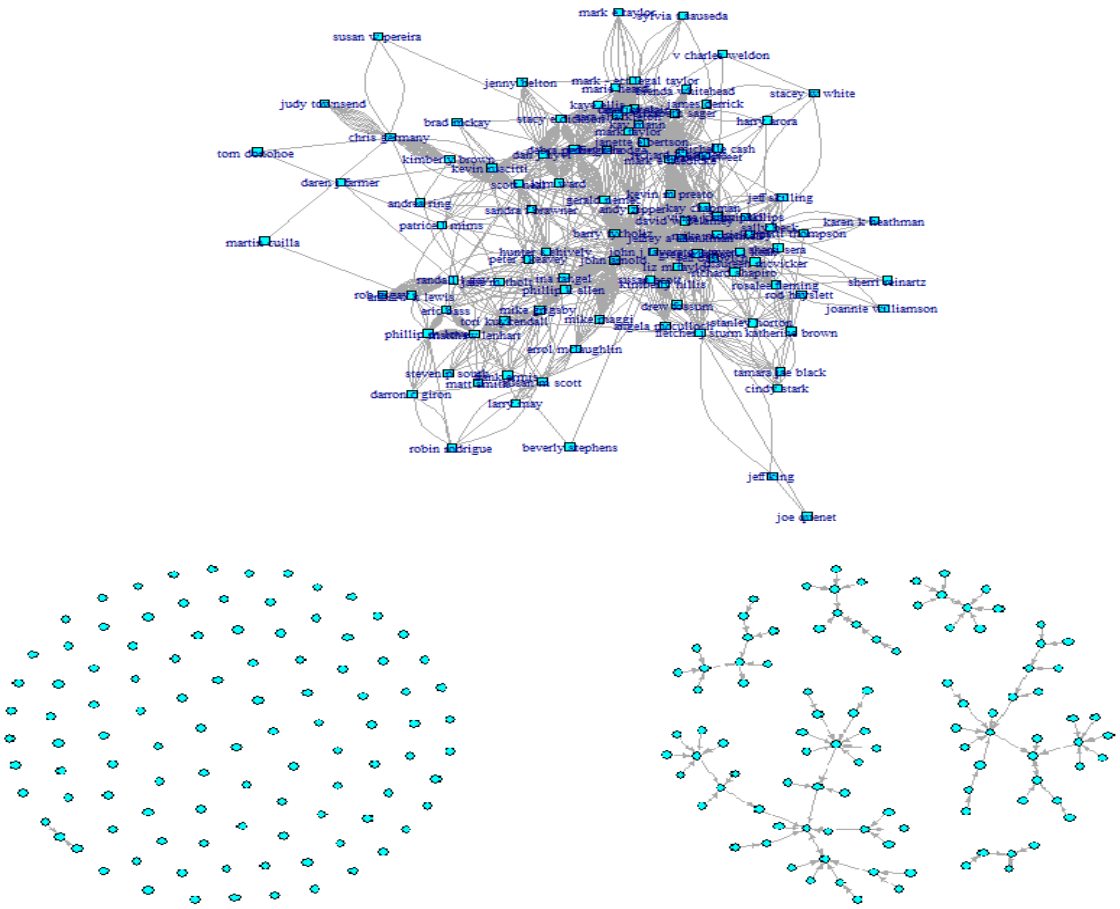

Figure 3: Enron employee-to-employee email network (November 1998 - July 2001). Top: Cumulative network from November 1998 to July 2001; Bottom (Left): Email network at November 1998, (Right): Email network at July 2001.

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 data set.

