Attribute fusion in a latent process model for time series of graphs

Minh Tang, Youngser Park, Nam H. Lee, and Carey E. Priebe*

***Abstract*—Hypothesis testing on time series of attributed graphs has applications in diverse areas, e.g., social network analysis (wherein vertices represent individual actors or organizations), connectome inference (wherein vertices are neurons or brain regions) and text processing (wherein vertices represent authors or documents). We consider the problem of anomaly/change point detection given the latent process model for time series of graphs with categorical attributes on the edges presented in [1]. Various attributed graph invariants are considered, and their power for detection as a function of a linear fusion parameter is presented. Our main result is that inferential performance in mathematically tractable first-order and second-order approximation models does provide guidance for methodological choices applicable to the exact (realistic but intractable) model. Furthermore, to the extent that the exact model is realistic, we may tentatively conclude that approximation model investigations have some bearing on real data applications.**

I. Introduction

The notion of data as a time series is fundamental to signal processing. However, time-series analysis had been mostly confined to data in Euclidean spaces, while the representation of data as graphs with the vertices as the entities and the edges as the relationships between the entities is now ubiquitous in many application domains, e.g., social networks (wherein vertices represent individual actors or organizations), neuroscience (wherein vertices are neurons or brain regions), and text processing (wherein vertices represent authors or documents). Furthermore, in many of these application domains, the edges between the vertices, in addition to representing the presence of a relationship, can also be use to identify the attributes associated with that relationship. For example, in social network graphs where edges represent email communications, the attributes signify the context of the email messages. Finally, in many of these application domains, the underlying data changes over time, and thus a representation of data as a time-series of attributed graphs is natural. The analysis of time-series of (attributed) graphs thus presents a natural and challenging extension to the signal processing community.

Minh Tang is with the Department of Applied Mathematics and Statistics, Johns Hopkins University, Baltimore, MD 21218 USA. (email: mtang10@jhu.edu)

Youngser Park is with the Center of Imaging Science, Johns Hopkins University, Baltimore, MD 21218 USA. (email:youngser@jhu.edu)

Nam H. Lee is with the Department of Applied Mathematics and Statistics, Johns Hopkins University, Baltimore, MD 21218 USA. (email: nhlee@jhu.edu)

Carey E. Priebe is with the Department of Applied Mathematics and Statistics, Johns Hopkins University, Baltimore, MD 21218 USA. (phone: 410-516-7200; fax:410-516-7459; email: cep@jhu.edu)

One of the main exploitation tasks in time-series analysis is the problem of anomaly/change-point detection, where anomaly is broadly interpreted to mean deviation from some "normal" pattern and change-point is the time window at which the anomalous deviation occurs. In the context of a time-series of graphs, an anomaly can be for example the presence of a specific subgraph, or frequent occurrences of some subgraphs, or other more vague notions such as existence of a community structure.

The notion of signal processing for graphs appears to be in an emerging state, with significant research activity in constructing wavelet representations for graphs, e.g., [2]–[5]. The formulation of graph analysis problems in a signal processing framework is also of potential interest, e.g, [6], wherein a subgraph detection problem is presented, test statistics are defined, noise and signal models are introduced, and ROC analysis is performed. The construction of wavelets representations on graphs is analogous to the classical wavelets constructions, but adapted to use the topology of the graphs instead of the usual Euclidean space topology. For example, the wavelet representations of [2] and [3] are constructed using the combinatorial and normalized Laplacian matrices of the graphs, respectively, while [4] defined wavelets via the n-hop neighbourhoods. The wavelet construction in [4] and its use in *spatial* traffic analysis (emphasis from [4]) is similar to the use of locality/scan statistics for anomaly detection in [22] and the current paper. Indeed, the locality statistic at a vertex v is the size of a n-hop neighbourhood starting at v defined via shortest path distance and the scan statistic is the maximum of the locality statistics over all the n-hop neighbourhood. Traffic analysis by tracking the coefficients of the wavelet representations is similar to anomaly detection by tracking the values of the scan statistics.

In this paper, we investigate the anomaly/change-point detection problem for time-series of attributed graphs. In the spirit of time-series analysis, we approach this problem via test statistics based on the moving average of some graph invariants. We derive limiting distributions for these test statistics under a generative model for the attributed graphs. The experimental evidence indicate that these test statistics have sufficient power against the null hypothesis for a general class of change-point problem but that many interesting questions remain to ponder.

The current paper builds upon ideas previously presented in [1] and [7], among others. The generative model for time-series of attributed graphs used in this paper is from [1]. Hypothesis testing for time-series of graphs and attribute fusion using a linear fusion parameter λ were considered in [1], [7]; limiting distribution for the size graph invariant was given in [1]. A version of the experimental results for simulated data included in this paper was first presented in [7].

The paper is structured as follows. We present an overview of the generative model for time-series of attributed graphs [1] in §II. The change-point detection problem along with the relevant test statistics are described in §III. The distributions of the test statistics and the power estimates for hypothesis testing using these test statistics are given in §IV. §V presents a synthetic data example and an example using the Enron email corpus to illustrate our methodology.

II. Latent Process Model

The abundance of graph representations in diverse application domains leads to a proliferation of random graph models to model the association among the entities. In addition to the classical Erdös-Rényi model [8], there is, the exponential graph models [9], the stochastic block model [10], [11], the latent position model [12], the dot product model for unattributed graphs [13] and the latent process model for attributed graphs [1]. We now describe briefly the latent position model of [12] and the dot product model of [13] as they are closely related to the latent process model in [1]. The latent process model serves as the generative model for our attributed graphs and thus plays a key role in our setup of the anomaly detection problem in § III.

The latent space model is one where each vertex is associated with a latent random vector. There may also be additional covariates information but those are not relevant to the current exposition. The vectors are independent and identically distributed and the probability of an edge between two nodes depends only on their latent vectors. Conditioned on the latent vectors, the presence of each edge is an independent Bernoulli trial. For example, if p_{uv} is the probability of an edge between the vertices u and v, then $p_{uv} = \frac{1}{1+\exp(-\|z_u - z_v\|)}$ where z_u and z_v are the latent vectors associated with vertex u and v, respectively. Another example of a latent space model is the random dot product graph model [13]. Under the random dot product graph model, the probability of an edge between two vertices is given by the dot product of their respective latent vectors, i.e., $p_{uv} = \langle z_u, z_v \rangle$. Parts of the motivation of the random dot product graph model is that, in a social network with edges indicating friendships, the components of the vector may be interpreted as the relative interest of the individual in various topics. The magnitude of the vector can be interpreted as how talkative the individual is, with more talkative individuals more likely to form relationships. Talkative individuals interested in the same topics are most likely to form relationships while individuals who do not share interest are unlikely to form relationships.

The latent process model for time-series of attributed graphs was presented in [1]. The main ideas underlying the model is as follows. The model is motivated by the assumption that each vertex is associated with a stochastic process, in this case a finite-state continuous time Markov chain. The stochastic processes then associate to each vertex v, for the time interval $(t-1,t]$, a (latent) random vector $X_v(t)$ in the unit simplex with the k-th entry of $X_v(t)$ denoting the proportion of time the stochastic process spent in state k. The probability of interactions between vertices u and v during $(t-1,t]$ is then given by the dot product of X_u and X_v a la the dot product model of [13], with a slight modification to account for the presence of attributes on the edges.

We now give the necessary details to make the above ideas precise. For any positive integer n, we shall denote by $[n]$ the set of numbers $\{1,2,\dots,n\}$. Let $G=(V,E)$ be an undirected graph, equipped with an edge-attribution function $\phi\colon \binom{V}{2} \mapsto [K+1]$ for some positive integer K. We shall require that ϕ satisfies $\phi(e) \leq K$ for $e \in E$ and $\phi(e) \equiv K+1$ for $e \notin E$. We refer to $\phi(e)$ as the attribute of edge $e \in E$, and the pair (G,ϕ) as an attributed graph. Let $\mathscr{S}$ be the unit simplex in $\mathbb{R}^K$, i.e.,

$$\mathscr{S} = \{\xi \in [0,1]^K \colon \sum_{k=1}^{K} \xi_k \leq 1\}. \tag{1}$$

A *random dot product space* for attributed graphs with vertices in $[n]$ and edge attributes $[K]$ is a pair $(\mathbf{X}, G)$ of random elements such that

1) $\mathbf{X} = \{X_v\}_{v=1}^n$ is a collection of independent $\mathscr{S}$-valued random vectors.
2) G is a random undirected graph with vertex set $[n]$ such that the edges of G are conditionally independent given $\mathbf{X}$ and that
$$\mathbb{P}(u \sim v \mid \mathbf{X}) = \mathbb{P}(u \sim v \mid X_u, X_v) = \langle X_u, X_v \rangle. \tag{2}$$
Furthermore, we also have
$$\mathbb{P}(\phi(\{u,v\}) = k \mid u \sim v) = \frac{X_{u,k} X_{v,k}}{\langle X_u, X_v \rangle} \tag{3}$$
i.e., the pair $\{u,v\}$ is an edge in G with attribute k with probability $X_{u,k}X_{v,k}$.

