Sequential Graph Scanning Statistic for Change-point Detection

Xi He, Yao Xie
School of Industrial and Systems Engineering
Georgia Institute of Technology, Georgia 30318
xi.he@gatech.edu, yao.xie@isye.gatech.edu

Sin-Mei Wu, Fan-Chi Lin
Department of Geology and Geophysics
University of Utah, Utah 84112
sinmei.wu, fanchi.lin@utah.edu

***Abstract*—Graph change-point detection problems have wide applications in graphical data types, such as social networks and sensor networks. Given a sequence of random graphs with fixed vertices and changing edges, we are interested in detecting a change that causes a shift in the distribution of a subgraph. We present two graph scanning statistics that can detect local changes in the distribution of edges in a subset of the graph. The first statistic assumes a parametric model, i.e., the observations on the edges are Gaussian random variables, and the change shifts the mean of a subgraph. We derive the scan statistic and present a theoretical approximation to the false alarm rate, which is verified to be accuracy numerically. The second statistic adopts a nonparametric approach based on k-Nearest Neighbors (k-NN). We demonstrate the efficiency of our detection statistics for ambient noise imaging, using a real dataset records real-time seismic signals around the Old Faithful Geyser in the Yellowstone National Park.**

I. Introduction

Change-point detection is a fundamental problem in social networks [1], sensor networks, and power networks. In this paper, we use graph scanning techniques [2], [3] to study the question of how to detect a change in the distribution of the graphs. In particular, we are interested in detecting a *local* change in the graph. This means, when the change happens, only a subset of the graph, or a subgraph, of known size is affected by the change and has a different distribution. The observed change in distribution for the graphs are caused by a local change, while the distribution for the rest of the graph remains the same. The problem of local change-point detection is challenging in that, first, we do not know whether there is a change, and second, if there is a change at some unknown time, it is not clear which subgraph contains the change.

A motivating application of our study is monitoring ambient noises in seismic sensor networks. In ambient noise imaging, because the signals are weak, it is difficult to observe any signal using observations from a single sensor. Fortunately, when we construct the pair-wise cross-correlation between the sensors, there will be coherent signals between affected sensors who observe changes in the subsurface structures. Specifically, at the time of the change, the cross-correlation function between the sensors affected by the change will have a significant peak. Between the affected sensors and the unaffected sensors, and among the unaffected sensors, such a waveform of the cross-correlation function does not exist. Therefore, this problem, mathematically, becomes detecting a local change in a sequence of graphs.

In this paper, we present two approaches for constructing *scan statistics* to detect a local change in a sequence of graphs, the parametric and the non-parametric approach. For the parametric approach, we assume Gaussian graphs and apply a scan statistic based on counting the maximum number of edges in a subgraph of fixed size. We derive an accurate theoretical approximation to the false alarm rate of the scan statistic based on selective inference [4], which can be used to set the threshold for the false alarm rate without large scale simulation. For the non-parametric approach, the scan statistic is constructed using similarity measures on the subgraphs and k-Nearest Neighbors (k-NN). We demonstrate the efficiency of the non-parametric approach on real data for the seismic sensor network in Yellowstone [5].

This paper is related to works in change-point detection, graph scan statistics, and community detection. Graph scan statistic for the stochastic block model, which counts the maximum number of edges in the subgraphs of an Erods-Renyi graph, has been considered in [6]. A likelihood ratio test for detecting communities in the Erdos-Renyi graph is studied in [7]. A non-parametric graph scan statistic based on k-NN is discussed in [8] and [9].

II. Problem Formulation

Suppose we observe a sequence of undirected graphs $G_1, \ldots, G_N$, where N is the time horizon. For $t = 1, \ldots, N$, let $G_t = \{V, E_t\}$, with V and E_t being the set of vertices and the set of edges respectively. Let V^i be a size-m subset of the nodes V, $i = 1, \ldots, d$, where $d = \binom{N}{m}$ if all possible subsets are considered. In networks, usually $d \ll \binom{N}{m}$. Let $S^i = \{V^i, E^i\}$ be the subgraph containing V^i and the edges between them, which change over time. Denote $\mathcal{S}$ as the set of all possible subgraphs, then $\mathcal{S} = \{S^i, \ldots, S^d\}$. Assume a

change-point happening at an unknown time τ and the change is contained in the graph $S^* = \{V^*, E^*\}$, such that before and after τ, the distribution of the edges in E^* are different. At time t, denote $S^i(t) = \{V^i, E^i_t\} \subset G_t$. When there is a change, we assume $E^*_1, \ldots, E^*_{\tau-1}$ are i.i.d. distributed according to some distribution P, and $E^*_\tau \ldots, E^*_T$ are i.i.d. distributed according to another distribution Q. The problem of detecting a local change becomes the following hypothesis testing problem.

