A latent process model for time series of attributed random graphs

N. H. Lee · C. E. Priebe

Abstract   We introduce a latent process model for time series of attributed random graphs for characterizing multiple modes of association among a collection of actors over time. Two mathematically tractable approximations are derived, and we examine the performance of a class of test statistics for an illustrative change-point detection problem and demonstrate that the analysis through approximation can provide valuable information regarding inference properties.

Keywords   Random graph · Latent position model · Latent process model · Inference · Change point

1 Introduction

1.1 Overview

Statistical analysis of scenarios comprising multiple modes of association among a collection of actors over time is of ever-increasing importance in a wide-ranging array of applications; e.g., communications analysis—who talks to whom, about what, and when. See the recent monograph (Kolaczyk 2009) and the survey papers (Olding and Wolfe submitted; Goldenberg et al. 2010; Porter et al. 2009), e.g., for a unified presentation of statistical models and methods for network data from across various disciplines. Random graph models are commonly used to model association among actors, and attributed random graphs, wherein edge attributes (categorical weights) encode modes of association between vertices, are a natural extension.

In this article, we propose a generative model for association patterns over time which allows for comparative inference analysis. Our model for time series of attributed random

N. H. Lee · C. E. Priebe (✉)
Johns Hopkins University, Baltimore, MD, USA
e-mail: cep@jhu.edu

N. H. Lee
e-mail: nhlee@jhu.edu

graphs on vertex set $\{1, \ldots, n\}$ is derived from n continuous-time finite-state stochastic processes—one for each vertex—taking their values in some (finite) attribute space, which gives rise to a (discrete-time) time series of attributed random graphs. The attribute on edge (i, j) at time $t \in \{1, 2, \ldots\}$ is a function of the (unobserved) states of the processes associated with vertices i and j prior to time t.

Of particular interest are two analytically tractable approximations to the exact model. A first approximation involves a sequence of independent random graphs with independent edges. A second approximation involves a sequence of independent latent position graphs. Through these approximations, comparative analysis of inference procedures may be performed.

We consider for illustration a change-point detection problem in which an unknown subset of the vertex processes change their behavior at some unknown point in time. Observing only the time series of attributed graphs, we consider the problem of testing for the change.

1.2 Outline

This article is organized as follows. Section 2 introduces our latent process model in the context of a change-point detection problem, and two mathematically simpler approximations are defined. Section 3 presents our main approximation results. In Sect. 4, as an illustrative application of our approximation results, we examine the performance of a simple class of test statistics for the change-point detection problem and demonstrate that analysis through approximation can provide valuable information about test statistic behavior.

2 Model for a change-point detection problem

2.1 Change-point detection problem

Fix constants $t_1 \in \{1, 2, 3, \ldots\} \cup \{\infty\}$ and $K \in \mathbb{N}$. For $n \in \mathbb{N}$, let $[n] = \{1, 2, \ldots, n\}$ and consider a time series $\{G(t)\}_{t=1}^{\infty}$ of random graphs, where each $G(t)$ is an attributed random graph $([n]; E(t))$. That is, all random graphs are on the same vertex set $[n]$ and the undirected edge set at time t, denoted by $E(t)$, specifies, for each pair of vertices for which an edge is present, an attribute $k \in \mathcal{K} = \{1, 2, \ldots, K\}$ associated with the pair.

The probabilistic properties of the edge sets $\{E(t)\}_{t=1}^{\infty}$ are determined by a collection of n vertex processes. The state of the collection of vertex processes is unobserved. Prior to time $t_1 - 1$, the latent state of each of the n vertex processes is described by identical probabilistic dynamics (the null dynamics); after time $t_1 - 1$, there is a (fixed but unknown) subset consisting of n_1 vertices whose latent state dynamics are different from the null dynamics (and the remaining $n - n_1$ vertex processes evolve as before).

The constant t_1 is assumed to be unknown but non-random, and t_1 will be said to be the *change-point*. Note that, as a degenerate case, if $t_1 = \infty$, then the all vertex processes evolve according to their null dynamics for all time. Observing only the time-indexed attributed graphs and without knowing directly any history of the vertex processes themselves, the task of testing for the change-point is of general interest, and we will call the task the *change-point detection problem*.