We say that a cádlág (right continuous with left limit) process $W\colon [0,\infty) \mapsto [K+1]^n$ induces the sequence of random dot product spaces $\mathscr{V} = (\mathbf{X}(t), G(t))_{t=1}^{\infty}$ if

1) Each $(\mathbf{X}(t), G(t))$ is a random dot product space with vertices $[n]$ and edge attributes $[K]$. Furthermore, for each $v \in [n]$, $k \in [K]$ and $t \in \mathbb{N}$, we have
$$X_{v,k}(t) = \int_{t-1}^{t} \mathbf{1}\{W_v(\omega) = k\}\, d\omega \tag{4}$$
where we denote $W_v(\omega)$ as the v-th component of $W(\omega)$ and $X_{v,k}$ as the k-th component of X_v.
2) For each $t \in \mathbb{N}$, we have
$$\mathbb{P}(G(t) = g \mid \mathscr{F}_{\leq t}) = \mathbb{P}(G(t) = g \mid \mathbf{X}(t)) \tag{5}$$
where $\mathscr{F}_{\leq t}$ is the σ-field generated by $\{W(s)\colon s \leq t\}$.

We will call any pair $(\mathscr{V}, W)$ that satisfies the above properties a random dot process model.

The random dot process model can be interpreted as follows. During each time interval $(t-1, t]$, a particular vertex is inclined to communicate with other vertices on a variety of topics (the topics are enumerated as the elements of $[K]$). As the inclination of a vertex v is governed by its underlying stochastic process W_v, we can view the $\{W_v\}$ as encapsulating a notion of changes in the inclination of the vertices over time. The propensity for vertex v to communicate about a particular topic k is $X_{v,k}(t)$ and two vertices u and v are more likely to exchange communication on topic k if the propensity to communicate about topic k is high for both u and v. Eq. (2) and Eq. (3) attempt to capture these two assumptions.

In this paper, we are interested in the pairs $(\mathscr{V}, W)$ possessing the following properties:

1) For each $t \in \mathbb{N}$ and vertex $v \in [n]$, there exists a matrix $\mathbf{Q}^{(v)}(t)$ such that W_v, when restricted to the interval $[t, t+1)$, is a stationary, continuous-time Markov chain with state space $[K+1]$, intensity matrix $\mathbf{Q}^{(v)}(t)$, and stationary distribution $\pi^{(v)}(t)$. W is thus a piecewise stationary, continuous-time Markov chain with state space $[K+1]^n$ and intensity matrix $\otimes_{v=1}^{n} \mathbf{Q}^{(v)}(t)$.
2) There exists a $t^* \in \mathbb{N}$ and a $m < n$ such that, for some $\pi_0, \pi_1, \mathbf{Q}_0$ and $\mathbf{Q}_1$ we have
   a) for $t < t^* - 1$,
$$\pi^{(1)}(t) = \cdots = \pi^{(n)}(t) = \pi_0$$
$$\mathbf{Q}^{(1)}(t) = \cdots = \mathbf{Q}^{(n)}(t) = \mathbf{Q}_0,$$
   b) for $t \geq t^* - 1$,
$$\pi^{(1)}((t)) = \cdots = \pi^{(m)}(t) = \pi_1$$
$$\pi^{(m+1)}(t) = \cdots = \pi^{(n)}(t) = \pi_0$$
$$\mathbf{Q}^{(1)}(t) = \cdots = \mathbf{Q}^{(m)}(t) = \mathbf{Q}_1$$
$$\mathbf{Q}^{(m+1)}(t) = \cdots = \mathbf{Q}^{(n)}(t) = \mathbf{Q}_0$$

The above properties characterize a random dot process model with a change-point phenomenon for $(X(t), G(t))$ at time $t = t^*$. We chose $\{1, 2, \dots, m\}$ as the set of vertices with a change in the stationary distribution and intensity matrix at time $t = t^* - 1$ for ease of discussion. Permutations of the vertex labels does not affect our subsequent analysis. We will refer to $(t^*, m, \pi_0, \pi_1, \mathbf{Q}_0, \mathbf{Q}_1)$ as the change-parameters. Because W is completely determined by the change-parameters, we often choose to omit W and only mention $\mathscr{V}$ when referring to random dot process models with change-point phenomenon. For a random dot process model $\mathscr{V}$ with change-parameters $(t^*, m, \pi_0, \pi_1, \mathbf{Q}_0, \mathbf{Q}_1)$, we can construct several approximations to $\mathscr{V}$. Of particular interest are the following two approximations.

Definition 1. Let $\mathscr{V}$ be a random dot process model (rdpm) with change parameters $(t^*, m, \pi_0, \pi_1, \mathbf{Q}_0, \mathbf{Q}_1)$. The first order approximation $\bar{\mathscr{V}}$ of $\mathscr{V}$ is the sequence $\{(\bar{X}(t), \bar{G}(t)\}_{t=1}^{\infty}$ of *independent* random dot product spaces such that

1) for $t < t^* - 1$,
$$\bar{X}_v(t) \equiv \hat{\pi}_0 \quad \text{for all } v \in [n] \tag{6}$$

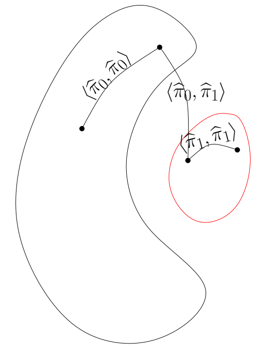


Fig. 1: An illustration of the kidney-egg model for $G(t^*)$. The probability of an edge is $\langle\hat{\pi}_0, \hat{\pi}_0\rangle$ for the connections inside the kidney set (black), $\langle\hat{\pi}_1, \hat{\pi}_1\rangle$ for the connections inside the egg set (red), and $\langle\hat{\pi}_0, \hat{\pi}_1\rangle$ for the connections between the kidney and egg set.

2) for $t \geq t^* - 1$
$$\bar{X}_v(t) \equiv \hat{\pi}_1 \quad \text{for } v \leq m$$
$$\bar{X}_v(t) \equiv \hat{\pi}_0 \quad \text{for } v > m$$

where $\hat{\pi}_0$ and $\hat{\pi}_1$ are sub-probability vectors obtained by removing the last coordinate of π_0 and π_1.

The first approximation yields a sequence of independent random graphs with independent edges. For $t \leq t^* - 1$, $G(t)$ is an attributed instance of the Erdös-Renyi random graphs. For $t \geq t^*$, $G(t)$ is an attributed instance of the kidney-egg model [14] (see Fig. 1), which in itself is a special instance of the stochastic block model [10]. The unattributed kidney-egg model can be denoted as $\kappa(n, m, p, q, s)$ with the quantities p, q, and s representing

$$p = \mathbb{P}(u \sim v \,|\, u \leq m, v \leq m) = \langle\hat{\pi}_1, \hat{\pi}_1\rangle$$
$$q = \mathbb{P}(u \sim v \,|\, u \leq m, v > m) = \langle\hat{\pi}_1, \hat{\pi}_0\rangle$$
$$s = \mathbb{P}(u \sim v \,|\, u > m, v > m) = \langle\hat{\pi}_0, \hat{\pi}_0\rangle.$$

Definition 2. Let $\mathscr{V}$ be a rdpm with change-parameters $(t^*, m, \pi_0, \pi_1, \mathbf{Q}_0, \mathbf{Q}_1)$. Define $\mathbf{Z}_0$ and $\mathbf{Z}_1$ by

$$\mathbf{Z}_0 = (\mathbf{1}\pi_0^T - \mathbf{Q}_0)^{-1}(\mathbf{I} - \mathbf{1}\pi_0^T) \tag{7}$$
$$\mathbf{Z}_1 = (\mathbf{1}\pi_1^T - \mathbf{Q}_1)^{-1}(\mathbf{I} - \mathbf{1}\pi_1^T). \tag{8}$$

$\mathbf{Z}_0$ and $\mathbf{Z}_1$ are the fundamental matrices for the continuous-time Markov chain on $[K+1]$ with intensity matrix $\mathbf{Q}_0$ and $\mathbf{Q}_1$ (see e.g. [16, p. 55]). Let Σ_0 and Σ_1 be given by

$$\Sigma_0 = \mathrm{diag}(\pi_0)\mathbf{Z}_0 + \mathbf{Z}_0^T \mathrm{diag}(\pi_0)$$
$$\Sigma_1 = \mathrm{diag}(\pi_1)\mathbf{Z}_1 + \mathbf{Z}_1^T \mathrm{diag}(\pi_1).$$

The second order approximation $\widehat{\mathscr{V}}$ of $\mathscr{V}$ is the sequence $\{\widehat{X}(t), \widehat{G}(t)\}_{t=1}^{\infty}$ where

1) For each t and each $v \in [n]$, $\widehat{X}_v(t)$ is a random vector obtained by truncating a multivariate normal random

vector $Z_v(t)$ to $\mathscr{S}$ with mean and covariance matrices given below.

2) For $t < t^* - 1$,

$$\mathbb{E}[Z_v(t)] \equiv \hat{\pi}_0, \ \mathrm{Var}[Z_v(t)] \equiv \widehat{\Sigma}_0 \quad \text{for all } v. \tag{9}$$

3) For $t \geq t^* - 1$,

$$\mathbb{E}[Z_v(t)] \equiv \hat{\pi}_1, \ \mathrm{Var}[Z_v(t)] \equiv \widehat{\Sigma}_1 \quad \text{for } v \leq m$$
$$\mathbb{E}[Z_v(t)] \equiv \hat{\pi}_0, \ \mathrm{Var}[Z_v(t)] \equiv \widehat{\Sigma}_0 \quad \text{for } v > m$$

where $\widehat{\Sigma}_0$ and $\widehat{\Sigma}_1$ are the $K \times K$ matrices obtained by removing the last row and column of Σ_0 and Σ_1, respectively.

