Sequential Changepoint Approach for Online Community Detection

David Marangoni-Simonsen and Yao Xie

***Abstract*—We present new algorithms for detecting the emergence of a community in large networks from sequential observations. The networks are modeled using Erdős-Renyi random graphs with edges forming between nodes in the community with higher probability. Based on statistical changepoint detection methodology, we develop three algorithms: the Exhaustive Search (ES), the Mixture, and the Hierarchical Mixture (H-Mix) methods. Performance of these methods is evaluated by the average run length (ARL), which captures the frequency of false alarms, and the detection delay. Numerical comparisons show that the ES method performs the best; however, it is exponentially complex. The Mixture method is polynomially complex by exploiting the fact that the size of the community is typically small in a large network. However, it may react to a group of active edges that do not form a community. This issue is resolved by the H-Mix method, which is based on a dendrogram decomposition of the network. We present an asymptotic analytical expression for ARL of the Mixture method when the threshold is large.**

***Index Terms*— Changepoint detection, community detection, sequential methods, social networks.**

I. Introduction

COMMUNITY detection within a network is a problem which arises from a wide variety of applications, including social networks, biology, and speech processing [1]–[4]. These problems often consist of some graph $\mathcal{G}$ which contains a community $\mathcal{C} \subset \mathcal{G}$ where $\mathcal{C}$ and $\mathcal{G}\backslash\mathcal{C}$ differ in some fundamental characteristic, such as the frequency of interaction (see [5] for more details). For example, in social networks, a node would be an individual and an edge between two nodes would represent a friendship or kinship of some sort shared by these two individuals, and community detection is about clustering the nodes into groups with strong inner interaction and weak outer interaction.

Community detection problems can be divided into either one-shot [5]–[10] or dynamic categories [11]–[13]. The more commonly considered one-shot setting assumes observations from static networks. The dynamic setting is concerned with sequential observations from possibly dynamic networks, and has become increasingly important since such scenarios become prevalent in social networks [12]. These dynamic categories can be further divided into networks with (1) structures that either continuously change over time [11], or (2) structures that change abruptly after some changepoint [13], the latter of which will be the focus of this paper. In online community detection problems, due to the real time processing requirement, we cannot simply adopt the exponentially complex algorithms, especially for large networks.

In this paper, we propose a new sequential changepoint detection framework for detecting an abrupt emergence of a *single* community using sequential observations of random graphs. We also adopt the Erdős-Renyi model, but our methods differ from [9] in that we use a sequential hypothesis testing formulation and the methods are based on sequential likelihood ratios, which have statistically optimal properties. On the other hand, our work is a new addition to the field of online community detection: previous work such as [6] focus on empirical study, whereas we take an approach from the rigorous statistical methodology. From the likelihood formulations, three sequential procedures are derived: the Exhaustive Search (ES), the Mixture, and the Hierarchical Mixture (H-Mix) methods. The ES method performs the best but it is exponentially complex even if the community size is known; the Mixture method is polynomially complex and it exploits the fact that the size of the community inside a network is typically small. A limit of the Mixture method is that it may raise a false alarm due to a set of highly active edges that do not form a community. The H-Mix method addresses this problem by imposing a dendrogram decomposition of the graph. The performance of the changepoint detection procedures are evaluated using the average-run-length (ARL) and the detection delay. We derived a theoretical asymptotic approximation of the ARL of the Mixture method, which was numerically verified to be accurate even in the non-asymptotic regime. Hence, the theoretical approximation can be used to determine the detection threshold efficiently. The complexity and performance of the three methods are also analyzed using numerical examples.

II. Formulation and Methods

Assume a network with N nodes and an observed sequence of independent adjacency matrices over time $X_1, X_2, \ldots$ with $X_t \in \mathbb{R}^{N\times N}$, where X_i represents the interaction of these nodes at time i. Also assume when there is no community, there are only random interactions between all nodes in the network with relatively low frequency. There may exist an (unknown) time at which a community emerges and nodes inside the community have much higher frequencies of interaction.

Manuscript received July 23, 2014; revised October 30, 2014; accepted December 04, 2014. Date of publication December 18, 2014; date of current version December 24, 2014. This work was supported in part by NSF Grant 1442635. The associate editor coordinating the review of this manuscript and approving it for publication was Prof. Nigar Kiyavash.