$$\begin{array}{lll} H_0: & E^i_t \sim P, & t = 1, \ldots, N, \ \forall\ S^i \in \mathcal{S}; \\ H_1: & E^i_t \sim Q, & t \geq \tau, \ S^i = S^*, \\ & E^i_t \sim P, & \text{otherwise.} \end{array} \tag{1}$$

E^i_t is also the adjacency matrix of the subgraph S^i at time t. The hypothesis testing problem is illustrated in Fig. 1.

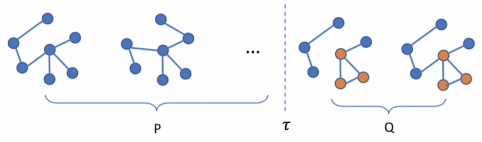

Fig. 1: Graphs prior to the change-point in time τ follow the distribution P, and graphs after time τ follow the distribution Q. We are particularly interested in detecting the local change in a subgraph (showed in highlight).

Assuming that the change happens at τ, at each time t, for each subgraph S^i, we form a test statistic $R(t, \tau, S^i)$. The change is detected when the test statistic exceeds a given threshold γ. Let w be a small sliding window, the test scheme can be formulated as

$$T = \inf\{t : \max_{t-w<\tau<t} \max_{S^i \in \mathcal{S}} R(t, \tau, S^i) > \gamma\}. \tag{2}$$

We are further interested in knowing which subgraph causes the change in the graph structure. The test statistic $R(t, \tau, S^i)$ is useful in localizing the change, as the subgraph S^* that maximizes $R(t, \tau, S^i)$ is the subgraph containing the change,

$$S^* = \arg\max_{S^i \in \mathcal{S}} R(t, \tau, S^i).$$

We present two possible approaches to this problem based on *scan statistic* in the next sections, a parametric approach and a non-parametric approach. Moreover, we will study real data for this problem in the numerical example section.

III. Parametric Approach

First, we consider a parametric approach to form the scan statistic $R(t, \tau, S^i)$ in (2) by introducing a probability model to the sequence of graphs. In particular, we assume that the entries of the adjacency matrices are Gaussian random variables. Before the change, the edges have smaller means (e.g., zero mean) to represent that there is no significant correlation between the sensors. After the change, a subset of the nodes, i.e. sensors containing the change, will have higher means on the edges between them. For any subgraph $S^i \in \mathcal{S}$, at time t, let $W_{u,v}(t)$ denote the probability of the edge formation between the vertices u and v, where $u, v \in V^i$, then $E^i_t = \{W_{u,v}(t) : u, v \in V^i\}$. In this case, in the hypothesis testing problem (1), P represents $\mathcal{N}(\mu_0, \sigma_0^2)$, and Q represents $\mathcal{N}(\mu_1, \sigma_0^2)$, where μ_0, μ_1, σ_0^2 are constants, and $\mu_1 > \mu_0$. We can re-write (1) as

$$\begin{array}{lll} H_0: & W_{u,v}(t) \sim \mathcal{N}(\mu_0, \sigma_0^2), & t = 1, \ldots, N, \ \forall\ u, v \in V; \\ H_1: & W_{u,v}(t) \sim \mathcal{N}(\mu_1, \sigma_0^2), & t \geq \tau, \ \mu_1 > \mu_0, \ u, v \in S^*, \\ & W_{u,v}(t) \sim \mathcal{N}(\mu_0, \sigma_0^2), & \text{otherwise.} \end{array}$$

In this section, we first set aside the time dimension and focus on detecting the subgraph S^* affected by the change. Once we formulate the subgraph detection scheme, we can repeatedly apply the test to the sequence of graphs as a Shewhart chart procedure.