2.2 Random dot product space and random dot process model

We begin by introducing preliminary terminology. Let

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

be the unit simplex in $\mathbb{R}^K$. Motivated by Young and Scheinerman (2009), Scheinerman and Tucker (2010), Kraetzl et al. submitted, and Bollobas et al. 2007, Sect. 16.4, we define a random dot product space for attributed graphs.

Definition 1 *A random dot product space with vertices* $[n]$ *and attributes* $\mathscr{K}$ is a pair (X, G) of random elements such that

(i) $X = \{X_i\}_{i=1}^n$ is a collection of $\mathscr{S}$-valued random vectors,
(ii) G is a random graph with vertices $[n]$ and attributes $\mathscr{K}$ such that, conditioning on the event that $(X_1, \ldots, X_n) = (x_1, \ldots, x_n)$ with $x_i \in \mathscr{S}$ for $i = 1, \ldots, n$, there is an edge between vertex i and vertex j in G with probability $\langle x_i, x_j \rangle = \sum_{k=1}^K x_{i,k} x_{j,k}$ and that edge has attribute $k \in \mathscr{K}$ with probability $x_{i,k} x_{j,k}$, independently of all other edges.

Next, let

$$\mathscr{K}_+ = \{1, \ldots, K + 1\}. \tag{2}$$

We will say that *an r.c.l.l. stochastic process* $W = \{W(u) : u \in [0, \infty)\}$ *with state space* $\mathscr{K}_+^n = \mathscr{K}_+ \times \cdots \times \mathscr{K}_+$ *induces the sequence* $\mathscr{V} = \{(X(t), G(t))\}_{t=1}^\infty$ *of random dot product spaces if*

(i)   each $(X(t), G(t))$ is a random dot product space with vertices $[n]$ and attributes $\mathscr{K}$, where $X_{i,k}(t) = \int_{t-1}^t 1_{\{W_i(u)=k\}} du$ almost surely for each vertex $i \in [n]$, attribute $k \in \mathscr{K}$, and time $t \in \mathbb{N}$,
(ii) for each t and graph g,

$$\mathbf{P}[G(t) = g \mid W(s), s \leq t] = \mathbf{P}[G(t) = g \mid X(t)].$$

Definition 2 Let W be an r.c.l.l. stochastic process with state space $\mathscr{K}_+^n$ and let $\mathscr{V}$ be a sequence of random dot product spaces. The pair $(\mathscr{V}, W)$ is a *random dot process model* if W induces $\mathscr{V}$.

2.3 Change parameter of a random dot process model

A random dot process model is general enough to incorporate various change-point phenomena. While there are various possibilities for extension, as hinted in Remark 1, we will restrict our attention to the class of random dot process models further specified by fixing certain parameters as described in Definition 3.

Definition 3 Let π_0 and π_1 be $(K + 1)$-dimensional probability vectors and let Q_0 and Q_1 be $(K + 1) \times (K + 1)$-dimensional matrices such that $\pi_0^T Q_0 = 0$ and $\pi_1^T Q_1 = 0$. A random dot process model $(\mathscr{V}, W)$ has the change parameter $(t_1, n_1, \pi_0, \pi_1, Q_0, Q_1)$ if

(i)   for each time t and vertex i, there exists a matrix $Q^{(i)}(t)$ such that for each k and $k' \in \mathscr{K}_+$,

$$\mathbf{P}[W_i(t + \varepsilon) = k \mid W_i(t) = k] = 1 - Q_{kk}^{(i)}(t)\varepsilon + o(\varepsilon), \tag{3}$$

$$\mathbf{P}[W_i(t + \varepsilon) = k' \mid W_i(t) = k] = Q_{kk'}^{(i)}(t)\varepsilon + o(\varepsilon), \tag{4}$$
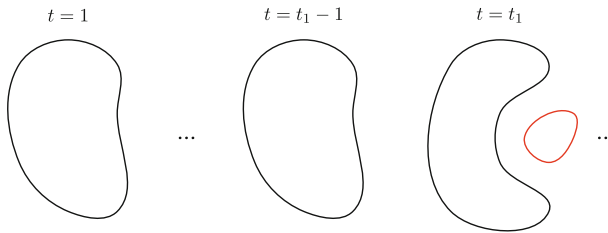
Fig. 1 Conceptual depiction of our change point model. The null hypothesis is that Q_0 governs all n vertex processes for all time. The alternative hypothesis (depicted) is that Q_0 governs all n vertex processes from time $t = 1$ through time $t = t_1 - 1$, and at time $t_1 - 1$ a small collection of n_1 vertex processes become governed by Q_1. This induces a change-point in the observed time series of graphs, characterized by altered probabilistic behavior for the n_1 vertices, at time t_1