A second-order approximation yields a sequence of independent latent position graphs [12], [17], [18].

Suppose that we have a sequence $\{\mathscr{V}^r : r > 0\}$ of rdpm with vertices $[n]$ and attributes $[K]$ where for each $r > 0$, $\mathscr{V}^r$ has change-parameters $(t^*, m, \pi_0, \pi_1, r\mathbf{Q}_0, r\mathbf{Q}_1)$. The parameter r can be thought of as the vertex process rate, i.e., the waiting time decreases exponentially as r increases. Let us now consider the sequence of first approximations $\bar{\mathscr{V}}^r$ and second approximations $\widehat{\mathscr{V}}^r$ of $\mathscr{V}^r$. We note that $\bar{\mathscr{V}}^{r_1} = \bar{\mathscr{V}}^{r_2}$ for any $r_1, r_2 > 0$. Let us then denote by $\bar{\mathscr{V}}$ the (a.e.) unique first order approximation of $\mathscr{V}^r$ for $r > 0$. If we denote by $\widehat{\Sigma}_0^{(r)}$ and $\widehat{\Sigma}_1^{(r)}$ the matrices $\widehat{\Sigma}_0$ and $\widehat{\Sigma}_1$ for $\widehat{\mathscr{V}}^r$, then $\widehat{\Sigma}_0^{(r)} = \widehat{\Sigma}_0^{(1)}/r$ and $\widehat{\Sigma}_1^{(r)} = \widehat{\Sigma}_1^{(1)}/r$. Therefore, as $r \to \infty$, $\widehat{\Sigma}_0^{(r)} \to \mathbf{0}$ and $\widehat{\Sigma}_1^{(r)} \to \mathbf{0}$ and so $\widehat{\mathscr{V}}^r \xrightarrow{\mathrm{d}} \bar{\mathscr{V}}$. This indicates that for sufficiently large r, there is little statistical difference among the $\bar{\mathscr{V}}$, $\mathscr{V}^r$ and $\widehat{\mathscr{V}}^r$ ([1, Theorem 2]). The behavior of $\widehat{\mathscr{V}}^r$ for "small" r is less clear. $\bar{\mathscr{V}}$ will thus serve as the basis for our subsequent analysis on graph invariants for attributed random graphs.

A comment should also be made about the utility of the exact generative model. It is computationally intractable to estimate the parameters of the underlying continuous time Markov chains, i.e. the $\{\mathbf{Q}^{(v)}(t)\}$, given a snapshot of the graphs. Therefore, the main interest of [1], and consequently of this paper, lies not in the exact generative model but in its first and second-order approximation. We will not be concerned with the estimation of the latent position vectors for the first and second-order approximation in this paper, but we note that they can be done via spectral embedding technique a la [19]. We also note that more refined models, where the messaging events between each pairs of vertices is modeled as point process, e.g. [20], [21], can also be formulated. The snapshot of the graphs is then a binning of these point processes and the issue of parameter estimation is fundamental but feasible for these point processes approaches.

III. Change-point detection

Let $\mathscr{V}$ be a random dot process model with change-parameters $(t^*, m, \pi_0, \pi_1, \mathbf{Q}_0, \mathbf{Q}_1)$. The change-parameters encapsulate a notion of chatter anomalies, i.e., a subset of vertices of $\mathscr{V}$ with altered communication behavior in an otherwise stationary setting as depicted in Fig. 2. We are interested in the problem

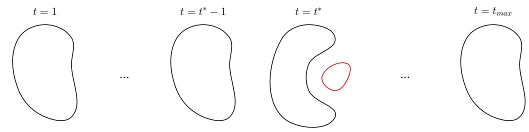


Fig. 2: Notional depiction of the problem of change-point detection in a time-series of graphs

of testing, for a $t \in \mathbb{N}$, the hypotheses that t is the change-point of $\mathscr{V}$, namely

$$\mathscr{H}_0 \colon t^* > t$$
$$\mathscr{H}_A \colon t^* = t$$

This will be done using the notion of fusion of attributed graph invariants. The particular invariants of interest are the size $\mathcal{E}$, number of triangles τ, scan Ψ, and max degree Δ. The scan statistics Ψ are introduced and applied to the problem of detecting chatter anomalies in [22]. The number of triangles is the simplest graph invariant based on clique counts. One can easily define test statistics

The use of various graph invariants as test statistics was considered in [23] and it was shown there, via Monte Carlo analysis, that no single invariant is uniformly most powerful.

Let (G, ϕ) be an attributed graph. For a given $\{u, v\} \in \binom{V}{2}$, we define a $\Gamma_{uv} \in \mathbb{R}^K$ by $\Gamma_{uv}(k) = \mathbb{I}\{\phi(\{u,v\}) = k\}$ where $\mathbb{I}\{\cdot\}$ is the indicator function. Thus $\Gamma_{uv} = \mathbf{0} = (0, 0, \dots, 0)$ unless $\{u, v\} \in E$. Under the independent edges assumption, the Γ_{uv} are independent. We consider linear attribute fusion of graph invariants with parameter $\lambda \in \mathbb{R}^K$ via

$$\mathcal{E}_\lambda(G) = \sum_{\{u,v\} \in \binom{V}{2}} \langle \lambda, \Gamma_{uv} \rangle \tag{10}$$

$$\tau_\lambda(G) = \sum_{\{u,v,w\} \in \binom{V}{3}} \langle \lambda, \Gamma_{uv} \rangle \langle \lambda, \Gamma_{uw} \rangle \langle \lambda, \Gamma_{vw} \rangle \tag{11}$$

$$\Delta_\lambda(G) = \max_{v \in V} \sum_{w \in N(v)} \langle \lambda, \Gamma_{vw} \rangle \tag{12}$$

$$\Psi_\lambda(G) = \max_{v \in V} \sum_{u,w \in N[v]} \langle \lambda, \Gamma_{uw} \rangle. \tag{13}$$

where $N(v) = \{u \colon u \sim v\}$ is the set of neighbors of v and $N[v] = v \cup N(v)$.

Let $\{G(t)\}$ be a time series of graphs. Let $J_\lambda(t)$ be a statistic for $G(t)$ of the form as in Eq. (10) through Eq. (13). We define the running average $\bar{J}_\lambda^l(t)$ of J_λ as

$$\bar{J}_\lambda^l(t) = \frac{1}{l} \sum_{s=1}^{l} J_\lambda(t - s) \tag{14}$$

where $l \in \mathbb{N}$ specified the width of the running-average window. Our main interest is in the normalized fusion statistic $T_\lambda^l(t)$ as depicted in Fig. 3, namely

$$T_\lambda^l(t) = \frac{J_\lambda(t) - \bar{J}_\lambda^l(t)}{\sqrt{\frac{1}{l-1} \sum_{s=1}^{l} (J_\lambda(t-s) - \bar{J}_\lambda^l(t))^2}} \tag{15}$$

for $l \geq 2$. The choice of values for l depends on the practitioner, in that small values of l could induce large fluctuations in the temporal standardization statistics, which might lead to a higher false positive rate while large values of l, on the other hand, could induce over-smoothing, leading to a higher false negative rate.

We are interested in finding the optimal fusion parameter λ, i.e., we want to find a λ such that the power of the test using T_λ^l is maximized. The theoretical results in § IV will be useful for this purpose. We note that T_λ^l is scale invariant in λ for all of our graph invariants. It was noted in [1] that the maximum of the power for scale invariant test statistics is attainable in the set $\{\lambda \colon \|\lambda\| = 1\}$.

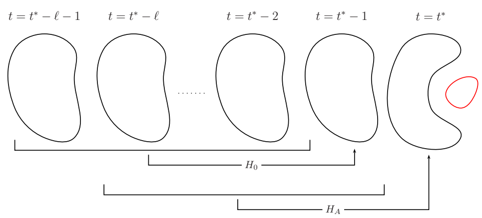


Fig. 3: Temporal standardization: when testing for change at time t, the recent past (graphs $G(t-l), \dots, G(t-1)$) is used to standardize the invariants

IV. Power estimates

§III described the framework of hypothesis testing for anomaly/change-point detection in a time-series of graphs. In particular, it was assumed that the graphs were generated according to a random dot process model with change-parameters $(t^*, m, \pi_0, \pi_1, \mathbf{Q}_0, \mathbf{Q}_1)$ and the test statistics are temporal standardization of several graph invariants, namely size, max degree, scan, and number of triangles.

We derive the limiting distributions for the above-mentioned test statistics in this section. As we mentioned earlier, we will assume that the number of vertices n is large so that the use of the first order approximation $\bar{\mathscr{V}}$ of $\mathscr{V}$ is appropriate for studying the asymptotic theory for the above-mentioned graphs invariants. We strive to obtain approximations for the power in testing $t^* > t$ against $t^* = t$ using $T_\lambda^l(t)$ and from these approximations, find the λ that maximize the power. The limiting distribution for graph invariants in the *unattributed* setting are well-known, see e.g., [26] for max-degree, [24] for subgraph counts, and [27] for scan. The results for the attributed and temporal standardization settings are thus straightforward extensions of their counterparts in the unattributed settings.