The authors are with the H. Milton Stewart School of Industrial and Systems Engineering, Georgia Institute of Technology, Atlanta, GA 30332 USA (e-mail: dmarangonisi@gatech.edu; yao.xie@isye.gatech.edu).

Color versions of one or more of the figures in this paper are available online at http://ieeexplore.ieee.org.

Digital Object Identifier 10.1109/LSP.2014.2381553

We formulate this problem as a sequential changepoint detection problem. The null hypothesis is that the graph corresponding to the network at each time step is a realization of an Erdős-Renyi random graph, i.e., edges are independent Bernoulli random variables that take values of 1 with probability p_0 and values of 0 with probability $1-p_0$. Let $[X]_{ij}$ denote the ijth element of a matrix X, then

$$[X_t]_{ij} = \begin{cases} 1 & \text{w. p. } p_0 \\ 0 & \text{otherwise} \end{cases} \quad \forall (i,j). \tag{1}$$

The alternative hypothesis is that there exists an unknown time κ such that after κ, an *unknown* subset of nodes $\mathcal{S}^*$ in the graph form edges between community nodes with a higher probability p_1, $p_1 > p_0$, implying that the interaction is stronger between nodes in the community:

$$[X_t]_{ij} = \begin{cases} 1 & \text{w. p. } p_1 \\ 0 & \text{otherwise} \end{cases} \quad \forall i,j \in \mathcal{S}^*, t > \kappa, \tag{2}$$

and for all others

$$[X_t]_{ij} = \begin{cases} 1 & \text{w. p. } p_0 \\ 0 & \text{otherwise} \end{cases} \quad \forall i \notin \mathcal{S}^* \text{or} j \notin \mathcal{S}^*, t > \kappa. \tag{3}$$

We assume that p_0 is known, as it is a baseline parameter which can be estimated from historic data. We will consider both cases when p_1 is either unknown or known. Let $|\mathcal{S}|$ denote the cardinality of a set $\mathcal{S}$. Here, to simplify problem, we assume that nodes within community take the a homogeneous probability of interaction p_1.

Our goal is to define a stopping rule T such that for a given large *average-run-length (ARL)* value, $\mathbb{E}^{\infty}\{T\}$, the *expected detection delay* $\mathbb{E}^{\kappa}\{T-\kappa | T > \kappa\}$ is small. Here $\mathbb{E}^{\infty}$ and $\mathbb{E}^{\kappa}$ consecutively denote the expectation when there is no changepoint, and when the changepoint occurs at time κ. This formulation corresponds the classic mini-max changepoint detection problem [14], and the desired stopping rules are determined by likelihood ratios that we describe below.

Define the following statistics for edge (i,j) and assumed changepoint time $\kappa = k$ for observations up to some time t,

$$U^{(i,j)}_{k,t,p_1} = \sum_{m=k+1}^{t} [X_m]_{ij} \log\left(\frac{p_1}{p_0}\right) + (1-[X_m]_{ij}) \log\left(\frac{1-p_1}{1-p_0}\right). \tag{4}$$

Then for a given changepoint time $\kappa = k$ and a community $\mathcal{S}$, we can write the log-likelihood ratio for (1), (2) and (3) as follows:

$$\ell(\kappa = k | p_1, \mathcal{S}) \triangleq \log\left(\prod_{m=k+1}^{t} \prod_{(i,j)\in\mathcal{S}} \frac{p_1^{[X_m]_{ij}}(1-p_1)^{1-[X_m]_{ij}}}{p_0^{[X_m]_{ij}}(1-p_0)^{1-[X_m]_{ij}}}\right) = \sum_{(i,j)\in\mathcal{S}} U^{(i,j)}_{k,t,p_1}. \tag{5}$$