6.1 Simulation study

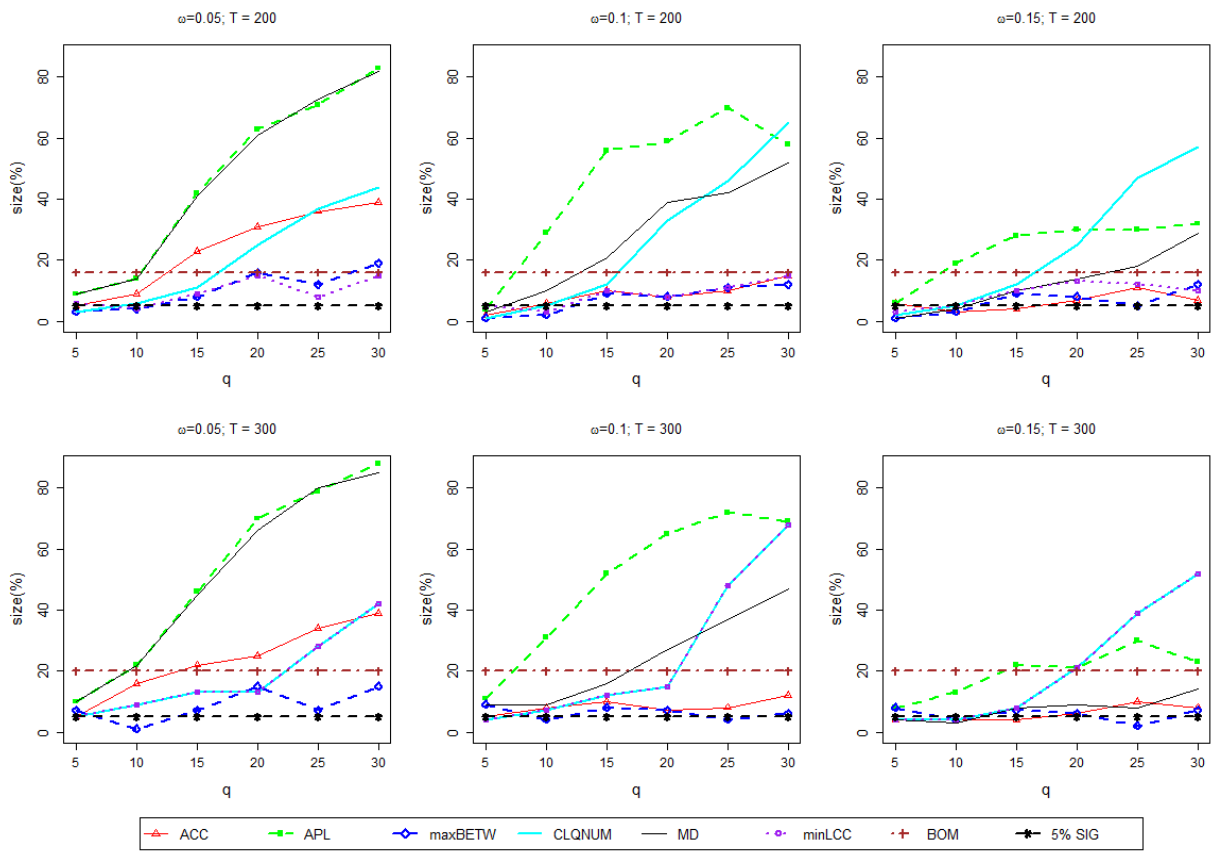

Figure 4: The size of the test for the no-change (change in mean) scenario with $T = \{200, 300\}$, $\omega = \{0.05, 0.1, 0.15\}$, $q = \{5, 10, 15, 20, 25, 30\}$ and $P = 5$.

Figure 4 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 ω increases, the size 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 ω, and that their size values improve substantially as ω 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 5. 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 utilized in the $NEDM$, we found that power drops as ω 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, ω 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 having 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 $KVFM$.

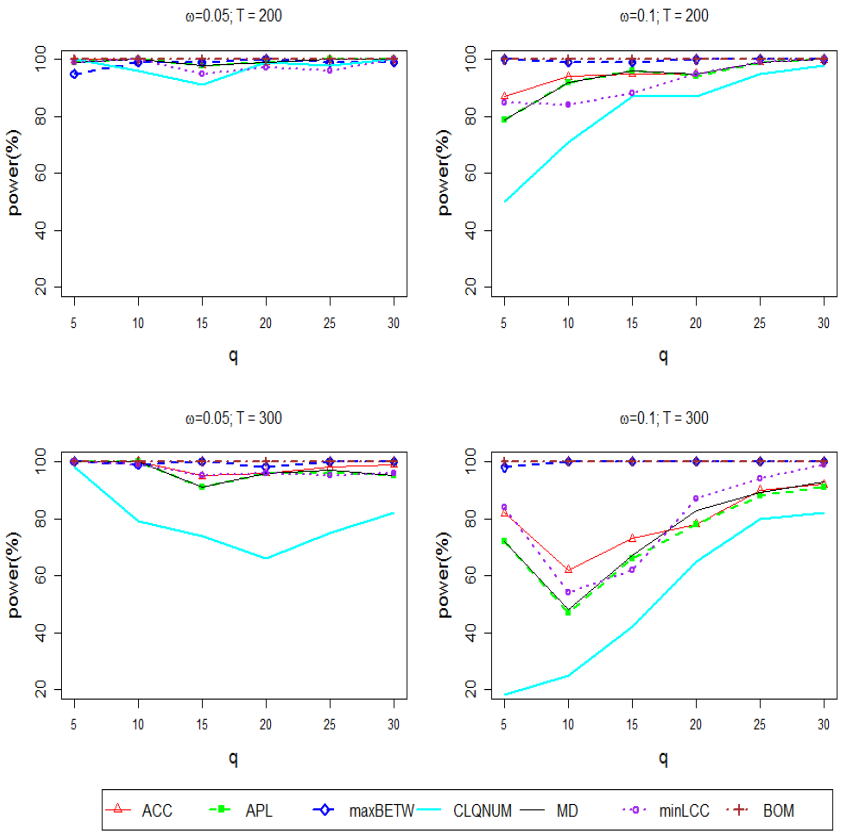

Figure 5: The power of the test for the one-change (change in variance) scenario with $\omega = \{0.05, 0.1\}$, $q = \{5, 10, 15, 20, 25, 30\}$, $T = \{200, 300\}$, $\tau = \{88, 156\}$ and $P = 10$.

Results generated by the $NEDM$ with $CLQNUM$ and MD as the network topological summary statistic are deferred to the Supplementary material. 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, 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