We now introduce additional notation that will be used later on in the paper. First, let π_{00}, π_{01}, and π_{11} be sub-probability vectors whose components are given by

$$\begin{aligned} \pi_{00}(k) &= \pi_0(k)\pi_0(k) \\ \pi_{01}(k) &= \pi_0(k)\pi_1(k) \\ \pi_{11}(k) &= \pi_1(k)\pi_1(k) \end{aligned}$$

where $k \in [K]$. We also let η_{00}, η_{01}, and η_{11} be matrices of size $K \times K$ defined by

$$\begin{aligned} \eta_{00} &= \operatorname{diag}(\pi_{00}) - \pi_{00}\pi_{00}^T \\ \eta_{01} &= \operatorname{diag}(\pi_{01}) - \pi_{01}\pi_{01}^T \\ \eta_{11} &= \operatorname{diag}(\pi_{11}) - \pi_{11}\pi_{11}^T. \end{aligned}$$

A. Power estimates for $\mathcal{E}_\lambda$

It was shown in [1] that if $J_\lambda(t) = \mathcal{E}_\lambda(G(t))$, then $T_\lambda^l(t)$ follows a t-distribution with $l-1$ degrees of freedom in the limit. Specifically, we have the following results

Theorem 1. *Let $t \in \{l+1, \dots, t^*\}$ and $\lambda \in \mathbb{R}^K$. Define the vector ζ and the matrix ξ by*

$$\begin{aligned} \zeta &= \tbinom{m}{2}(\pi_{11} - \pi_{00}) + (n-m)m(\pi_{01} - \pi_{00}) \\ \xi &= \tfrac{l+1}{l}\tbinom{n}{2}\eta_{00} + \tbinom{m}{2}(\eta_{11} - \eta_{00}) + (n-m)m(\eta_{01} - \eta_{00}). \end{aligned}$$

Suppose that ξ is positive definite. Define the random variable $\psi_\lambda^l(t)$ by

$$\psi_\lambda^l(t) = \begin{cases} \sqrt{\frac{l}{l+1}} T_\lambda^l(t) & \text{if } t < t^* \\ \sqrt{\frac{\langle \lambda, \binom{n}{2}\eta_{00}\lambda\rangle}{\langle \lambda, \xi\lambda\rangle}} T_\lambda^l(t) & \text{if } t = t^*. \end{cases} \tag{16}$$

As $n \to \infty$, $\psi_\lambda^l(t)$ converges weakly to the Student t-distribution with $l-1$ degrees of freedom and non-centrality parameter μ_λ, where

$$\mu_\lambda = \begin{cases} 0 & \text{if } t < t^* \\ \frac{\langle \lambda, \zeta\rangle}{\sqrt{\langle \lambda, \xi\lambda\rangle}} & \text{if } t = t^* \end{cases} \tag{17}$$

We are interested in finding the λ that will maximize the power of the test. The power approximation is determined by various factors but the dominating factor is μ_λ (for large n). The following corollary extends a result in [1] for $K = 2$ to the case where $K \geq 2$.

Corollary 2. *Let ζ and ξ be as defined in Theorem 1. Suppose that ξ is also positive definite. Let ν^* be the normalized eigenvector corresponding to the largest eigenvalue of $\xi^{-1/2}\zeta\zeta^T\xi^{-1/2}$, i.e.,*

$$\nu^* = \operatorname*{argmax}_{\nu \colon \|\nu\|=1} \nu^T \xi^{-1/2}\zeta\zeta^T\xi^{-1/2}\nu. \tag{18}$$

Then $\lambda^* = \frac{\xi^{-1/2}\nu^*}{\|\xi^{-1/2}\nu^*\|}$ satisfies

$$\operatorname*{argmax}_{\|\lambda\|=1} \mu_\lambda \cap \{\lambda^*, -\lambda^*\} \neq \emptyset \tag{19}$$

$$\operatorname*{argmin}_{\|\lambda\|=1} \mu_\lambda \cap \{\lambda^*, -\lambda^*\} \neq \emptyset. \tag{20}$$

B. Power estimates for τ_λ

The limiting distribution for the number of triangles in unattributed random graphs was considered in [24] for the Erdös-Renyi and in [25] for the kidney-egg model. We note here the small changes that allow us to extend the results in [24], [25] to our attributed graphs model. Let $Y_{uv} = \langle \lambda, \Gamma_{uv} \rangle$ for $\{u,v\} \in \binom{V}{2}$. We can now write $\tau_\lambda(G(t))$ as

$$\frac{\tau_\lambda(G(t))}{\binom{n}{3}} = \binom{n}{3}^{-1} \sum_{\{u,v,w\} \in \binom{V}{3}} Y_{uv} Y_{uw} Y_{vw}; \tag{21}$$

$\tau_\lambda(G(t))/\binom{n}{3}$ is then an U-statistic on $\{Y_e\}$, with kernel function $h(Y_1, Y_2, Y_3) = Y_1 Y_2 Y_3$. By using Hajek's projection method, we can show that $\tau_\lambda(G(t))$ converges to a normal distribution as $n \to \infty$.

Lemma 3. *Let τ_λ^* be the Hajek's projection of τ_λ, i.e.,*

$$\tau_\lambda^*(G) - \mathbb{E}[\tau_\lambda(G)] = \sum_{\{u,v\} \in \binom{V}{2}} \Big(\mathbb{E}[\tau_\lambda(G) \,|\, Y_{uv}] - \mathbb{E}[\tau_\lambda(G)] \Big). \tag{22}$$

For $t < t^*$, we have

$$\begin{aligned} \mathbb{E}[\tau_\lambda^*(G(t))] &= \tbinom{n}{3} \langle \lambda, \pi_{00} \rangle^3 \\ \sqrt{\mathrm{Var}[\tau_\lambda^*(G(t))]} &= (n-2) \langle \lambda, \pi_{00} \rangle^2 \sqrt{\tbinom{n}{2} \langle \lambda, \eta_{00} \lambda \rangle}. \end{aligned}$$

For $t = t^*$, we have

$$\begin{aligned} \mathbb{E}[\tau_\lambda^*(G(t))] &= \tbinom{m}{3} \langle \lambda, \pi_{11} \rangle^3 + \tbinom{m}{2} (n-m) \langle \lambda, \pi_{11} \rangle \langle \lambda, \pi_{01} \rangle^2 \\ &\quad + m \tbinom{n-m}{2} \langle \lambda, \pi_{01} \rangle^2 \langle \lambda, \pi_{00} \rangle \\ &\quad + \tbinom{n-m}{3} \langle \lambda, \pi_{00} \rangle^3 \\ \mathrm{Var}[\tau_\lambda^*(G(t))] &= \tbinom{m}{2} \langle \lambda, \eta_{11} \lambda \rangle S_1^2 + m(n-m) \langle \lambda, \eta_{01} \lambda \rangle S_2^2 \\ &\quad + \tbinom{n-m}{2} \langle \lambda, \eta_{00} \lambda \rangle S_3^2 \end{aligned}$$

where S_1, S_2, and S_3 are given by

$$\begin{aligned} S_1 &= (m-2) \langle \lambda, \pi_{11} \rangle^2 + (n-m) \langle \lambda, \pi_{01} \rangle^2 \\ S_2 &= (m-1) \langle \lambda, \pi_{11} \rangle \langle \lambda, \pi_{01} \rangle + (n-m-1) \langle \lambda, \pi_{00} \rangle \langle \lambda, \pi_{01} \rangle \\ S_3 &= m \langle \lambda, \pi_{01} \rangle^2 + (n-m-2) \langle \lambda, \pi_{00} \rangle^2. \end{aligned}$$

As $n \to \infty$, we have

$$\frac{\tau_\lambda(t) - \mathbb{E}[\tau_\lambda^*(G(t))]}{\sqrt{\mathrm{Var}[\tau_\lambda^*(G(t))]}} \xrightarrow{\mathrm{d}} \mathcal{N}(0,1). \tag{23}$$

From Lemma 3, we can show that the limiting distribution of $T_\lambda^l(t)$ is once again a Student t-distribution with $l-1$ degrees of freedom.

Theorem 4. *Let $t \in \{l+1, \dots, t^*\}$ and $\lambda \in \mathbb{R}^K$. Let $\mu_0 = \mathbb{E}[\tau_\lambda^*(G(t))]$ and $\sigma_0^2 = \mathrm{Var}[\tau_\lambda^*(G(t))]$ for $t < t^*$. Let $\mu_A = \mathbb{E}[\tau_\lambda^*(G(t^*))]$ and $\sigma_A^2 = \mathrm{Var}[\tau_\lambda^*(G(t^*))]$. Define the random variable $\psi_\lambda^l(t)$ by*

$$\psi_\lambda^l(t) = \begin{cases} \sqrt{\frac{l}{l+1}}\, T_\lambda^l(t) & \text{if } t < t^* \\ \sqrt{\frac{l\sigma_0^2}{l\sigma_A^2 + \sigma_0^2}}\, T_\lambda^l(t) & \text{if } t = t^*. \end{cases} \tag{24}$$

As $n \to \infty$, $\psi_\lambda^l(t)$ converges to a Student t-distribution with $l-1$ degrees of freedom and non-centrality parameter μ_λ where

$$\mu_\lambda = \begin{cases} 0 & \text{if } t < t^* \\ \frac{\mu_A - \mu_0}{\sqrt{l\sigma_A^2 + \sigma_0^2}} & \text{if } t = t^*. \end{cases} \tag{25}$$

The power approximation for τ_λ is determined by various factors, but similar to the power approximation for $\mathcal{E}_\lambda$, the dominating factor is μ_λ (for large n). Thus, we are also interested in finding the λ that will maximize μ_λ. However, because of the high power of λ in the expression for μ_λ, an exact solution to $\mathrm{argmax}_\lambda\, \mu_\lambda$ is more challenging for τ than for $\mathcal{E}$.