Now we present the construction of the scan statistic in the parametric setting. Let x_i denote the number of edges in a subgraph S^i with m vertices. Then x_i follows a Gaussian distribution with mean μ_{x_i} and covariance Σ_{x_i}.

$$x_i = \sum_{u,v \in S^i} W_{u,v} \sim \mathcal{N}(\mu_i, \Sigma_i).$$

Under the null hypothesis,

$$\mu_i = \frac{m(m-1)}{2}\mu_0, \quad \Sigma_i = \frac{m(m-1)}{2}\sigma_0^2.$$

A change is detected when the maximum number of edges in a subgraph exceeds a pre-specified threshold γ, i.e.

$$\max_{S^i \in \mathcal{S}} x_i > \gamma.$$

We estimate the false alarm rate: $\mathbb{P}_0\{\max_{S^i \in \mathcal{S}} x_i > \gamma\}$. Recall $|\mathcal{S}| = d$. So the false alarm rate can also be written as

$$\mathbb{P}_0\{\max_{i=1,\ldots,d} x_i > \gamma\}. \tag{3}$$

A. Determining the Threshold by Controlling the False-alarm

We observe that (3) is the tail probability of the maximum of a series of correlated Gaussian random variables. In this section, we transform the false alarm rate formula using Bayes rule, and then apply the idea of selective inference [4] to estimate the probability.

Notice that we can decompose the event in (3) as the union of polyhedrons:

$$\begin{aligned} \left\{\max_{i=1,\ldots,d} x_i > \gamma\right\} &= \bigcup_{i=1,\ldots,d} \{x_i > \gamma, x_i \geq x_j, j \neq i\} \\ &\triangleq \bigcup_{i=1,\ldots,d} \{A_i \mathbf{x} \geq \mathbf{b}\}, \end{aligned}$$

where $\mathbf{x} = [x_1, \ldots, x_d]^N \in \mathbb{R}^d$, $\mathbf{b} = [\gamma, 0, \ldots, 0]^N \in \mathbb{R}^d$,

and $A_i = AP_i$. Here,

$$A = \begin{pmatrix} 1 & 0 & 0 & \cdots & 0 \\ 1 & -1 & 0 & \cdots & 0 \\ 1 & 0 & -1 & \cdots & 0 \\ \vdots & & & \ddots & \\ 1 & 0 & 0 & \cdots & -1 \end{pmatrix} \in \mathbb{R}^{d\times d},$$

and P_i is the permutation matrix swapping the 1^{st} and the i^{th} entry of $\mathbf{x}$. Similar decomposition appears in [10]. Thus,

$$\mathbb{P}_0\{\max_{i=1,\dots,d} x_i > \gamma\} = \frac{\beta}{\alpha}, \tag{4}$$

where

$$\begin{aligned}\alpha &= \mathbb{P}_0\{x_1 > \gamma | \max_{i=1,\dots,d} x_i > \gamma\} \\ &= \mathbb{P}_0\Big\{x_1 > \gamma | \bigcup_{i=1,\dots,d} \{-A_i\mathbf{x} \le -\mathbf{b}\}\Big\}, \\ \beta &= \mathbb{P}_0\{x_1 > \gamma\} \\ &= 1 - \Phi\Big(\gamma; \frac{m(m-1)}{2}\mu_0, \frac{m(m-1)}{2}\sigma_0^2\Big),\end{aligned}$$

where Φ is the CDF of the standard normal distribution, and α can be evaluated using selective inference as Theorem 5.3 in [4]. Our result is summarized in Lemma 1.

Lemma 1. *Let F^B_{μ,σ^2} denote the CDF of a normal random variable with mean μ and variance σ^2 truncated to the set B, and let $\mathbf{x} \sim N(\boldsymbol{\mu}, \Sigma)$. Then*

$$\begin{aligned}\boldsymbol{\mu} &= \frac{m(m-1)}{2}\mu_0 \mathbb{1}_d, \\ \Sigma_{(i,i)} &= \frac{m(m-1)}{2}\sigma_0^2, \\ \Sigma_{(i,i')} &= \frac{l_{i,i'}(l_{i,i'}-1)}{2}\sigma_0^2, \ i \ne i',\end{aligned}$$

where $\mathbb{1}_d$ is the d-dimensional vector of all 1's, and $l_{i,i'}$ is the number of overlapping nodes between two subgraphs S^i and $S^{i'}$. Then we have the following conclusion.