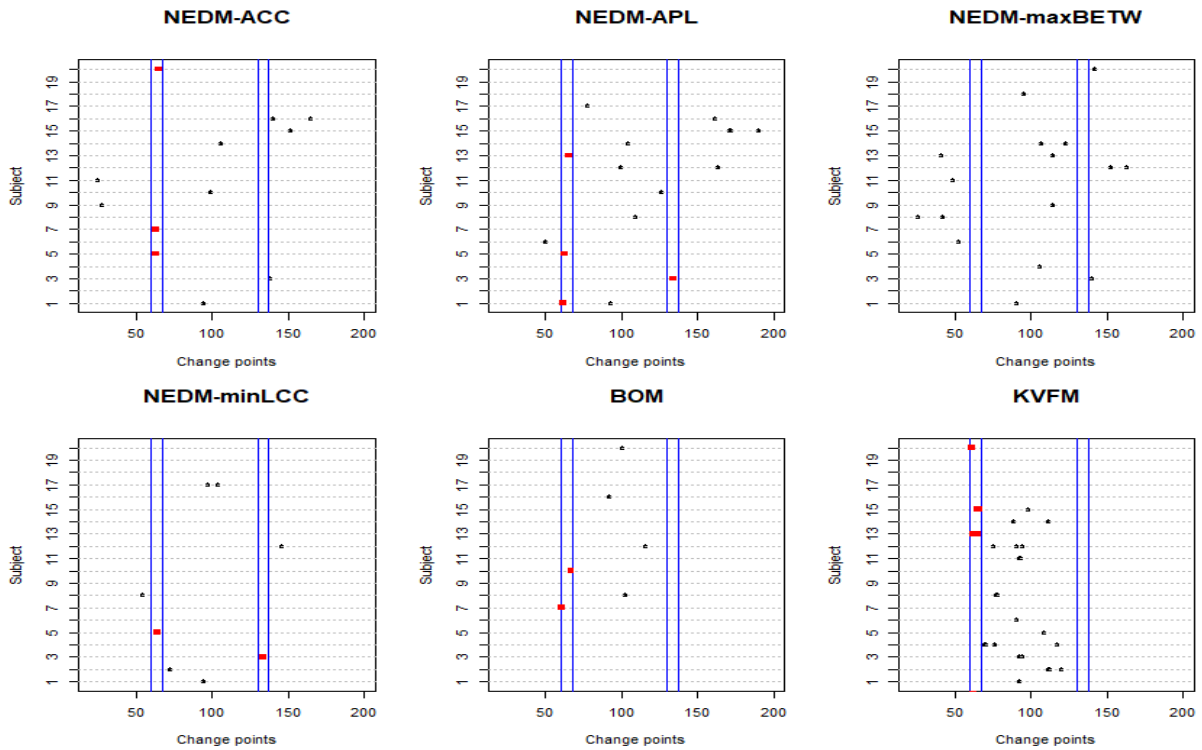

Figure 6: Change point detection for the anxiety fMRI data.

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 significant 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 email networks

We now turn to the anomaly detection problem for data in the form of a graph, that is, the Enron email network. From our list of network summary statistics, we utilize 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. In comparison, the time period for month 2 (November 1998 to December 1998) is linked to the hire 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.

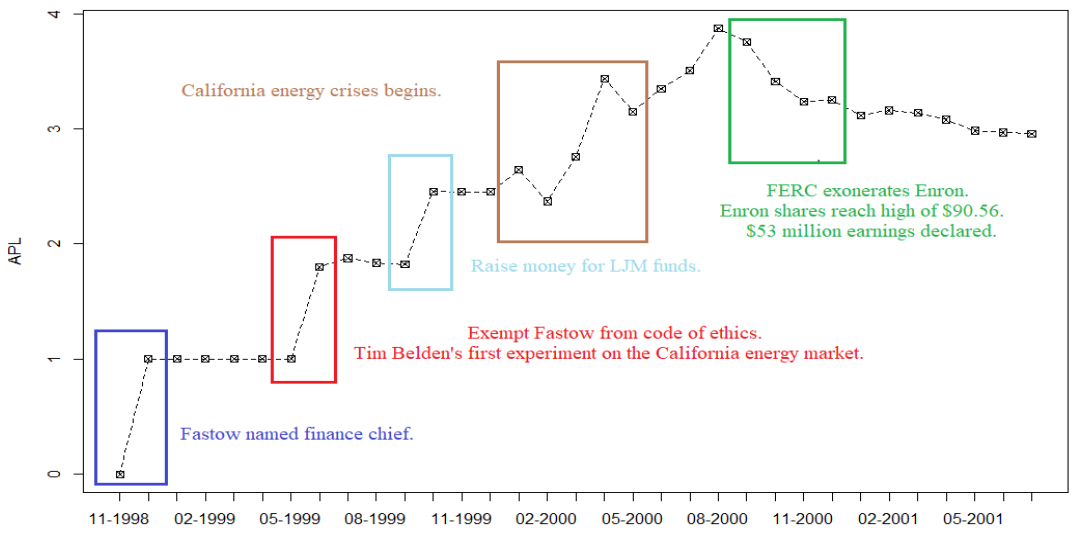

Figure 7: Detection of Enron Email Networks from November 1998 - July 2001.

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 leaded 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 California energy crisis. The time period around month 24 is a very interesting one not only because it is connected with FERC investigation, but also because Enron's share price hit 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 2 in the Appendix displays the outcome of the $KVFM$ analysis. 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 paper, 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 paper 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, e.g., in conjunction with telecommunication, financial (Cribben, 2019), and blockchain networks.

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 behavioural 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 et al., 2019; Sarkar et al., 2019) and the concepts of topological data analysis (TDA), particularly, persistent homology, in the derivation of graph summary statistics (Carlsson, 2009; Patania et al., 2017; Carlsson, 2019). Indeed, tracking local network topological summaries based

on graph persistent homology offers multi-fold 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 Notations 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 email 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 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. The authors would like to thank Gagan S. Wig for stimulating discussions on brain networks.

References

Abay, Nazmiye Ceren, Akcora, Cuneyt Gurcan, Gel, Yulia R, Kantarcioglu, Murat, Islambekov, Umar D, Tian, Yahui, and Thuraisingham, Bhavani (2019),, Chainnet: Learning on blockchain graphs with topological features. in *2019 IEEE International Conference on Data Mining (ICDM)*, 946–951. IEEE.

Akcora, Cuneyt Gurcan, Li, Yitao, Gel, Yulia R, and Kantarcioglu, Murat (2020),, Bitcoin-Heist: Topological Data Analysis for Ransomware Detection on the Bitcoin Blockchain. in *International Joint Conference on Artificial Intelligence (IJCAI)*.

Akoglu, Leman and Faloutsos, Christos (2010),, Event detection in time series of mobile communication graphs. 77–79. Army science conference.

Aue, Alexander, Hörmann, Siegfried, Horváth, Lajos, Reimherr, Matthew, et al. (2009), "Break detection in the covariance structure of multivariate time series models," *The Annals of Statistics*, 37, 6B, 4046–4087.

Barabási, Albert-László and Albert, Réka (1999), "Emergence of scaling in random networks," *science*, 286, 5439, 509–512.

Barabási, Albert-László and Pósfai, Márton (2016), *Network science*, Cambridge university press.

Barigozzi, Matteo, Cho, Haeran, and Fryzlewicz, Piotr (2018), "Simultaneous multiple change-point and factor analysis for high-dimensional time series," *Journal of Econometrics*.

Barnett, Ian and Onnela, Jukka-Pekka (2016), "Change point detection in correlation networks," *Scientific reports*, 6, 18893.