C. Power estimates for $\Delta_\lambda(t)$

The limiting distribution for maximum degree in unattributed random graphs was considered in [26] for the Erdös-Renyi and in [14] for the kidney-egg model. We note here the necessary changes that allow us to to extend the results in [14], [26] to our attributed graphs model. We denote by $\mathcal{G}(\alpha, \beta)$ the Gumbel distribution with location parameter α and scale parameter β.

Proposition 5. *Let a_n and b_n be functions of n given by*

$$\begin{aligned} a_n &= (2 \log n)^{1/2} \Big(1 - \frac{\log \log n + \log 4\pi}{4 \log n} \Big) \\ b_n &= (2 \log n)^{-1/2}. \end{aligned}$$

Then as $n \to \infty$ and $m = \Omega(\sqrt{n \log n})$, we have

$$\frac{\Delta_\lambda(t) - a_n \sigma_0 - \mu_0}{b_n \sigma_0} \xrightarrow{\mathrm{d}} \mathcal{G}(0,1) \quad \text{for } t < t^* \tag{26}$$

$$\frac{\Delta_\lambda(t) - a_m \sigma_A - \mu_A}{b_m \sigma_A} \xrightarrow{\mathrm{d}} \mathcal{G}(0,1) \quad \text{for } t = t^* \tag{27}$$

where

$$\begin{aligned} \mu_0 &= (n-1) \langle \lambda, \pi_{00} \rangle \\ \sigma_A &= \sqrt{(n-1) \langle \lambda, \eta_{00} \lambda \rangle} \\ \mu_0 &= (m-1) \langle \lambda, \pi_{11} \rangle + (n-m) \langle \lambda, \pi_{01} \rangle \\ \sigma_A &= \sqrt{(m-1) \langle \lambda, \eta_{11} \lambda \rangle + (n-m) \langle \lambda, \eta_{01} \lambda \rangle}. \end{aligned}$$

$T_\lambda^l(t)$ based on the Gumbel distributed $\Delta_\lambda(t)$, in contrast to $T_\lambda^l(t)$ based on the normally distributed $\mathcal{E}_\lambda$ and $\tau_\lambda(t)$, does not have a simple distribution for small or moderate values of l. The following result give the limiting distribution for $T_\lambda^l(t)$ under the assumption that l is sufficiently large.

Theorem 6. *Let $t \in \{l+1, \dots, t^*\}$ and $\lambda \in \mathbb{R}^K$. For sufficiently large n and sufficiently large l, the variable $T_\lambda^l(t)$ has approximately a $\mathcal{G}(\rho_\lambda, \varsigma_\lambda)$ distribution with*

$$\rho_\lambda = \begin{cases} -\frac{\sqrt{6}}{\pi} \gamma & \text{if } t < t^* \\ \frac{\sqrt{6}}{\pi} \frac{\mu_A - \mu_0 + a_m \sigma_A - (a_n + b_n \gamma) \sigma_0}{b_n \sigma_0} & \text{if } t = t^* \end{cases} \tag{28}$$

$$\varsigma_\lambda = \begin{cases} \frac{\sqrt{6}}{\pi} & \text{if } t < t^* \\ \frac{\sqrt{6}}{\pi} \frac{b_m \sigma_A}{b_n \sigma_0} & \text{if } t = t^* \end{cases} \tag{29}$$

where $\gamma \approx 0.57721$ is the Euler-Mascheroni constant.

We are interested in finding the λ that will maximize the power of the test using Δ_λ. The power approximation is determined by various factors, but the dominating factor is $\frac{\rho_\lambda}{\varsigma_\lambda}$. For sufficiently large n and m, we have for $t = t^*$,

$$\begin{aligned}\frac{\rho_\lambda}{\varsigma_\lambda} &= (1+o(1))\frac{\mu_A - \mu_0}{b_m \sigma_A} \\ &= (1+o(1))\frac{\langle \lambda, (m-1)\pi_{11} + (n-m)\pi_{01} - (n-1)\pi_{00}\rangle}{b_m\sqrt{\langle \lambda, ((m-1)\eta_{11} + (n-m)\eta_{01})\lambda\rangle}}.\end{aligned}$$

We can thus find the maximum and minimum of $\frac{\rho_\lambda}{\varsigma_\lambda}$ by solving an eigenvalue problem as in Corollary 2.

D. Power estimates for $\Psi_\lambda(t)$

The limiting distribution for the scan statistics in unattributed random graphs was considered in [27]. We note here the changes that allow us to extend the results in [27] to our attributed graphs model. Let z_n be defined, for $n \in \mathbb{N}$, by

$$z_n = \sqrt{2\log n}\Big(1 - \frac{\log\log n + \log 4\pi}{4\log n}\Big). \quad (30)$$

Let us also define two random variables E and F by

$$\begin{aligned}E &\sim \mathrm{Bin}(m-1, \langle 1, \pi_{11}\rangle) \\ F &\sim \mathrm{Bin}(n-m, \langle 1, \pi_{01}\rangle).\end{aligned}$$

Let $E+F$ be the convolution of E and F. Denote by μ_{E+F} and σ^2_{E+F} the mean and variance of $E+F$. Let $N_\kappa = \mu_{E+F} + z_m\sigma_{E+F}$. We then have the following results.

Lemma 7. *Let $\Psi_\lambda(t)$ be the scan statistic for $t < t^*$. Let N_0, a_n, and b_n be defined by*

$$\begin{aligned}\mu_0 &= (n-1)\langle 1, \pi_{00}\rangle \\ \sigma_0 &= \sqrt{(n-1)(\langle 1, \pi_{00}\rangle - \langle 1, \pi_{00}\rangle^2)} \\ N_0 &= \mu_0 + z_n\sigma_0 \\ a_n &= \langle \lambda, \pi_{00}\rangle\tbinom{N_0}{2} \\ b_n &= \langle \lambda, \pi_{00}\rangle N_0 \frac{\sigma_0}{\sqrt{2\log n}}.\end{aligned}$$

Then we have

$$\frac{\Psi_\lambda(t) - a_n}{b_n} \xrightarrow{\mathrm{d}} \mathcal{G}(0,1). \quad (31)$$

Lemma 8. *Let $a_{n,m}$ and $b_{n,m}$ be given by*

$$\begin{aligned}a_{n,m} = \langle \lambda, \pi_{00}\rangle\tbinom{N_\kappa}{2} + \langle \lambda, \pi_{11} - \pi_{00}\rangle\tbinom{\mu_E}{2} \\ + \langle \lambda, \pi_{01} - \pi_{00}\rangle \mu_E\mu_F\end{aligned} \quad (32)$$

$$b_{n,m} = \langle \lambda, \pi_{00}\rangle N_\kappa \frac{\sigma_{E+F}}{\sqrt{2\log m}}. \quad (33)$$

If $m = \Omega(\sqrt{n\log n})$ and $m = O(n^{k/(k+1)})$ for some $k \in \mathbb{N}$ then

$$\frac{\Psi_\lambda(t^*) - a_{n,m}}{b_{n,m}} \xrightarrow{\mathrm{d}} \mathcal{G}(0,1). \quad (34)$$

The next result is analogous to Theorem 6 and gives the limiting distribution for $T^l_\lambda(t)$ based on Ψ_λ for the case of large l.

Theorem 9. *Let $t \in \{l+1, \dots, t^*\}$ and $\lambda \in \mathbb{R}^K$. For sufficiently large n and sufficiently large l, $T^l_\lambda(t)$ has approximately a $\mathcal{G}(\rho_\lambda, \varsigma_l)$ distribution with*

$$\rho_\lambda = \begin{cases} -\frac{\sqrt{6}}{\pi}\gamma & \text{if } t < t^* \\ \frac{\sqrt{6}}{\pi}\frac{a_{n,m} - a_n - b_n\gamma}{b_n} & \text{if } t = t^* \end{cases} \quad (35)$$

$$\varsigma_\lambda = \begin{cases} \frac{\sqrt{6}}{\pi} & \text{if } t < t^* \\ \frac{\sqrt{6}}{\pi}\frac{b_{n,m}}{6b_n} & \text{if } t = t^*. \end{cases} \quad (36)$$

The dominating factor in the power approximation for Ψ_λ is $\frac{\rho_\lambda}{\varsigma_\lambda}$. For $t = t^*$ and sufficiently large n and l, we have

$$\begin{aligned}\frac{\rho_\lambda}{\varsigma_l} &= \frac{a_{n,m} - a_n - b_n\gamma}{b_{n,m}} \\ &= (1+o(1))\frac{\langle \lambda, \xi\rangle}{\langle \lambda, N_\kappa \frac{\sigma_{E+F}}{\sqrt{2\log n}}\pi_{00}\rangle}\end{aligned} \quad (37)$$

where ξ is given by

$$\xi = \Big(\tbinom{N_\kappa}{2} - \tbinom{N_0}{2} - \tbinom{\mu_E}{2} - \mu_E\mu_F\Big)\pi_{00} + \tbinom{\mu_E}{2}\pi_{11} + \mu_E\mu_F\pi_{10}.$$

From Eq. (37), we see that there exists (as $n \to \infty$) a λ that maximizes $\frac{\rho_\lambda}{\varsigma_\lambda}$ and is on the boundary, i.e., $\lambda_k \neq 0$ for exactly one k. Therefore the asymptotic theory for scan statistics, in contrast with the other graph invariants that were considered, indicates that there may be no benefits in fusing attributes. However, as Fig. 6c shows, this phenomenon does not hold in general for moderate values of n. Similar observations about potential inaccuracies in using asymptotic results to predict finite but large samples behavior in testing using graph invariants were discussed in [14], [28].