$$F^{\bigcup_i [\mathcal{V}_i^-(\mathbf{z}), \mathcal{V}_i^+(\mathbf{z})]}_{\boldsymbol{\eta}^N\boldsymbol{\mu}, \boldsymbol{\eta}^N\Sigma\boldsymbol{\eta}}(\boldsymbol{\eta}^N\mathbf{x}) | \bigcup_{i=1,\dots,d} \{-A_i\mathbf{x} \le -\mathbf{b}\} \sim \textit{Unif}(0,1)$$

with the specification $\boldsymbol{\eta} = [1, 0, \dots, 0]^N \in \mathbb{R}^d$, and the set boundaries

$$\begin{aligned}\mathcal{V}_i^-(\mathbf{z}) &\equiv \max_{j:(A_i\mathbf{c})_j > 0} \frac{b_j - (A_i\mathbf{z})_j}{(A_i\mathbf{c})_j}, \\ \mathcal{V}_i^+(\mathbf{z}) &\equiv \min_{j:(A_i\mathbf{c})_j < 0} \frac{b_j - (A_i\mathbf{z})_j}{(A_i\mathbf{c})_j},\end{aligned}$$

where

$$\boldsymbol{c} \equiv \Sigma\boldsymbol{\eta}(\boldsymbol{\eta}^N\Sigma\boldsymbol{\eta})^{-1} = \Sigma\boldsymbol{\eta}\Sigma_{1,1}^{-1} = a\Sigma_{(:,1)},$$

$$\begin{aligned}\mathbf{z} &\equiv (\boldsymbol{I}_d - \boldsymbol{c}\boldsymbol{\eta}^N)\mathbf{x} = \mathbf{x} - \boldsymbol{c}\boldsymbol{\eta}^N\mathbf{x} = \mathbf{x} - a\Sigma_{(:,1)}x_1, \\ a &= \frac{2}{k(k-1)\sigma_0^2}.\end{aligned}$$

For $i \ne i'$,

$$\begin{aligned}A_i\mathbf{c} &= AP_i\mathbf{c} = a\begin{pmatrix}\Sigma_{(i,1)} \\ \Sigma_{(i,1)} - \Sigma_{(i',1)}\end{pmatrix}, \\ A_i\mathbf{z} &= AP_i\mathbf{z} = \begin{pmatrix} x_i - a\Sigma_{(i,1)}x_i \\ \big(x_i - a\Sigma_{(i,1)}x_i\big) - \big(x_{i'} - a\Sigma_{(i',1)}x_i\big)\end{pmatrix}, \\ \mathbf{b} - A_i\mathbf{z} &= \begin{pmatrix} \gamma - \big(x_i - a\Sigma_{(i,1)}x_i\big) \\ \big(x_{i'} - a\Sigma_{(i',1)}x_i\big) - \big(x_i - a\Sigma_{(i,1)}x_i\big)\end{pmatrix}.\end{aligned}$$

Therefore, we can estimate the false alarm rate using Lemma 1 and (4) and set the threshold γ accordingly. The performance of the estimation is presented in the next section.

B. Verification of Numerical Accuracy

In this section, we conduct a numerical experiment to verify the numerical accuracy of our estimation of the false alarm rate. Assuming standard normal distribution under the null hypothesis, we generate α according to Lemma 1 and compute the false alarm rate based on (4). The resulting false alarm rate curve by changing the threshold γ is plotted in Fig. 2. The result is based on 500 experiments, and the standard error, which is small, is shown as the shaded area in the plot.

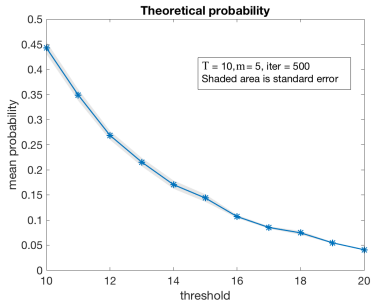

Fig. 2: Simulated false alarm rate for the detection statistic by (4).

We also compare the theoretically estimated γ from using formula (4) with the simulated γ in Table I. The two γ's are quite close in this case, showing good approximation of the theoretical result.