Bassett, Danielle Smith and Bullmore, ED (2006), "Small-world brain networks," *The neuroscientist*, 12, 6, 512–523.

Bickel, Peter J and Gel, Yulia 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 (Statistical Methodology)*, 73, 5, 711–728.

Bühlmann, Peter et al. (1997), "Sieve bootstrap for time series," *Bernoulli*, 3, 2, 123–148.

Carlsson, G. (2009), "Topology and data," *Bulletin of the American Mathematical Society*, 46, 2.

Carlsson, Gunnar (2019), "Persistent homology and applied homotopy theory," *Handbook of Homotopy Theory*.

Cribben, Ivor (2019), "Change points in heavy-tailed multivariate time series: Methods using precision matrices," *Applied Stochastic Models in Business and Industry*, 35, 2, 299–320.

Cribben, Ivor, Haraldsdottir, Ragnheidur, Atlas, Lauren Y, Wager, Tor D, and Lindquist, Martin A (2012), "Dynamic connectivity regression: determining state-related changes in brain connectivity," *Neuroimage*, 61, 4, 907–920.

Cribben, Ivor, Wager, Tor, and Lindquist, Martin (2013), "Detecting functional connectivity change points for single-subject fMRI data," *Frontiers in computational neuroscience*, 7, 143.

Cribben, Ivor and Yu, Yi (2017), "Estimating whole-brain dynamics by using spectral clustering," *Journal of the Royal Statistical Society: Series C (Applied Statistics)*, 66, 3, 607–627.

Dette, Holger and Wied, Dominik (2016), "Detecting relevant changes in time series models," *Journal of the Royal Statistical Society: Series B (Statistical Methodology)*, 78, 2, 371–394.

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, www.pnas.org/cgi/doi/10.1073/pnas.1819529116*.

Diesner, Jana, Frantz, Terrill L., and Carley, Kathleen M. (2005), "Communication networks from the Enron email corpus It's always about the people. Enron is no different," *Computational & Mathematical Organization Theory*, 11, 3, 201–228.

Elliott, Matthew, Golub, Benjamin, and Jackson, Matthew O (2014), "Financial networks and contagion," *American Economic Review*, 104, 10, 3115–53.

Fragkeskou, Maria and Paparoditis, Efstathios (2018), "Extending the range of validity of the autoregressive (sieve) bootstrap," *Journal of Time Series Analysis*, 39, 3, 356–379.

Gel, Yulia R and Barabanov, Andrey (2007), "Strong consistency of the regularized least-squares estimates of infinite autoregressive models," *Journal of statistical planning and inference*, 137, 4, 1260–1277.

Gombay, Edit (2008), "Change detection in autoregressive time series," *Journal of Multivariate Analysis*, 99, 3, 451–464.

Hamilton, James D (1995), "Time series analysis," *Economic Theory. II, Princeton University Press, USA*, 625–630.

Harshaw, Christopher R, Bridges, Robert A, Iannacone, Michael D, Reed, Joel W, and Goodall, John R (2016),, Graphprints: Towards a graph analytic method for network anomaly detection. in *Proceedings of the 11th Annual Cyber and Information Security Research Conference*, 15. ACM.

Host-Madsen, Anders and Zhang, June (2018), "Coding of Graphs with Application to Graph Anomaly Detection," *arXiv preprint arXiv:1804.02469*.

Inoue, Atsushi (2001), "Testing for distributional change in time series," *Econometric theory*, 17, 1, 156–187.

Keogh, Eamonn, Chu, Selina, Hart, David, and Pazzani, Michael (2001),, An online algorithm for segmenting time series. in *Data Mining, 2001. ICDM 2001, Proceedings IEEE International Conference on*, 289–296. IEEE.

Koutra, Danai, Vogelstein, Joshua T, and Faloutsos, Christos (2013),, Deltacon: A principled massive-graph similarity function. in *Proceedings of the 2013 SIAM International Conference on Data Mining*, 162–170. SIAM.

Kreiss, Jens-Peter (1988), *Asymptotic statistical inference for a class of stochastic processes*, D.Phil. thesis, Verlag nicht ermittelbar.

— (1992), "Bootstrap procedures for AR ()processes," In *Bootstrapping and Related Techniques*, ed. I. *Bootstrapping and Related Techniques*, Springer, pp. 107–113.

Kreiss, Jens-Peter, Paparoditis, Efstathios, Politis, Dimitris N, et al. (2011), "On the range of validity of the autoregressive sieve bootstrap," *The Annals of Statistics*, 39, 4, 2103–2130.

Messer, Michael, Albert, Stefan, and Schneider, Gaby (2018), "The multiple filter test for change point detection in time series," *Metrika*, 81, 6, 589–607.

Milo, Ron, Shen-Orr, Shai, Itzkovitz, Shalev, Kashtan, Nadav, Chklovskii, Dmitri, and Alon, Uri (2002), "Network motifs: simple building blocks of complex networks," *Science*, 298, 5594, 824–827.

Newman, Mark EJ (2003), "The structure and function of complex networks," *SIAM review*, 45, 2, 167–256.

Ozkaya, Yusuf, Sariyuce, A Erdem, Catalyurek, Umit V, and Pinar, Ali (2017), "Active betweenness cardinality: Algorithms and applications," *arXiv preprint arXiv:1711.10634*.

Park, Youngser, Priebe, Carey E, and Youssef, Abdou (2012), "Anomaly detection in time series of graphs using fusion of graph invariants," *IEEE journal of selected topics in signal processing*, 7, 1, 67–75.

Patania, A., Vaccarino, F., and Petri, G. (2017), "Topological analysis of data," *EPJ Data Science*, 6, 7.

Peel, Leto and Clauset, Aaron (2015),, Detecting Change Points in the Large-Scale Structure of Evolving Networks. in *AAAI*, vol. 15, 1–11.

Picard, Dominique (1985), "Testing and estimating change-points in time series," *Advances in applied probability*, 17, 4, 841–867.