V. Inference Examples

We present here two inference examples, one on simulated data and the other on the Enron email data set. In the first inference example, we simulated data from the model in § II with the following parameters:

$$K = 2, n = 100, m = 9, l = 10,$$

transition matrices

$$\mathbf{Q}_0 = \begin{bmatrix} -\frac{2}{3} & \frac{1}{6} & \frac{1}{2} \\ 1 & -1 & 0 \\ -1 & 0 & -1 \end{bmatrix}, \quad \mathbf{Q}_1 = \begin{bmatrix} -\frac{13}{7} & \frac{5}{7} & \frac{8}{7} \\ 1 & -1 & 0 \\ 1 & 0 & -1 \end{bmatrix},$$

and stationary probability vectors

$$\pi_0 = (0.10, 0.30, 0.60)^T, \quad \pi_1 = (0.25, 0.40, 0.35)^T.$$

The change-point t^* for this example is $t^* = 11$. Power estimates for our attribute fusion statistics for this example are presented in Fig. 4 and Fig. 6. We consider the asymptotic, first approximation, second approximation, and exact model power estimates. Fig. 4 shows power as a function of the vertex process parameter r at a specific point, namely at $\lambda^* = \mathrm{argmax}_\lambda\, \beta(\lambda)$. Fig. 6 shows power as a function of angle θ, where $\lambda = (cos(\theta), sin(\theta))$. Finally, Fig. 5 presents Q-Q plots illustrating convergence of the temporal standardization statistics to the limiting distributions given in § IV.

A comment should be made regarding the application of the theoretical results in § IV to this simulation example. The results in § IV on the limiting distributions of the temporal standardization statistics specified both the type of distribution, i.e., Gumbel or Student's t-distribution, along with the associated parameters, i.e., location and scale for the Gumbel distribution and degrees of freedom and non-centrality parameter for the Student's t-distribution. Empirical investigations demonstrate that, for the temporal standardization statistics based on max degree, number of triangles, and scan, the limiting distributions with the parameters as specified in Theorem 4, Theorem 6 and Theorem 9 are accurate large-sample approximations for these statistics provided that the graphs have $n \geq 1000$ and $m \geq 50$. The variance parameters as specified by the theoretical result require correction for small n, as used in our experiment ($n = 100$). For example, $\mathrm{Var}[\tau^*_\lambda(G(t))]$ in Lemma 3 is a good approximation of the variance term in the normal approximation only if the ratio of $\mathrm{Var}[\tau^*_\lambda(G(t))] - \mathrm{Var}[\tau_\lambda(G(t))]$, the difference in the variance between the Hajek's projection and the true variance, to $\mathrm{Var}[\tau_\lambda(G(t))]$ is small; this does not hold for $n = 100$. Thus, the results for the limiting distributions for the above experiment are obtained by correcting for the variance. In particular, we used the sample variance of the graph invariants under the first-order approximation as the variance for the limiting distribution.

The main implication that can be inferred from Fig. 4, Fig. 5 and Fig. 6 is that inferential performance in the mathematically tractable first-order and second-order approximation models does provide guidance for methodological choices applicable to the exact (realistic but intractable) model. Furthermore, to the extent that the exact model is realistic, we may tentatively conclude that approximation model investigations have some bearing on real data applications.

Our second example uses the Enron data set which consists of email messages between 184 executives of the Enron corporation during a time period from 1998 to 2002. The Enron email messages for most of the 2001 calendar year have been annotated into 32 different topics [29] and thus form the basis of our analysis. We chose to condense these 32 different topics into two groups. The first group consists of general energy-related topics and the second group consists of topics that are specifically related to the Enron corporation. From the collection of email messages, we construct a time series of graphs. Each graph consists of email messages sent during a week in 2001. The vertices of the graphs represent the executives and the edges represent email communication between the executives. The attributes on the edges are given by the grouping of the email topics as mentioned above.

The detection of a chatter anomaly in the Enron data set during the first week of May 2001 ($t = 152$) was previously reported in [22], but with unattributed edges. We chose to investigate the effect of attribute fusion for that same week in our inference example. Fig. 7 presents a plot of the normalized

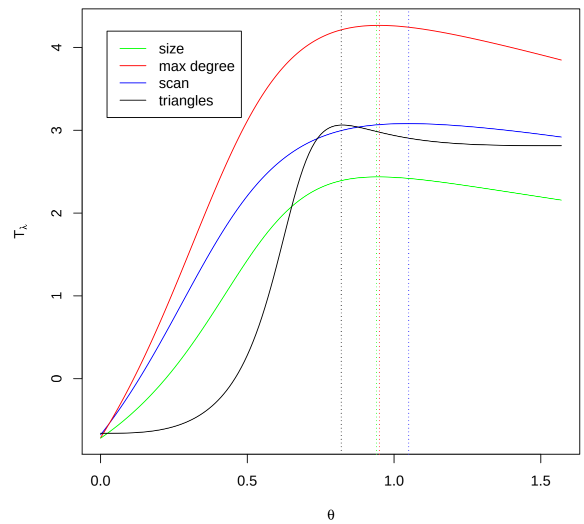


Fig. 7: Detecting chatter anomaly in the Enron time series of graphs. The plot shows the normalized statistics $T^l_\lambda(t)$ as a function of the fusion parameter $\lambda = (\cos(\theta), sin(\theta))$ for $l = 20$ and $t = 152$. The graph in question ($t = 152$) corresponds to the first week of May 2001. The dashed lines correspond to the θ^* for the different graph invariants ($\theta^*_{\mathcal{E}_\lambda} \approx \theta^*_{\Delta_\lambda} \approx 0.96$, $\theta^*_{\Psi_\lambda} \approx 1.06$ and $\theta^*_{\tau_\lambda} \approx 0.83$). Attribute fusion provides superior detection.

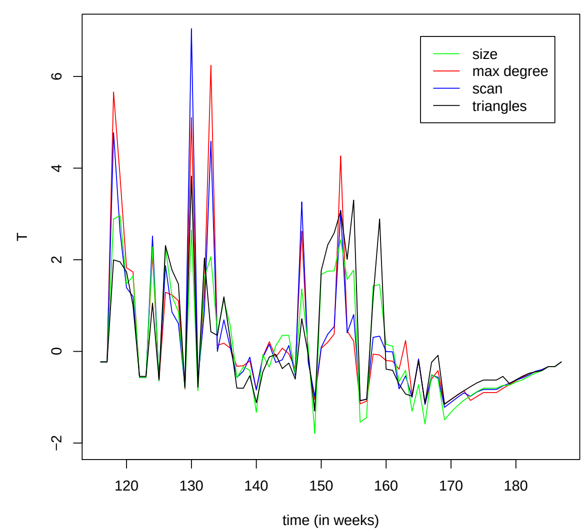
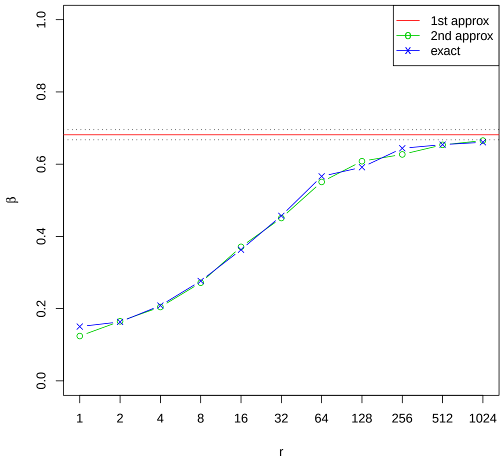
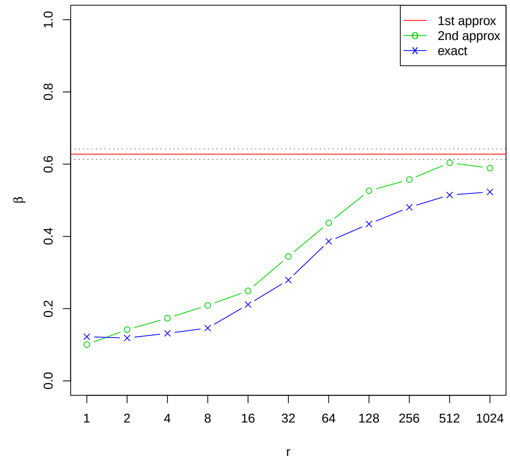


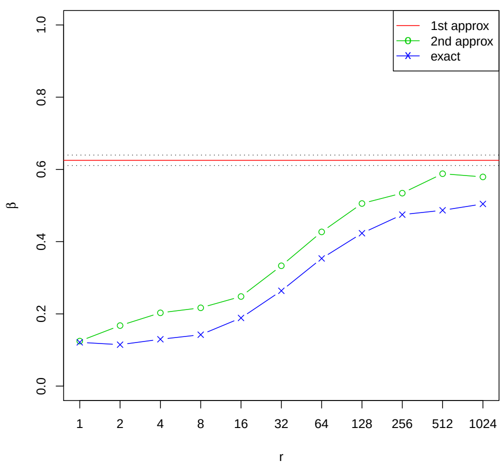
Fig. 8: The fluctuations of $T^l_\lambda(t)$ over time for the four graph invariants. The value of $\lambda = (\cos(\theta), sin(\theta))$ for each graph invariants correspond to the optimal fusion parameter for the detection as illustrated in Fig. 7 and they are $\theta^*_{\mathcal{E}_\lambda} \approx \theta^*_{\Delta_\lambda} \approx 0.96$, $\theta^*_{\Psi_\lambda} \approx 1.06$ and $\theta^*_{\tau_\lambda} \approx 0.83$. The fluctuations indicate that there are possible anomalies around the weeks 120, 132, and 146, which were also reported and analyzed in [22].