Probability	Theory γ	Simulated γ
0.2	13.34	14.43
0.15	14.77	14.68
0.1	16.00	15.93

TABLE I: $N = 50, m = 5, d = 2,118,760$

IV. Non-parametric Approach based on Similarity

In this section, we describe a non-parametric detection statistic based on the similarity measure between subgraphs at different time. The idea is to compare the subgraphs formed

with the same set of nodes occurring before and after time t to check for their graph structure similarity. If the graph structures are similar, they are likely from the same distribution, and if if the dissimilarity is large enough, we declare a change-point at t. For $i = 1, \ldots, d$, at time $t = 1, \ldots, N$, we check the similarity between $S^i(1), \ldots, S^i(t-1)$ and $S^i(t), \ldots, S^i(N)$. For simplicity, denote an arbitrary subgraph S^i as S in the rest of the analysis.

We use a k-NN based change-point detection statistic introduced by [8], [9], which is developed from the two-sample test statistic in [11] and [12]. At time t, the sequence of subgraphs can be divided into two groups, those happening before t: $S(1), \ldots, S(t-1)$, and those at or after t: $S(t), \ldots, S(N)$. We use $b_{n,n'}(t)$, $1 \le n, n' \le N$, to indicate whether the subgraphs at time n and n' are in the same group or not when the separator is at time t.

$$b_{n,n'}(t) = \mathbf{I}\Big((n \le t, n' > t) \text{ or } (n > t, n' \le t)\Big),$$

where $\mathbf{I}(\cdot)$ is the indicator function. Then $b_{n,n'}(t) = 1$ if at time t, $S(n)$ and $S(n')$ belong to different groups. Define the similarity between $S(n)$ and $S(n')$ as $d_{n,n'}$. Assume the neighbors are unique. Let

$$A^{(r)}_{n,n'}(S(n')) = \mathbf{I}(S(n') \text{ is the } r\text{th nearest neighbor of } S(n)),$$

$$A^{+}_{n,n'}(S(n')) = \sum_{r=1}^{k} A^{(r)}_{n,n'}(S(n')),$$

Then $A^{+}_{n,n'}(S(n')) = 1$ means that $S(n')$ is among the k-NN of $S(n)$. Define the test statistic as

$$R(t, \tau, S) = \sum_{n=1}^{N} \sum_{n'=1}^{N} (A^{+}_{n,n'} S(n') + A^{+}_{n',n} S(n)) b_{n,n'}(t).$$

We now treat the subgraphs $S(t)$'s as nodes, and form a k-NN graph in which an edge forms between any nodes and their closest k neighbors. Then $R(t, \tau, S)$ represents twice the number of edges connecting nodes in the two groups separated by t. H_0 is rejected when $R(t, \tau, S)$ is significantly *smaller* than its expectation under the permutation null distribution. When $R(t, \tau, S)$ is small, it means that the number of edges connecting the two groups in the k-NN graph is small, and the two samples are likely from different distributions. If $R(t, \tau, S)$ is large, it implies that the samples are well-mixed and are likely to be from the same distribution.

It is shown in [11] and [12] that the standardized test statistic

$$\frac{R(t, \tau, S) - \mathbb{E}[R(t, \tau, S)]}{\sqrt{Var(R(t, \tau, S)}}$$

converges to the standard normal distribution under H_0 when $\frac{t}{N-t} \to \lambda \in (0, \infty)$ for multivariate data. The mean and variance for the statistics are

$$\mathbb{E}[R(t, \tau, S)] = \frac{4kt(N-t)}{N-1},$$

$$Var(R(t, \tau, S)) = \frac{4kt(N-t)}{N-1}\Big(h\big(t, (N-t)\big)$$

$$\big(\frac{1}{N}\sum_{n,n'=1}^{N} A^{+}_{n,n'} A^{+}_{n',n} + k - \frac{2k^2}{N-1}\big)$$

$$+ \big(1 - h(t, N-t)\big) + \big(\frac{1}{N}\sum_{n,n',n''=1}^{N} A^{+}_{n',n} A^{+}_{n'',n} - k^2\big)\Big),$$

where $h(t, N-t) = \frac{4(t-1)(N-t-1)}{(N-2)(N-3)}$.