Politis, Dimitris N (2015), "Model-Free Prediction in Regression," In *Model-Free Prediction and Regression*, ed. I. *Model-Free Prediction and Regression*, Springer, pp. 57–80.

Pourahmadi, Mohsen (2001), *Foundations of time series analysis and prediction theory*, vol. 379, John Wiley & Sons.

Priebe, Carey E., Conroy, John M., Marchette, David J., and Park, Youngser (2005), "Scan statistics on enron graphs," *Computational & Mathematical Organization Theory*, 11, 3, 229–247.

Rubinov, Mikail and Sporns, Olaf (2010), "Complex network measures of brain connectivity: uses and interpretations," *Neuroimage*, 52, 3, 1059–1069.

Sarkar, Soumajyoti, Guo, Ruocheng, and Shakarian, Paulo (2019), "Using network motifs to characterize temporal network evolution leading to diffusion inhibition," *Social Network Analysis and Mining*, 9, 1, 14.

Shumway, Robert H and Stoffer, David S (2017), *Time series analysis and its applications: with R examples*, Springer.

Sun, Jimeng, Faloutsos, Christos, Papadimitriou, Spiros, and Yu, Philip 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*, 687–696. ACM.

Vandermarliere, Benjamin, Karas, Alexei, Ryckebusch, Jan, and Schoors, Koen (2015), "Beyond the power law: Uncovering stylized facts in interbank networks," *Physica A: Statistical Mechanics and its Applications*, 428, 443–457.

Wager, Tor D, Waugh, Christian E, Lindquist, Martin, Noll, Doug C, Fredrickson, Barbara L, and Taylor, Stephan 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, 3, 821–835.

Wig, Gagan S, Schlaggar, Bradley L, and Petersen, Steven E (2011), "Concepts and principles in the analysis of brain networks," *Annals of the New York Academy of Sciences*, 1224, 1, 126–146.

Wolff, S, Gastwirth, J, and Rubin, Herman (1967), "The effect of autoregressive dependence on a nonparametric test (corresp.)," *IEEE Transactions on Information Theory*, 13, 2, 311–313.

Zellner, Arnold (1962), "An efficient method of estimating seemingly unrelated regressions and tests for aggregation bias," *Journal of the American statistical Association*, 57, 298, 348–368.

Zivot, Eric and Wang, Jiahui (2007), *Modeling financial time series with S-Plus®*, vol. 191, Springer Science & Business Media.

Supplement to "Nonparametric Anomaly Detection on Time Series of Graphs"

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

1 Appendix A: Notation and Theorem

Table 1: Notation

Symbol	Description
$G = (V, E)$	graph/network; V set of vertices, E set of edges
$q; \omega$	window length, threshold parameter
$\boldsymbol{D}$	$T \times P$ multivariate time series; T number of time points, P number of time series
$\boldsymbol{F}$	q-row fold of $\boldsymbol{D}$
$\boldsymbol{R}$; $\boldsymbol{A}$	correlation matrix; adjacency matrix
$CLQNUM$	clique number
MD	mean node degree
ACC	average clustering coefficient
APL	average path length coefficient
BOM	Barnett & Onnela's change detection method
$KVFM$	Koutra et al.'s DELTACON method
$NEDM$	network-evolution detection method
$minLCC$	minimum local clustering coefficient
$maxBETW$	maximum node betweenness centrality

Theorem 1 (Gombay (2008)) *Assume the sequence of observations* $\{Y_t\}$ *satisfy equation (1.1) with Gaussian i.i.d. white noise* $\{\varepsilon_t\}$, $var(\varepsilon_t) = \sigma^2$, *and characteristic polynomial*

$\varphi(z) = 1 - \varphi_1 z - ... - \varphi_p z^p$ *has roots outside the unit circle. Then there exists a* $(p+2)$ *- dimensional Gaussian process* $\mathbf{B}(u)$ *with independent Brownian bridge components* $B^{(r)}(u), j = 1, ..., p+2$, *such that :*

$$\max_{1 \le j \le p+2} \sup_{0 \le u \le 1} |\hat{B}^{(r)}(u) - B^{(r)}(u)| = o_p(1). \tag{1.11}$$

2 Appendix B: Experimental data analysis

2.1 Simulation study: Size & Power of the $NEDM$ with $P = 50$

We performed Monte Carlo simulations (5000 iterations) under a no-change point (change in mean) scenario and a one-change point (change in variance) scenario with $P = 50$ using the same parameters stated in Pages 13-14 described under Section 4 of the manuscript. The results are presented in Figures 1 and 2, respectively.

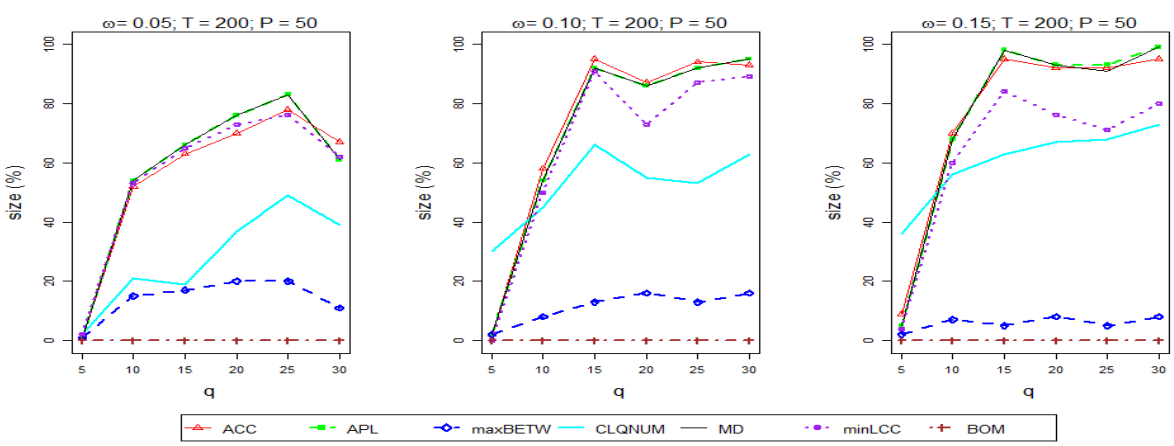

Figure 1: The size of the test for the no-change (change in mean) scenario with $T = 200$, $\omega = \{0.05, 0.1, 0.15\}$, $q = \{5, 10, 15, 20, 25, 30\}$ and $P = 50$.