A. Exhaustive Search (ES) method

Often, the probability of two community members interacting p_1 is unknown since it typically represents an anomaly (or new information) in the network. One approach is to set p_1 equal to a nominal value, say δ, which would be important to detect, and set a targeted size of the community s that may be a guess for $|\mathcal{S}^*|$. Thus, we define a stopping rule that detects a community whenever the likelihood ratio exceeds a threshold b at certain time t:

$$T_{\text{ES},1} = \inf\left\{ t : \max_{t-m_1 \le k \le t-m_0} \max_{\mathcal{S}\subset[[N]]:|\mathcal{S}|=s} \sum_{(i,j)\in\mathcal{S}} U^{(i,j)}_{k,t,\delta} \ge b \right\}, \tag{6}$$

which we refer to as the *exhaustive search (ES)* procedure. Here $[[N]] \triangleq \{1,\ldots,N\}$. In (6), the test statistic is the maximum log likelihood ratio (5) over all possible sets $\mathcal{S}$ of size s and all possible changepoint locations in a time window $k \in [t-m_1, t-m_0]$, with m_0 being the start and m_1 being the end of the window. There exists a recursive way to calculate the test statistic in (6), known as the CUSUM statistic [14]. If $m_0 = 0$, for each possible $\mathcal{S}$, we calculate $W_{\mathcal{S},t+1} = \max\{W_{\mathcal{S},t} + \sum_{(i,j)\in\mathcal{S}} U^{(i,j)}_{t,t+1,\delta}, 0\}$, with $W_{\mathcal{S},0} = 0$, and (6) can be implemented as

$$T_{\text{ES},1} = \inf\left\{ t : \max_{\mathcal{S}\subset[[N]]:|\mathcal{S}|=s} W_{\mathcal{S},t} \ge b \right\}. \tag{7}$$

An alternative approach is to replace p_1 by its maximum likelihood estimate, which can be found by taking the derivative of $\ell(\kappa = k | p_1, \mathcal{S})$ in (5) with respect to p_1 for fixed $\mathcal{S}$, k and t, setting it equal to 0 and solving for p_1:

$$\widehat{p}_1(\mathcal{S}) = \frac{2}{|\mathcal{S}|(|\mathcal{S}|-1)(t-k)} \sum_{(i,j)\in\mathcal{S}} \sum_{m=k+1}^{t} [X_m]_{ij}, \tag{8}$$

and uses $U^{(i,j)}_{k,t,\widehat{p}_1}$ when forming the detection statistic, which we refer to as $T_{\text{ES},2}$:

$$T_{\text{ES},2} = \inf\left\{ t : \max_{t-m_1 \le k \le t-m_0} \max_{\mathcal{S}\subset[[N]]:|\mathcal{S}|=s} \sum_{(i,j)\in\mathcal{S}} U^{(i,j)}_{k,t,\widehat{p}_1(\mathcal{S})} \ge b \right\}. \tag{9}$$

However, for this approach, there is no recursive formula for calculating the statistic, due to a nonlinearity resulting from substituting $\widehat{p}_1$ for p_1.

B. Mixture method

Note that the testing statistic of the ES method in (6) searches over all 2^s possible communities. The Mixture method avoids the exponential complexity of the ES method by introducing a simple probabilistic mixture model, which exploits the fact that typically the size of the community is small, i.e. $|\mathcal{S}^*|/N \ll 1$. It is motivated by the mixture method developed for detecting a changepoint using multiple sensors [15] and detecting aligned changepoints in multiple DNA sequences [16]. Assume two nodes (i,j) are both in the community with probability α, which is indicated by a set of i.i.d. Bernoulli random variables Q_{ij}

$$Q_{ij} = \begin{cases} 1 & \text{w. p. } \alpha \\ 0 & \text{otherwise} \end{cases} \quad \forall i,j \in \mathcal{S}^*. \tag{10}$$

Here α can be interpreted as a guess for $|\mathcal{S}^*|/N$. Let

$$h(x) \triangleq \log\{1-\alpha+\alpha\exp(x)\}. \tag{11}$$

Then after the changepoint, if nodes i and j both belong to the community, the likelihood ratio is given by exponential of (5);

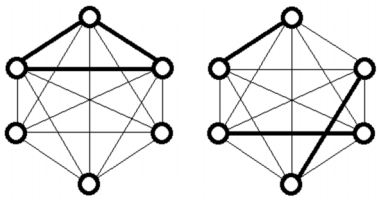

Fig. 1. (a): a community where all nodes in the community are connected with a higher probability than under the null hypothesis. (b): a model which would output the same mixture statistic that does not correspond to a community.