Define the test statistic

$$R'(t, \tau, S) = -\frac{R(t, \tau, S) - \mathbb{E}[R(t, \tau, S)]}{\sqrt{Var(R(t, \tau, S)}}.$$

Suppose the change occurs at time τ, then $R'(t, \tau, S)$ will be large when t is close to τ (note the negative sign in the standardization). The testing procedure can be written as

$$T(t, \tau, S) = \inf\{t : \max_{S^i \in \mathcal{S}} \max_{n_0 \le t \le N - n_0} R'(t, \tau, S^i) > \gamma\}, \quad (5)$$

where $1 < n_0 < N$.

V. Real-data example

In this section, we demonstrate how the proposed detection statistics could be used in solving the local change-point detection problem in a seismic sensor networks using real data. We first check whether there is a change in the graphs, and then narrow down the change to a subgraph. For simplicity, we only apply the nonparametric approach.

A. Description of the Dataset

The seismic sensor network that we study is illustrated in Fig. 3. It shows the physical location of the sensors measuring signals around the Old Faithful Geyser in the Yellowstone National Park. There are 18 sensors in the network, and edge information is contained in the pair-wise cross-correlation function between the sensors. The cross-correlation function is then transformed to a value called peak lag time, which is shown on the y-axis in Fig. 3. We observe a sequence of 101 graphs on this network over time, one at each "stage", ranging from stage -50 to 50 (shown as the x-axis). The nodes, or sensors, in the networks remain the same, while the edge value fluctuates as the peak lag time among the sensors changes. At stage 0, the geyser erupts, and the distribution of the peak lag time among the sensors affected by the eruption changes. Our goal is to detect the change in the sequence of the graphs at stage 0 and find the sensors responsible for the change. We have data on 11 stations: 001, 002, 003, 005, 006, 008, 009, 010, 014, 015, 016, and the peak lag time on 10 pairs of the stations. For the other 45 pairs without data, we assume that no edge forms between the sensors.

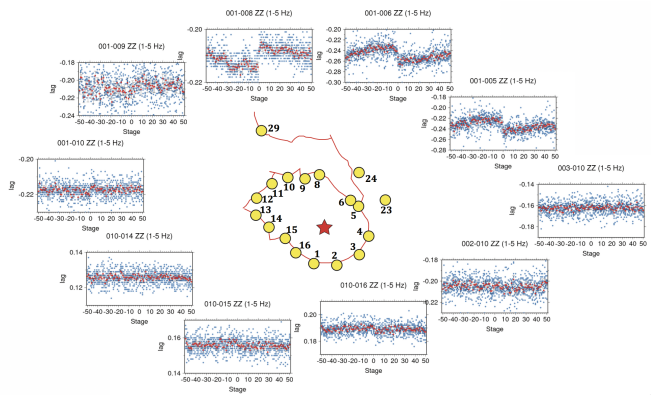

Fig. 3: Peak lag time in the seismic sensor network which measures the geyser activity in the Yellowstone National Park.

B. Change-point Detection

First we detect whether there is a change-point in the sequence of graphs. Two types of graphs are considered, the unweighted graph and the weighted graph.

1) Unweighted Graph: Denote the *mean* peak lag time (red points in Fig. 3) of a pair of sensors u, v at time t as $y_{u,v}(t)$. Assume that an edge forms between u, v at time t if $y_{u,v}(t)$ is greater than the average $\bar{y}_{u,v}$, that is, $y_{u,v}(t) > \bar{y}_{u,v}$, where $\bar{y}_{u,v} = \frac{1}{101}\sum_{t=1}^{101} y_{u,v}(t)$. We use the Weisfeiler-Lehman edge graph kernel [13] to measure the closeness of the graphs and find the k-NN as described in the non-parametric section. The test statistic $-R(t, \tau, S)$ is plotted in Fig. 4, and it peaks at stage 0, corresponding to the true change-point.

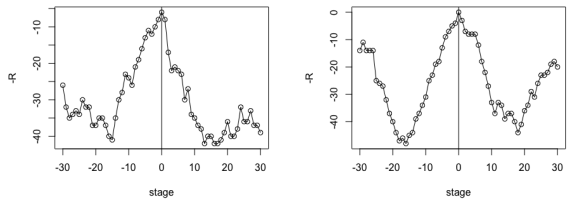

Fig. 4: Test statistic for: (Left) unweighted graph, (Right) weighted graph.