In Figure 1, we find that as the window length (q) and the threshold ω increase the size of the test reported by all network summary statistics (with the exception of $maxBETW$) under $NEDM$ also increase, leading to more liberal results. With $maxBETW$, however, we find that as both q and ω increase, the size of the test steadily approaches the declared

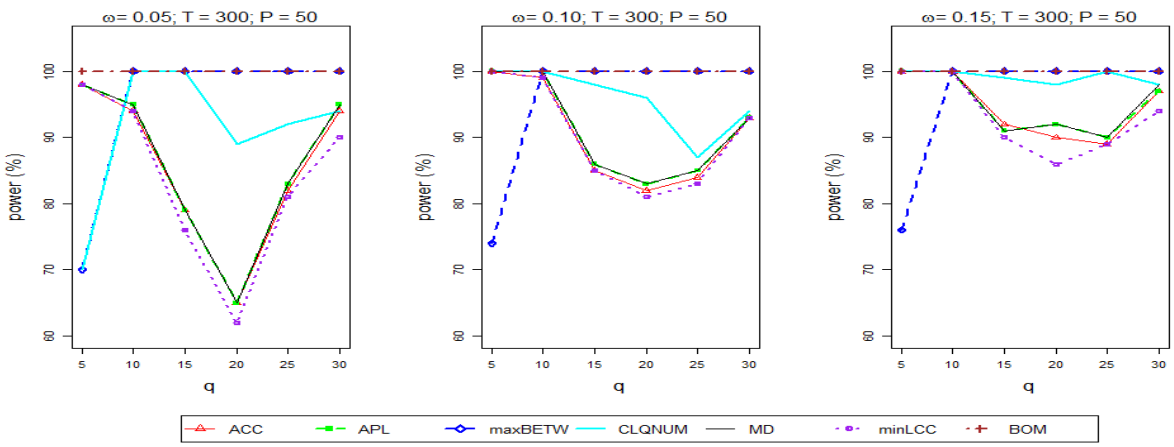

Figure 2: The power of the test for the one-change (change in variance) scenario with $T = 300$, $\omega = \{0.05, 0.1, 0.15\}$, $q = \{5, 10, 15, 20, 25, 30\}$, $\tau = 157$ and $P = 50$.

5% level of significance. In turn, BOM appears to be very conservative, delivering a size of the test close to 0% for all q and ω. This is in contrast to our analysis in the manuscript with $P = 5$. Here we found that the size of the test under BOM tends to be more liberal (ranging between 18% to 20%).

Figure 2 shows that, for $P = 50$, $NEDM$ with the $maxBETW$ network summary performs as well as BOM in terms of power (*i.e.,* it correctly flags the one-change scenario every time) except for one combination. This is very similar to the results for $P = 5$. In summary, we can conclude that both the power and the size of the test tend to be sensitive to the choice of parameters (q, ω and T) and that the performance of the $NEDM$ also depends on the choice of network summary statistic which in turn is likely to be driven by the underlying network topology.

2.2 Case study: Anxiety fMRI data

Figure 3 presents additional results for the Anxiety fMRI data based on the $NEDM$ with $CLQNUM$ and MD as the network summary statistics. To this end we find that the $NEDM$ is able to detect 3 change locations with the MD summary, and 1 change location with the $CLQNUM$ network summary. Table 2 presents a summary of the numerical results obtained based on this data.

Results for Anxiety fMRI data.

Table 2: Summary of change points detected for the Anxiety fMRI data.

Subject	ACC	APL	$maxBETW$	$CLQNUM$	MD	$minLCC$	BOM	$KVFM$
1	94	62,93				94		92
2						72		111,112,120
3	138	134	134	72	134	134		92,94
4					115			69,70,76,117
5	63	63,70	68,131	63	64	64		108
6			106					90
7	63						61	
8		109					102	77,78
10	99	126					67	
11								92,93
12		99					116	75,90,94
13		65,66			71			62,65
14	106	104		97				88,111
15								65,66,98
16	140						92	
17		77	97			97,104		
18					130			
19			93					
20	65				70		100	61,62,86

* Subjects not included have no change points detected across all methods.

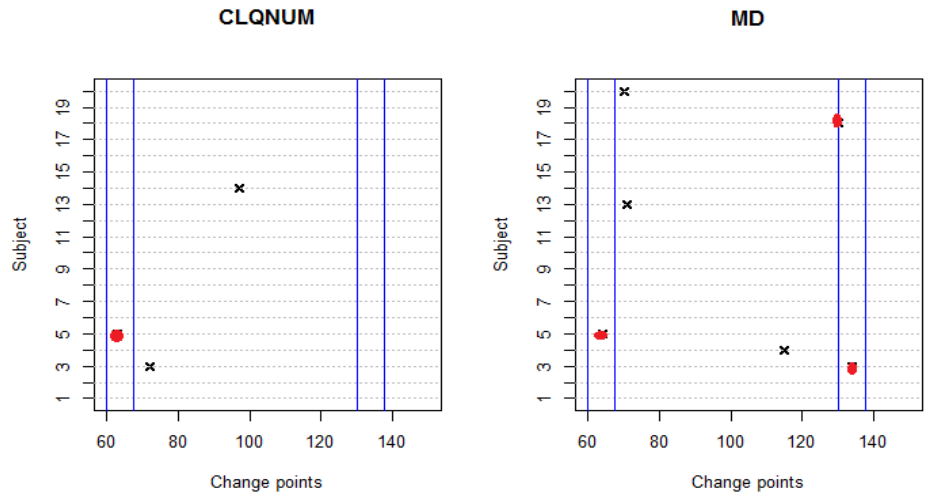

Figure 3: The change points detected for the Anxiety fMRI data using the $NEDM$ and the $CLQNUM$ and the MD summary statistics.