otherwise the likelihood ratio is 1. Hence, the likelihood ratio for the mixture model is given by

$$\ell(\kappa = k|p_1, \mathcal{S}) = \sum_{1\le i<j\le N} \log\left\{\mathbb{E}_{Q_{ij}}\left[(1-Q_{ij}) + Q_{ij}\prod_{m=k+1}^{t}\frac{p_1^{[X_m]_{ij}}(1-p_1)^{1-[X_m]_{ij}}}{p_0^{[X_m]_{ij}}(1-p_0)^{1-[X_m]_{ij}}}\right]\right\} = \sum_{1\le i<j\le N} h(U_{k,t,p_1}^{(i,j)}). \tag{12}$$

Again, similarly, for the unknown p_1 we can either replace it with a fixed nominal value δ, or with a maximum likelihood estimate. Forming the exact maximum likelihood estimate for p_1 in the mixture model is hard. One possible estimate would be $\tilde{p}_1 = \max_{\mathcal{S}\subset[[N]]:|\mathcal{S}|=s} p_1(\mathcal{S})$, which, however, requires searching over all possible sets. Therefore, hereafter we only consider the first approach

$$T_{\text{Mix}} = \inf\left\{t : \max_{t-m_1\le k\le t-m_0} \sum_{1\le i<j\le N} h(U_{k,t,\delta}^{(i,j)}) \ge b\right\}, \tag{13}$$

where b is the threshold. Here $h(x)$ can be viewed as a soft-thresholding function [15] that selects the edges which are more likely to be between community members.

C. Hierarchical Mixture method (H-Mix)

One problem with the Mixture method is that its statistic can be gathered from edges that do not form a community. Fig. 1 below displays two scenarios where the mixture statistics will be identical, but Fig. 1(b) does not correspond to a network forming a community. To solve this problem, we introduce the hierarchical Mixture method (H-Mix) that takes advantage of the low computational complexity of the Mixture method while enforcing the statistics to be calculated only over meaningful communities. The H-Mix method requires setting a targeted size of the community s that may be a guess for $|\mathcal{S}^*|$.

The H-Mix method enforces the community structure by constructing a dendrogram decomposition of the network, which is a hierarchical partitioning of the graphical network [17]. The hierarchical structure provided by dendrogram enables us to systematically remove nodes from being considered for the community. Suppose a network has a community of size s. Starting from the root level with all nodes belonging to the community, each of the nodes in the dendrogram tree decomposition is a subgraph of the entire network that contains all but one node. Then the mixture statistic from (13) is evaluated for each subgraph: using $h(x)$ defined in (11), for a given set of nodes $\mathcal{S}_0$, nominal value $p_1 = \delta$, and a hypothesized changepoint location k, the mixture statistic is calculated as

$$M(\mathcal{S}_0) = \sum_{(i,j)\in\mathcal{S}_0} h\left(U_{k,t,\delta}^{(i,j)}\right). \tag{14}$$

We iteratively select the subgraph with the highest mixture statistic value, since it indicates that the associated node removed is most likely to be a non-member of the community and will be eliminated from subsequent steps. The algorithm repeats until there are only s nodes remaining in the subgraph. Denote the mixture statistic for the selected subgraph as P_k. Then $\{P_k\}_{k=1}^t$ is a series of test statistics at each hypothesized changepoint location k. Finally, the H-Mix method is given by

$$T_{\text{H-Mix}} = \inf\left\{t : \max_{t-m_1\le k\le t-m_0} P_k \ge b\right\}, \tag{15}$$

where b is the threshold. The idea for a dendrogram decomposition is similar to the edge removal method [18]. An illustration for the procedure described above and Algorithm 1 summarizes the H-Mix method can be found in [19].

Algorithm 1 Hierarchical Mixture Method

Input: $\{X_m\}_{m=1}^t, X_m \in \mathbb{R}^{N\times N}$
Output: $\{P_k\}_{k=1}^t \in \mathbb{R}^t$, a set of statistics for each hypothesized changepoint location k.
for $k = 1 \to t$ **do**
  $\mathcal{S} = [[N]]$
  while $|\mathcal{S}| > s$ **do**
    $i^* = \text{argmax}_{i\in\mathcal{S}} M(\mathcal{S}\backslash\{i\})$
    $\mathcal{S} = \mathcal{S}\backslash\{i^*\}$
  end while
  $P_k = M(\mathcal{S})$
end for

The complexity for the algorithms are summarized as follows (derivations can be found in [19]): (1) when $s \ll N/2$, the complexity of the ES algorithm is $\mathcal{O}(N^s)$, of the Mixture method is $\mathcal{O}(N^2)$, of the H-Mix method is $\mathcal{O}(N^4)$; and (2) when s is on the order of $N/2$, the complexity of the ES algorithm is $\mathcal{O}(2^{s/2})$, of the Mixture method is $\mathcal{O}(N^2)$, of the H-Mix method is $\mathcal{O}(N^4)$. Note that the mixture and the H-Mix both have much lower complexity than the ES method.