2) Weighted Graph: To construct weighted graphs, at each time t, we use the peak lag time between the two stations u, v as the "weight" on the edge between the nodes. The test statistic $-R(t, \tau, S)$ is plotted in Fig. 4. Comparing with the previous experiment on unweighted graphs, we find that although both methods successfully identifies the change-point at stage 0, there are also two other local maxima for the weighted graph, which may interfere with the detection.

C. Determining the Change Location

We are further interested in finding the location within the graph where the change happens. In other words, we identify a subset of m nodes that contribute to the overall change in the graphs. Ideally, the data on those nodes would be sufficient for the overall change detection. In this example, we assume $m = 3$ by observing Fig. 3. We have data for 11 nodes, and therefore $\binom{11}{3} = 165$ possible subsets of nodes. However, recall that only 10 edges are available. So in reality, only 56 subsets are considered. Given each subset of nodes, we preserve the edge information among the 3 nodes, and set the weight on other edges to 0. For each subset, we repeat the steps in the last example on weighted graphs as if the graphs only contain 3 nodes in the subset. Following the testing procedure in (5), we find that the subgraph maximizing the test statistic is formed by nodes 001, 008, and 009.

ACKNOWLEDGEMENT

This work was partially supported by National Science Foundation (NSF) CCF-1442635.

REFERENCES

[1] Y. Wang, A. Chakrabarti, D. Sivakoff, and S. Parthasarathy, "Fast change point detection on dynamic social networks," *arXiv preprint arXiv:1705.07325*, 2017.

[2] C. E. Priebe, J. M. Conroy, D. J. Marchette, and Y. Park, "Scan statistics on enron graphs," *Computational & Mathematical Organization Theory*, vol. 11, no. 3, pp. 229–247, 2005.

[3] J. Sharpnack, A. Rinaldo, and A. Singh, "Detecting anomalous activity on networks with the graph fourier scan statistic," *IEEE Transactions on Signal Processing*, vol. 64, no. 2, pp. 364–379, 2016.

[4] J. D. Lee, D. L. Sun, Y. Sun, J. E. Taylor, *et al.*, "Exact post-selection inference, with application to the lasso," *The Annals of Statistics*, vol. 44, no. 3, pp. 907–927, 2016.

[5] S.-M. Wu, K. M. Ward, J. Farrell, F.-C. Lin, M. Karplus, and R. B. Smith, "Anatomy of old faithful from subsurface seismic imaging of the yellowstone upper geyser basin," *Geophysical Research Letters*, vol. 44, no. 20, 2017.

[6] N. Verzelen, E. Arias-Castro, *et al.*, "Community detection in sparse random networks," *The Annals of Applied Probability*, vol. 25, no. 6, pp. 3465–3510, 2015.

[7] D. Marangoni-Simonsen and Y. Xie, "Sequential changepoint approach for online community detection.," *IEEE Signal Process. Lett.*, vol. 22, no. 8, pp. 1035–1039, 2015.

[8] H. Chen, N. Zhang, *et al.*, "Graph-based change-point detection," *The Annals of Statistics*, vol. 43, no. 1, pp. 139–176, 2015.

[9] H. Chen, "Sequential change-point detection based on nearest neighbors," *arXiv preprint arXiv:1604.03611*, 2016.

[10] Z. I. Botev, M. Mandjes, and A. Ridder, "Tail distribution of the maximum of correlated gaussian random variables," in *Proceedings of the 2015 Winter Simulation Conference*, pp. 633–642, IEEE Press, 2015.

[11] M. F. Schilling, "Multivariate two-sample tests based on nearest neighbors," *Journal of the American Statistical Association*, vol. 81, no. 395, pp. 799–806, 1986.

[12] N. Henze, "A multivariate two-sample test based on the number of nearest neighbor type coincidences," *The Annals of Statistics*, pp. 772–783, 1988.

[13] N. Shervashidze, P. Schweitzer, E. J. v. Leeuwen, K. Mehlhorn, and K. M. Borgwardt, "Weisfeiler-lehman graph kernels," *Journal of Machine Learning Research*, vol. 12, no. Sep, pp. 2539–2561, 2011.