2.3 Case study: Enron email network using the $KVFM$

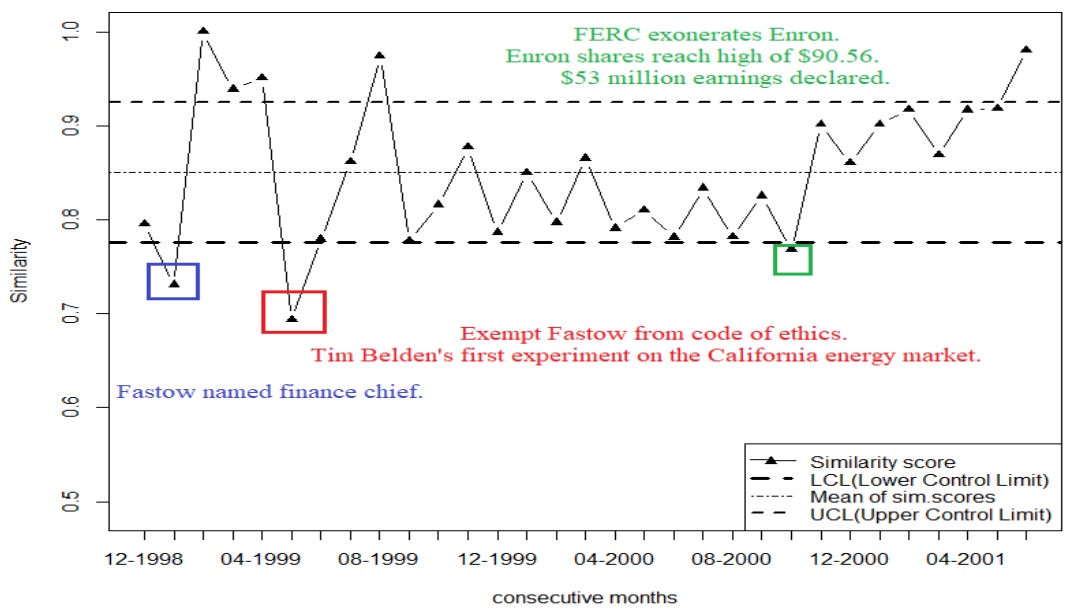

Figure 4: The detected change points for the Enron email data set from November 1998 – July 2001 using the $KVFM$.

2.4 Case study: resting state fMRI data

We apply the $NEDM$, with all network summary statistics except the $CLQNUM$, to a resting-state fMRI data as described in Cribben and Yu (2017). The $CLQNUM$ summary is omitted due to computational challenges encountered during the analysis. Participants ($n = 45$) are instructed to rest in the scanner for 9.5 minutes, with the instruction to keep their eyes open for the duration of the scan. We apply the Anatomical Automatic Labeling (Tzourio-Mazoyer et al., 2002) atlas to the adjusted voxel-wise time series and produce time series for 116 Regions of Interest (ROIs) for each subject by averaging the voxel time series within the ROIs. In total, each time series contained 285 time points (9.5 minutes with $TR = 2\ s$).

Results

Results obtained in this case study (as seen in Figure 5) show significant change point locations for all 45 subjects. As resting state data is unconstrained, we do not know where the true network change points occur. However, the $NEDM$ finds significant change points in the network structure for many subjects with the maximum number of change points being 4 which lines up with the results in Cribben and Yu (2017). From Figure 5, we find a high indication that change points vary highly across network summaries and that some change states endure longer periods while others transition more quickly. On the other hand, results by the BOM and by the $KVFM$ appear to suggest the detection of a much larger number of significant change points for each subject. This is expected for the $KVFM$ given that it does not provide a quanitifier for statistical significance. Furthermore, judging from the close proximity and the total number of change points reported by these two methodologies, we deduce that some of the change points for the subjects are unrealistic for resting state fMRI data. As noted above, having change points so close to one another makes it impossible to estimate the network structure between each pair of significant change points.