III. Theoretical Analysis for the Mixture Method

In this section, we present a theoretical approximation for the ARL of the Mixture method with a nominal value $p_1 = \delta$ using techniques outlined as follows. In [16] a general expression for the tail probability of scan statistics is given, which can be used to derive the ARL of a related changepoint detection procedure. For example, in [15] a generalized form for ARL was found using the expression in [16]. The basic idea is to relate the probability of stopping a changepoint detection procedure when there is no change, $\mathbb{P}^\infty\{T \le m\}$, to the tail probability of the maxima of a random field: $\mathbb{P}^\infty\{S \ge b\}$, where S is the statistic used for the changepoint detection procedure, b is the threshold, and $\mathbb{P}^\infty$ denotes the probability measure when there is no change. Hence, if we can write $\mathbb{P}^\infty\{S \ge b\} \approx m\lambda$ for some λ, by relying on the assumption that the stopping time is asymptotically

exponentially distributed when $b \to \infty$, we can find the ARL is $1/\lambda$. However, the analysis for the Mixture method here differs from that in [15] in two major aspects: (1) the detection statistics here involve a binomial random variable, and we will use a normal random variable to approximate its distribution; (2) the change-of-variable parameter θ depends on $t-k$ and, hence, the expression for ARL will be more complicated than that in [15].

Theorem 1: When $b \to \infty$, an upper approximation to the ARL $\mathbb{E}^{\infty}[T_{\text{mix}}]$ of the Mixture method with known p_1 is given by:

$$\text{ARL}_{\text{UA}} = \left[\int_{\sqrt{2N/m_1}}^{\sqrt{2N/m_0}} \frac{y\nu^2(y\sqrt{\gamma(\theta_y)})}{H(N,\theta_y)}dy\right]^{-1}, \quad (16)$$

and a lower approximation to the ARL is given by:

$$\text{ARL}_{\text{LA}} = \left[\sum_{\tau=m_0}^{m_1} \frac{2N\nu^2(2N\sqrt{\gamma(\theta_\tau)}/\tau^2)}{\tau^2 H(N,\theta_\tau)}\right]^{-1}, \quad (17)$$

where

$$\begin{aligned}
c_0 &= \log(p_1/p_0),\ c_1 = \log[(1-p_1)/(1-p_0)],\\
g_\tau(x) &= \tau[p_0(c_0-c_1)+c_1] + x\sqrt{\tau(c_0-c_1)^2 p_0(1-p_0)},\\
h_\tau(x) &\triangleq h(g_\tau(x)), \text{for } h(x) \text{defined in} (11),\\
\psi_\tau(\theta) &= \log \mathbb{E}\{e^{\theta h_\tau(Z)}\},\\
\dot{\psi}_\tau(\theta) &= \frac{\mathbb{E}\{h_\tau(Z)e^{\theta h_\tau(Z)}\}}{\mathbb{E}\{e^{\theta h_\tau(Z)}\}},\\
\ddot{\psi}_\tau(\theta) &= \frac{\mathbb{E}\{h_\tau^2(Z)e^{\theta h_\tau(Z)}\}}{\mathbb{E}\{e^{\theta h_\tau(Z)}\}} - \frac{\left(\mathbb{E}\{h_\tau(Z)e^{\theta h_\tau(Z)}\}\right)^2}{\left(\mathbb{E}\{e^{\theta h_\tau(Z)}\}\right)^2},\\
\gamma(\theta) &= \frac{1}{2}\theta^2 \mathbb{E}\left\{[\dot{h}_\tau(Z)]^2 \exp\{\theta h_\tau(Z) - \psi_\tau(\theta)\}\right.,\\
&\theta_\tau \text{is solution to} \dot{\psi}_\tau(\theta_\tau) = b/N,\\
H(N,\theta_\tau) &= \frac{\theta[2\pi\ddot{\psi}(\theta_\tau)]^{1/2}}{\gamma^2(\theta_\tau)\sqrt{N}} e^{N[\theta\dot{\psi}(\theta_\tau)-\psi(\theta_\tau)]},
\end{aligned}$$