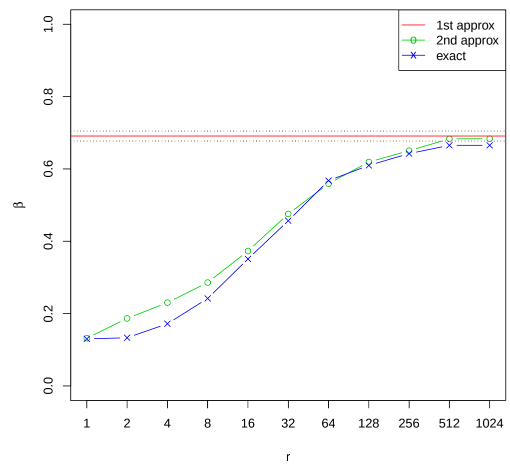
(a) $\mathcal{E}_\lambda$



(b) Δ_λ



(c) τ_λ



(d) Ψ_λ

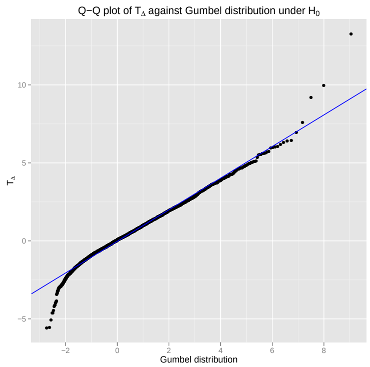
Fig. 4: Power $\beta(\theta_r^*, r)$ at $\theta_r^* = \operatorname{argmax}_\theta \beta(\theta, r)$ as a function of r for the four invariants at test size $\alpha = 0.05$. The horizontal lines represents first-order approximation $\pm$ three standard deviations, and the two curves represent the second approximation (green) and exact model (blue). The 10000 Monte Carlo replicates yield standard deviations not exceeding 0.005 for the power estimates . The second approximation results match well with the exact model results, and both match well for large r with the first-order approximation results.

statistic $T_\lambda^l(t)$ at $t = 152$ with $l = 20$ for the four graph invariants against the fusion parameter $\lambda = (cos(\theta), sin(\theta))$. Fig. 8 presents a plot of the normalized statistics $T_\lambda^l(t)$ with $l = 20$ for the four graph invariants over the time interval corresponding to the weeks $t = 96$ through $t = 168$. The value of λ used for each graph invariants is the λ that maximizes $T_\lambda^l(152)$. Fig. 7 indicates that the fusing of attributes leads to better detection as exemplified by the fact that the maximum for each of the normalized statistics T_λ occurs for $\theta \notin \{0, \pi/2\}$. Furthermore, the optimal fusion parameters are dependent on the specific graph invariants.

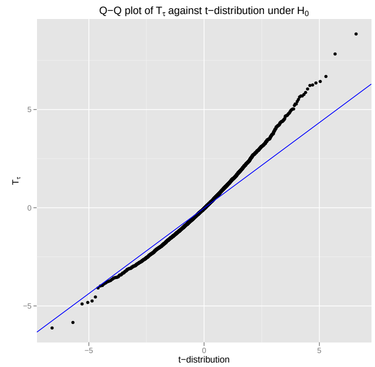
VI. Conclusions

We have presented an analysis of change-point detection for time series of attributed graphs through the use of test statistics which are based on linear attribute fusion of some graph invariants. We derived the limiting distribution of these test statistics under the assumptions that n, the number of vertices, is sufficiently large and r, the process parameter rate, is also sufficiently large. The limiting distribution for these test statistics are then used to derive estimates for the power of the tests.

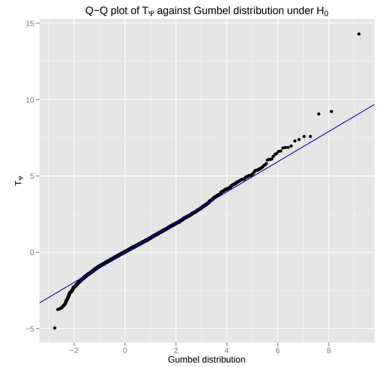
The simulation experiment in §V indicates that the power estimates are accurate, even for the moderate value of $n = 100$.



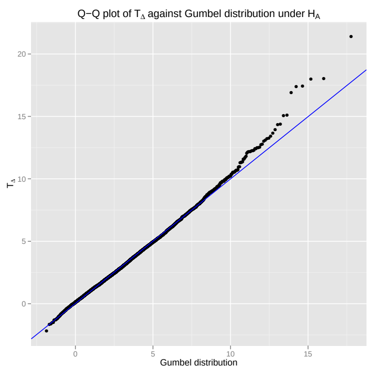
(a) T_Δ under H_0



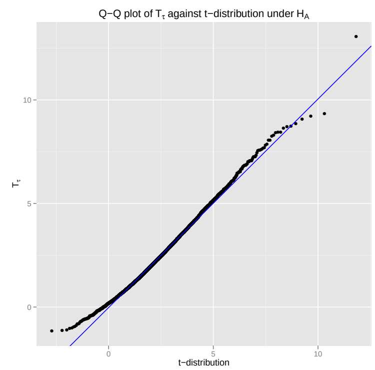
(b) T_τ under H_0



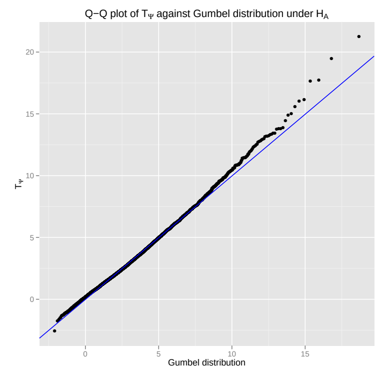
(c) T_ψ under H_0



(d) T_Δ under H_A



(e) T_τ under H_A



(f) T_ψ under H_A

Fig. 5: Q-Q plots of the temporal standardization statistics T_Δ, T_τ and T_Ψ against the corresponding limiting distributions as given in Theorem 4, Theorem 6 and Theorem 9, respectively. The sample quantile for the temporal standardization statistics are based on 10000 Monte Carlo replicates of the time-series of graphs. The graphs are generated according to the second-order approximation with $r = 1024$. Refer to the text at the beginning of § V for details on the parameters of the random graph model. The fusion parameter $\lambda = (cos(\theta), sin(\theta))$ for each temporal standardization statistics correspond to the θ^* that maximizes the power as depicted in Fig. 6, i.e., $\theta^*_\Delta \approx 0.32$, $\theta^*_\tau \approx 0.3$, $\theta^*_\psi \approx 0.14$. For each plot, a linear regression between the sample quantile and the theoretical quantile was performed and the resulting best linear fit is plotted. The R^2 statistics for the linear regressions exceed 0.995 in all cases.

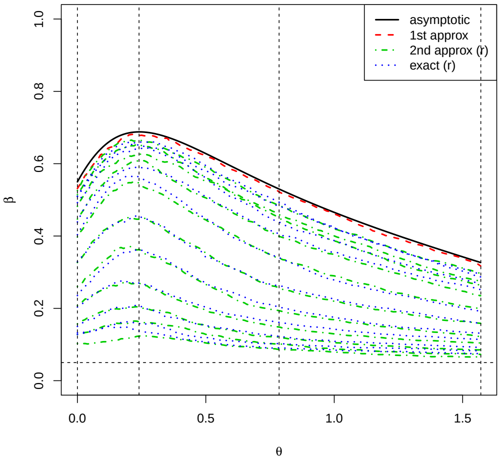
Furthermore, it was also indicated that the optimal linear fusion parameter depends on the graph invariant considered. In particular, the results depicted in Fig. 6 yield $\theta^*_{\mathcal{E}_\lambda} \approx 0.24$, $\theta^*_{\Delta_\lambda} \approx 0.32$, $\theta^*_{\Psi_\lambda} \approx 0.14$ and $\theta^*_{\tau_\lambda} \approx 0.38$. These optimal fusion parameter differences are statistically significant, and combining this result with the "no uniformly most powerful invariant" result [23], we conclude that optimal linear attribute fusion theory requires significant additional development. Toward this end, the approximation models from [1] promise to be of assistance.

Hypothesis testing on time series of attributed graphs has applications in diverse areas, e.g., social network analysis (wherein vertices represent individual actors or organizations), connectome inference (wherein vertices are neurons or brain regions) and text processing (wherein vertices represent authors or documents). These and many other applications may benefit from generalizations of our results and the model in [1] to directed, multi, and weighted graphs, to other class of graph invariants, and to consideration of inference with errorful edge attributes through an attribute confusion matrix. For example, the paper considered a test statistic based on the number of triangles which is the simplest clique counts statistic. One can define attribute fusion for k-clique counts for $k \geq 4$. Using the theory of U-statistics for subgraphs counts (cf. [24]), the limiting distribution for the temporal standardization statistics based on these k-clique counts can be derived in an analogous manner to the case for $k = 3$ in this paper.