(ii)   for each $t \leq t_1 - 1$,

$$\mathbf{P}[W_1(t) = k] = \cdots = \mathbf{P}[W_n(t) = k] = \pi_{0,k}, \quad \text{for each } k \in \mathscr{K}_+, \tag{5}$$
$$Q^{(1)}(t) = \cdots = Q^{(n)}(t) = Q_0, \tag{6}$$

(iii)   for each $t > t_1 - 1$,

$$\mathbf{P}[W_1(t) = k] = \cdots = \mathbf{P}[W_{n_1}(t) = k] = \pi_{1,k}, \quad \text{for each } k \in \mathscr{K}_+, \tag{7}$$
$$\mathbf{P}[W_{n_1+1}(t) = k] = \cdots = \mathbf{P}[W_n(t) = k] = \pi_{0,k}, \quad \text{for each } k \in \mathscr{K}_+, \tag{8}$$
$$Q^{(1)}(t) = \cdots = Q^{(n_1)}(t) = Q_1, \tag{9}$$
$$Q^{(n_1+1)}(t) = \cdots = Q^{(n)}(t) = Q_0. \tag{10}$$

(iv) $W_1, \ldots, W_n$ are mutually independent.

Remark 1 Definition 3 specifies a narrow class of random dot process models. An alternative description of a random dot process model $(\mathscr{V}, W)$ with change parameter $(t_1, n_1, \pi_0, \pi_1, Q_0, Q_1)$ is that $(\mathscr{V}, W)$ is a *piecewise exponential, block-homogeneous, independent and piecewise stationary random graph time series model with dot product kernel* (See Fig. 1). First, W is *piecewise exponential* because Part (i) of Definition 3 says that any vertex process can stay contiguously in any particular state only for an exponential amount of time unless interrupted by the change-point. Next, $W = (W_1, \ldots, W_n)$ is *block-homogeneous* because Part (ii) and Part (iii) say that if $t < t_1 - 1$, then the block of random variables $W_1(t), \ldots, W_n(t)$ are identical in distribution, and also that if $t \geq t_1 - 1$, then while $W_{n_1}(t)$ and $W_{n_1+1}(t)$ are different in distribution, the block of random variables $W_1(t), \ldots, W_{n_1}(t)$ are identical in distribution and the block of random variables $W_{n_1+1}(t), \ldots, W_n$ are identical in distribution. Moreover, W is *piecewise stationary* because for each $t \in \mathbb{N}$, Part (ii) and Part (iii) say that W is a stationary process over the interval $[t - 1, t)$. The term *independent* is due to Part (iv), and the term *dot product kernel* is because of the way edge formation probability is specified in our definition of random dot product spaces (cf. Sect. 16.4 in Bollobas et al. 2007). While it is convenient to direct our attention only to this special class of random dot process models, it would be of interest to investigate how the results in this article extend to a larger class of random dot process models.

2.4 Approximations

Fix a random dot process model $(\mathcal{V}, W)$ with change parameter

$$(t_1, n_1, \pi_0, \pi_1, Q_0, Q_1).$$

In this section, we introduce two approximations of the random dot process model. To begin, we introduce some new notation. First, let

$$\xi_0 = (\mathbf{1}\pi_0^T - Q_0)^{-1}(diag(\mathbf{1}) - \mathbf{1}\pi_0^T), \tag{11}$$
$$\xi_1 = (\mathbf{1}\pi_1^T - Q_1)^{-1}(diag(\mathbf{1}) - \mathbf{1}\pi_1^T). \tag{12}$$

For existence of the inverses in (11) and (12), we refer the reader to Corollary II.4.9 of Asmussen (2003). Now, let

$$\Sigma_0 = diag(\pi_0)\xi_0 + \xi_0^T diag(\pi_0), \tag{13}$$
$$\Sigma_1 = diag(\pi_1)\xi_1 + \xi_1^T diag(\pi_1). \tag{14}$$

Next, let $\overline{\pi}_0$ (resp. $\overline{\pi}_1$) be the sub-probability vector obtained by removing the last coordinate of π_0 (resp. π_1), and let $\widehat{\Sigma}_0$ (resp. $\widehat{\Sigma}_1$) be the square matrix obtained by removing both the last row and the last column of Σ_0 (resp. Σ_1). To distinguish $\mathcal{V}$ from its approximations, we will denote by $\overline{\mathcal{V}}$ our first-order approximation of $\mathcal{V}$, and denote by $\widehat{\mathcal{V}}$ our second-order approximation of $\mathcal{V}$.