and $\dot{f}$ and $\ddot{f}$ denote the first and second order derivatives of a function f, Z is a normal random variable with zero mean and unit variance, the expectation is with respect to Z, and the special function $\nu(x)$ is given by [15]

$$\nu(x) \approx \frac{(2/x)[\Phi(x/2) - 1/2]}{(x/2)\Phi(x/2) + \phi(x/2)}.$$

Here θ_τ is the solution to

$$\dot{\psi}_\tau(\theta_\tau) = b/N.$$

Proof for this theorem can be found in [19]. We verify the theoretical upper and lower approximations for ARL of the Mixture method versus the simulated values, and consider a case with $p_0 = 0.3$, $p_1 = 0.8$, and $N = 6$. The comparison results are shown in Fig. 2. These comparisons show that the lower approximation is an especially good approximation to the simulated ARL and, hence, it can be used to efficiently determine a threshold corresponding a desired ARL (which is typically set to a large number around 5000 or 10000), as shown in Table I.

IV. Numerical Examples

In this section, we compare the performance of our three methods via numerical simulations. We first use simulations to

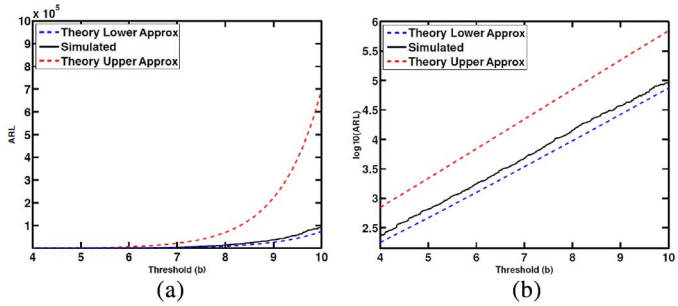

Fig. 2. Comparison of the theoretical upper and lower approximations with the simulated ARL for a case with $N = 6$, $p_0 = 0.3$, and $p_1 = 0.8$.

TABLE I
THEORETICAL VS. SIMULATED THRESHOLDS FOR $p_0 = 0.3$, $p_1 = 0.8$, AND $N = 6$. THE THRESHOLD b CALCULATED USING THEORY IS VERY CLOSE TO THE CORRESPONDING THRESHOLD OBTAINED USING SIMULATION

ARL	Theory b	Simulated ARL	Simulated b
5000	7.37	5049	7.04
10000	8.05	10210	7.64

TABLE II
COMPARISON OF DETECTION DELAYS FOR VARIOUS CASES WHEN $N = 6$. THE NUMBERS INSIDE THE BRACKETS ARE THE THRESHOLD b SUCH THAT ARL $= 5000$

	$T_{\text{ES},1}$ $\delta = p_1$	T_{Mix} $\delta = p_1$	$T_{\text{H-Mix}}$	T_{Mix} $\delta = p_1 - 0.1$
$N = 6$, $s = 3$, $p_0 = 0.2$, $p_1 = 0.9$	3.8 (9.96)	4.3 (6.71)	3.8 (9.95)	6.0 (6.71)
$N = 6$, $s = 3$, $p_0 = 0.3$, $p_1 = 0.7$	9.5 (10.17)	12.8 (6.77)	10.8 (10.18)	23.3 (6.77)
$N = 6$, $s = 4$, $p_0 = 0.3$, $p_1 = 0.7$	5.0 (8.48)	6.7 (6.88)	6.4 (10.17)	11.0 (6.88)
$N = 50$, $s = 10$, $p_0 = 0.3$, $p_1 = 0.7$	-	27.5 (-7.44)	-	-
$N = 50$, $s = 20$, $p_0 = 0.3$, $p_1 = 0.7$	-	1.1 (-7.41)	-	-

determine the threshold b for each method, so these methods all have the same average run length (ARL) which is equal to 5000, and then estimate the detection delays using these b's under different scenarios. The results are shown in Table II, including the detection delay and the thresholds (shown in brackets). The $\alpha = 0.3$ was set for all mixture model statistics and communities were formed. Note that the low-complexity mixture and H-Mix methods can both detect the community quickly. The detection delays in the first row are smaller than those in the the second row, as the first row represents an easier case to detect. Similar explanation can be applied to the second and third row. Also note that when δ does not equal the true p_1, the Mixture method can still be applied but the detection delay will be longer. For larger N, we only show the Mixture method as it has the lowest complexity.