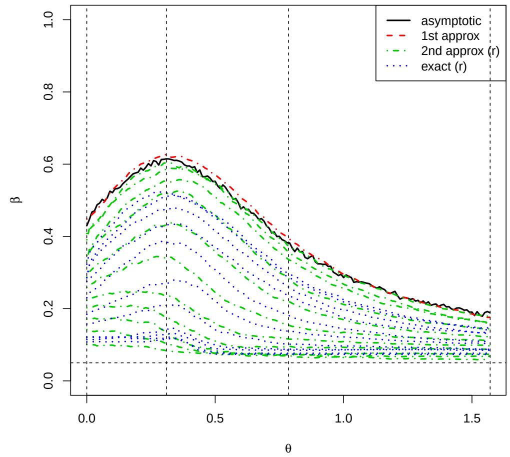
Finally, we remark that applicability of the statistics described in § III extends beyond the statistical inference perspective. Indeed, temporal standardization of graph invariants can be computed irrespective of any underlying model on the graphs. As such, these statistics can serve as the basis for simple and robust inference procedures on time-series of (attributed) graphs.

References

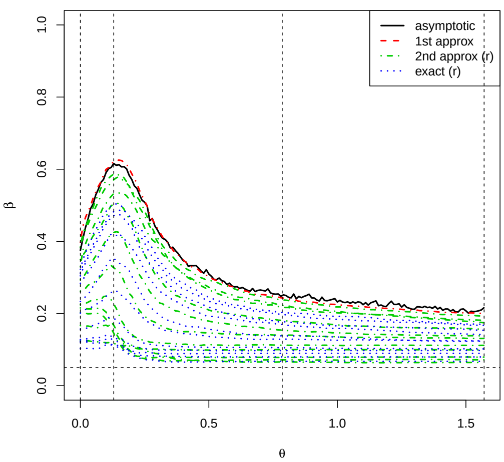
[1] N. H. Lee and C. E. Priebe, "A latent process model for time series of attributed random graphs," *Statistical Inference for Stochastic Processes*,



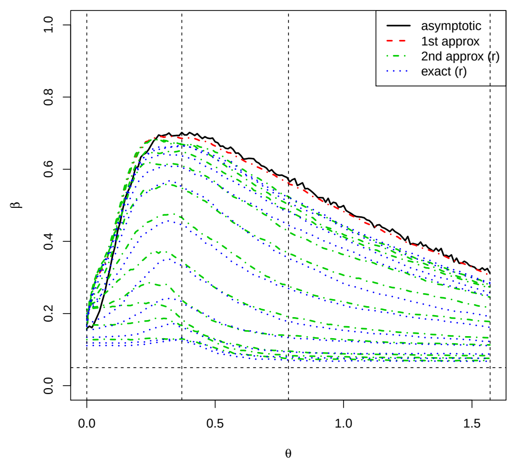
(a) $\mathcal{E}_\lambda$



(b) Δ_λ



(c) Ψ_λ



(d) τ_λ

Fig. 6: Power β as a function of angle θ for $\lambda = (cos(\theta), sin(\theta))$. The plot shows analytic asymptotic result (black), and first approximation (red), second approximation (green), and exact model (blue) estimates for the various invariants via Monte Carlo. The multiple green and blue lines correspond to different values of the vertex process rate r (here $r \in \{2^0, 2^1, \ldots, 2^{10}\}$). The 10000 Monte Carlo replicates yields standard deviations not exceeding 0.005 for the power estimates. The four vertical lines in each plot correspond to $\theta \in \{0, \pi/2, \pi/4, \theta^*\}$. Because we are considering a one-sided test and because $\pi_{1,k} > \pi_{0,k}$ for both $k = 1$ and $k = 2$, the power is maximized in the first quadrant, i.e., $\theta \in (0, \pi/2)$. The second approximation results match well with the exact model results, and both match well for large r with the first order approximation and the asymptotic results. The optimal θ^* is apparently different for the four different graph invariants.

vol. 14, pp. 231–253, 2011.

[2] D. K. Hammond, P. Vandergheynst, and R. Gribonval, "Wavelets on graphs via spectral graph theory," *Applied and Computational Harmonic Analysis*, vol. 30, pp. 129–150, 2010.

[3] R. R. Coifman and M. Maggioni, "Diffusion wavelets," *Applied and Computational Harmonic Analysis*, vol. 21, pp. 53–94, 2006.

[4] M. Crovella and E. D. Kolaczyk, "Graph wavelets for spatial traffic analysis," in *Proceedings of the twenty-second annual joint conference of the IEEE computer and communications societies*, 2003, pp. 1848–1857.

[5] S. K. Narang and A. Ortega, "Perfect reconstruction two-channel wavelet filter-baks for graph structured data," *IEEE Transactions on Signal Processing*, In press.

[6] B. A. Miller, N. T. Bliss, and P. J. Wolfe, "Toward signal processing theory for graphs and non-Euclidean data," in *Proceedings of the 35th IEEE international conference on acoustics, speech, and signal processing*, 2010, pp. 5414–5417.

[7] C. E. Priebe, N. H. Lee, Y. Park, and M. Tang, "Attribute fusion in a latent process model for time series of graphs," in *IEEE workshop on statistical signal processing*, 2011.

[8] P. Erdös and A. Rényi, "On random graphs," *Publicationes Mathematicae*, vol. 6, pp. 290–297, 1959.

[9] P. W. Holland and S. Leinhardt, "An exponential family of of probability distributions for directed graphs," *Journal of the American Statistical Association*, vol. 76, pp. 33–50, 1981.

[10] P. Holland, K. Laskey, and S. Leinhardt, "Stochastic blockmodels: First steps," *Social Networks*, vol. 5, no. 2, pp. 109–137, 1983.

[11] Y. J. Wang and G. Y. Wong, "Stochastic Blockmodels for Directed Graphs," *Journal of the American Statistical Association*, vol. 82, no. 397, pp. 8–19, 1987.

[12] P. Hoff, A. E. Raftery, and J. M. Tantrum, "Latent space approaches to social network analysis," *Journal of the American Statistical Association*, vol. 97, pp. 1090–1098, 2002.

[13] S. Young and E. Scheinerman, "Random dot product models for social networks," in *Proceedings of the 5th international conference on algorithms and models for the web-graph*, 2007, pp. 138–149.

[14] A. Rukhin and C. E. Priebe, "A comparative power analysis of the maximum degree and size invariants for random graph inference," *Journal of Statistical Planning and Inference*, vol. 141, pp. 1041–1046, 2011.

[15] E. Airoldi, D. Blei, S. Fienberg, and E. Xing, "Mixed membership stochastic blockmodels," *Journal of Machine Learning Research*, vol. 9, pp. 1981–2014, 2008.

[16] S. Asmussen, *Applied Probability and Queues*. Springer-Verlag, 2003.

[17] D. J. Marchette and C. E. Priebe, "Predicting unobserved links in incompletely observed networks," *Computational Statistics & Data Analysis*, vol. 52, pp. 1373–1386, 2008.

[18] E. R. Scheinerman and K. Tucker, "Modeling graphs using dot product representations," *Computational Statistics*, vol. 25, pp. 1–16, 2010.

[19] D. L. Sussman, M. Tang, D. E. Fishkind, and C. E. Priebe, "A consistent dot product embedding for stochastic blockmodel graphs," *Submitted*, August 2011. [Online]. Available: http://arxiv.org/abs/1108.2228

[20] N. Heard, D. Weston, K. Platanioti, and D. Hand, "Bayesian anomaly detection methods for social networks." *Annals of Applied Statistics*, vol. 4, pp. 645–662, 2010.

[21] P. Perry and P. J. Wolfe, "Point process modeling for directed interaction networks," *Submitted*, 2010. [Online]. Available: http://arxiv.org/abs/1011.1703v2

[22] C. E. Priebe, J. M. Conroy, D. J. Marchette, and Y. Park, "Scan statistics on Enron graphs," *Computational and Mathematical Organization Theory*, vol. 11, pp. 229–247, 2005.

[23] H. Pao, G. A. Coppersmith, and C. E. Priebe, "Statistical inference on random graphs: Comparative power analyses via Monte Carlo," *Journal of Computational and Graphical Statistics*, vol. 20, pp. 395–416, 2011.

[24] K. Nowicki and J. C. Wierman, "Subgraph counts in random graphs using incomplete U-statistics methods," *Discrete Mathematics*, vol. 72, pp. 299–310, 1988.

[25] A. Rukhin, "Asymptotic analysis of various statistics for random graph inference," Ph.D. dissertation, Johns Hopkins University, 2009.

[26] B. Bollobás, *Random Graphs*. Academic Press, 1985.

[27] A. Rukhin and C. E. Priebe, "On the limiting distribution of a graph scan statistic," *Communications in Statistics: Theory and Methods*, vol. 41, pp. 1151–1170, 2012.

[28] C. E. Priebe, G. C. Coppersmith, and A. Rukhin, "You say graph invariant, I say test statistic." *Statistical Computing and Graphics Newsletter*, vol. 21, 2010.

[29] M. W. Berry, M. Browne, and B. Signer, *2001 Topic Annotated Enron Email Data Set*. Philadelphia: Linguistic Data Consortium, 2001.