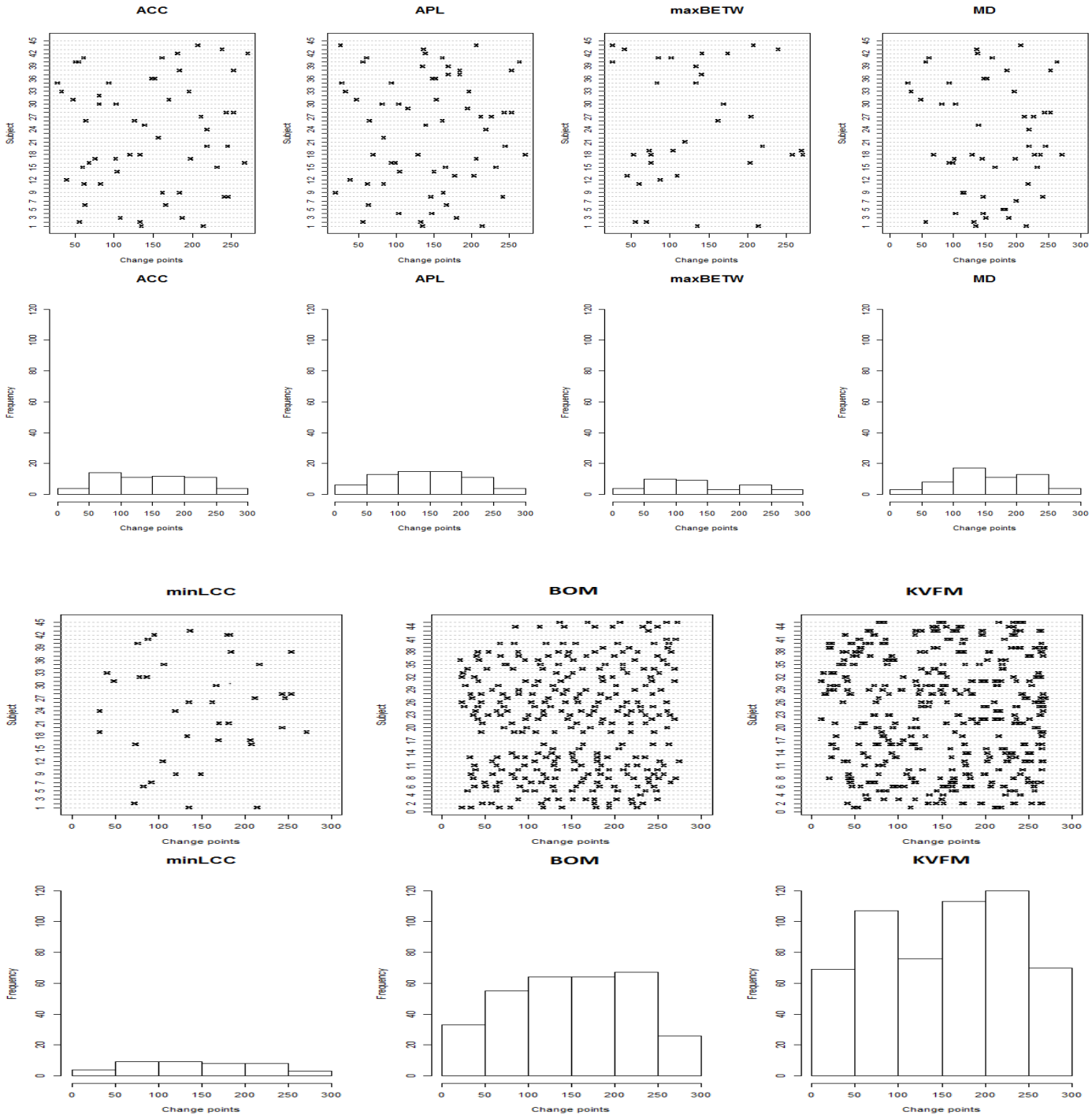

Figure 5: The detected change points for the resting state fMRI data set.

3 Appendix C: Proofs for Theorem 1 and Corollary 1

Theorem 1

Proof. Note that the assumptions on Y_t imply that it satisfies assumptions $A1'$, $A2$, and B of Bühlmann et al. (1997).

From Equation (1.2), we know that

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

Hence, each individual component of $B^{(r)}(u)$ and its bootstrap counterpart $B^{(r)*}(u)$ also satisfy the assumption C of Bühlmann et al. (1997). As a result, by invoking Theorem 3.3 of Bühlmann et al. (1997) (as $n \to \infty$), we obtain

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

□

Corollary 1

Proof. Note that s_n^2 is a consistent estimator of $E(\varepsilon_t^2) = \sigma^2$ (Lehmann and Casella, 2006). Hence, given that $\widetilde{\varepsilon^{\diamond}}_t$ are independent random variables drawn from $N(0, s_n^2)$,

$$\{\widetilde{\varepsilon_t^{\diamond}}\}_{t=1}^n \xrightarrow{D} \{\widetilde{\varepsilon}_t\}_{t=1}^n.$$

As a result,

$$X_t^{\diamond} \xrightarrow{d^*} X_t \ \ in \ \ probability, \tag{1.16}$$

and in view of the proof for Theorem 1

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

□

4 Appendix D: Asymptotic and Sieve bootstrap distribution simulations

The procedure for assessing the performance of the sieve bootstrap technique to the asymptotic distribution is set up as follows:

1. Simulate $5,000$ Monte Carlo simulations of an $AR(1)$ time series with different samples ($n = 100,\ 200,\ 300$) and change point times ($\tau = 50,\ 100,\ 180$).

2. Consider the following scenario for identifying change points in the $AR(1)$ time series: change in mean (μ). We use $\Gamma = 5000$ replications for the sieve bootstrap procedure (as expressed in Algorithm 1).

3. With equation (1.1) as reference, the $AR(1)$ model is defined as

$$Y_t - \mu = \varphi(Y_{t-1} - \mu) + \varepsilon_t; \quad t \in [1, n]; \quad \varepsilon_t \sim \text{Gaussian } WN(0, \sigma^2).$$

4. For each sample n, with σ^2 and φ as nuisance parameters, we test the following:

$$\begin{aligned} H_0 :\ & \mu = \mu_0, \forall\ t \in [1, n] \quad \text{against} \\ H_a :\ & \mu = \mu_0,\ \forall\ 1 \leqslant t \leqslant \tau - 1,\ at\ \tau = [\rho n],\ \ \rho \in (0, 1)\ ; \\ & \mu = \mu_a,\ \ \forall\, \tau \geqslant n. \end{aligned}$$

5. With $\sigma^2 = 1$ and $\mu_0 = 0$, we use the values $\varphi = (-0.5,\ 0.1,\ 0.5)$.

References

Bühlmann, P. et al. (1997), "Sieve bootstrap for time series," *Bernoulli*, 3, 123–148.

Cribben, I., and Yu, Y. (2017), "Estimating whole-brain dynamics by using spectral clustering," *Journal of the Royal Statistical Society: Series C (Applied Statistics)*, 66, 607–627.

Gombay, E. (2008), "Change detection in autoregressive time series," *Journal of Multivariate Analysis*, 99, 451–464.

Lehmann, E. L., and Casella, G. (2006), *Theory of point estimation*, Springer Science & Business Media.

Tzourio-Mazoyer, N., Landeau, B., Papathanassiou, D., Crivello, F., Etard, O., Delcroix, N., Mazoyer, B., and Joliot, M. (2002), "Automated anatomical labeling of activations in SPM using a macroscopic anatomical parcellation of the MNI MRI single-subject brain," *Neuroimage*, 15, 273–289.