V. Discussions

As we use simple stylized Erdős-Renyi models for the networks, this framework works for the communities with noticeable change of the frequency of members. Other kinds of communities with different characteristics such as friends, colleagues, special groups, etc. cannot be detected using this framework.

REFERENCES

[1] M. Girvan and M. E. J. Newman, "Community structure in social and biological networks," in *Proc. Nat. Acad. Sci.*, USA, 2002, vol. 99, pp. 7821–7826.

[2] M. E. J. Newman and M. Girvan, "Finding and evaluating community structure in networks," *Phys. Rev. E.*, vol. 69, no. 026113, 2004.

[3] S. H. Shum and W. M. Campbell, "Large-scale community detection on speaker content graphs," in *IEEE Int. Conf. Acoustics, Speeches and Sig. Proc. (ICASSP)*, May 2013, pp. 7716–7720.

[4] V. Kawadia and S. Screenivasan, "Sequential detection of temporal communities by estrangement confinement," *Scientific Reports*, vol. 2, 2012.

[5] E. Arias-Castro and N. Verzelen, "Community detection in dense random networks," *Ann. Stats.*, vol. 42, no. 3, pp. 940–969, 2014.

[6] I. X. Leung, P. Hui, P. Liò, and J. Crowcroft, "Towards real-time community detection in large networks," *Phys. Rev. E*, vol. 6, Jun. 2009.

[7] J. Leskovec, K. J. Lang, and M. W. Mahoney, "Empirical comparison of algorithms for network community detection," in *Proc. 19th Int. Conf. World wide web*, Apr. 2010, pp. 631–640.

[8] S. Bhattacharyya and P. J. Bickel, *Community detection in networks using graph distance*, Jan. 2014, ArXiv ID: 14013915.

[9] N. Verzelen and E. Arias-Castro, *Community detection in sparse random networks*, Aug. 2013, ArXiv ID: 13082955.

[10] T. Cai and X. Li, *Robust and computationally feasible community detection in the presence of arbitrary outlier nodes*, Apr. 2014, ArXiv ID: 14046000.

[11] S. Koujaku, M. Kudo, I. Takigawa, and H. Imai, "Structrual change point detection for evolutional networks," in *Proc. World Cong. on Engr.*, Jul. 2013, vol. 1.

[12] W. Zhang, G. Pan, Z. Wu, and S. Li, "Online community detection for large complex networks," in *Proc. 23rd Int. Conf. Artificial Intelligence*, 2013, pp. 1903–1909.

[13] D. Duan, Y. Li, Y. Jin, and Z. Lu, "Community mining on dynamic weighted directed graphs," in *Proc. on Complex Networks Meet Information, CNIKM '09, ACM*, New York, NY, USA, Nov. 2009, pp. 11–18.

[14] D. Siegmund, *Sequential analysis: Tests and confidence intervals*. New York, NY, USA: Springer, 1985.

[15] Y. Xie and D. Siegmund, "Sequential multi-sensor change-point detection," *Ann. Stats.*, vol. 41, pp. 670–692, May 2013.

[16] D. Siegmund, B. Yakir, and N. R. Zhang, "Detecting simultaneous variant intervals in aligned sequences," *Ann. Appl. Stats.*, vol. 5, pp. 645–668, Aug. 2011.

[17] A. Krishnamurthy, J. Sharpnack, and A. Singh, "Recovering graph-structured activations using adaptive compressive measurements," in *Asilomar Conf. Sig. Sys. and Computers*, Nov. 2013, pp. 765–769.

[18] M. Newman, "Detecting community structure in networks," *Euro. Phys. Jour. B - Condensed Matter and Complex Systems*, vol. 38, pp. 321–330, Mar. 2004.

[19] D. Maragoni-Simonsen and Y. Xie, *Sequential changepoint approach for community detection*, Jul. 2014, ArXiv:1407.5978.