Definition 4 The first-order approximation $\overline{\mathcal{V}}$ of $\mathcal{V}$ is the sequence $\{(\overline{X}(t), \overline{G}(t))\}_{t=1}^{\infty}$ of independent random dot product spaces such that

(i)   for each $t \leq t_1 - 1$,

$$\overline{X}_1(t) = \cdots = \overline{X}_n(t) = \overline{\pi}_0, \tag{15}$$

(ii) for each $t \geq t_1$,

$$\overline{X}_1(t) = \cdots = \overline{X}_{n_1}(t) = \overline{\pi}_1, \tag{16}$$
$$\overline{X}_{n_1+1}(t) = \cdots = \overline{X}_n(t) = \overline{\pi}_0. \tag{17}$$

The first approximation, then, yields a sequence of independent random graphs with independent edges. For $t \leq t_1 - 1$, $G(t)$ is an attributed version of the so-called Erdös-Renyi random graph (see, e.g., Bollobas 2001) where edges (ignoring the attribute value) are present independently with probability $\langle \overline{\pi}_0, \overline{\pi}_0 \rangle$. For $t \geq t_1$, $G(t)$ is an attributed blockmodel (see, e.g., Doreian et al. 2005) as illustrated in Fig. 2; again, edges are present independently, with edge probabilities between two of the $n - n_1$ "null process" vertices specified deterministically by $\langle \overline{\pi}_0, \overline{\pi}_0 \rangle$, edge probabilities between two of the n_1 "alternative process" vertices specified deterministically by $\langle \overline{\pi}_1, \overline{\pi}_1 \rangle$, and edges between a "null process" vertex and an "alternative process" vertex specified deterministically by $\langle \overline{\pi}_0, \overline{\pi}_1 \rangle$. (A relevant recent development for such models is Airoldi et al. (2008), where the so-called *mixed membership stochastic blockmodel* is proposed and studied for parameter estimation.)

Definition 5 A second-order approximation $\widehat{\mathcal{V}}$ of $\mathcal{V}$ is a sequence $\{(\widehat{X}(t), \widehat{G}(t))\}_{t=1}^{\infty}$ of independent random dot product spaces such that

(i)   for each time t and vertex i, $\widehat{X}_i(t)$ is the random vector obtained by truncating a multivariate normal random vector $Z_i(t)$ to $\mathscr{S}$,

Fig. 2 An illustration of the probabilistic blockmodel structure for $G(t)$, $t \geq t_1$. Edge probabilities (and their attributes) depend on the vertex processes. This illustration depicts $n - n_1$ "null process" vertices and a small collection of n_1 "alternative process" vertices
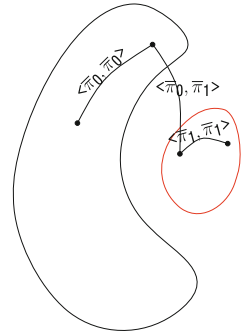
(ii) for each $t \leq t_1 - 1$,

$$\mathbf{E}[Z_1(t)] = \cdots = \mathbf{E}[Z_n(t)] = \overline{\pi}_0, \tag{18}$$

$$Var(Z_1(t)) = \cdots = Var(Z_n(t)) = \widehat{\Sigma}_0, \tag{19}$$

(iii) for each $t \geq t_1$,

$$\mathbf{E}[Z_1(t)] = \cdots = \mathbf{E}[Z_{n_1}(t)] = \overline{\pi}_1, \tag{20}$$

$$\mathbf{E}[Z_{n_1+1}(t)] = \cdots = \mathbf{E}[Z_n(t)] = \overline{\pi}_0, \tag{21}$$

$$Var(Z_1(t)) = \cdots = Var(Z_{n_1}(t)) = \widehat{\Sigma}_1, \tag{22}$$

$$Var(Z_{n_1+1}(t)) = \cdots = Var(Z_n(t)) = \widehat{\Sigma}_0. \tag{23}$$

A second-order approximation yields a sequence of independent latent position graphs (Hoff et al. 2002; Handcock et al. 2007; Bollobas et al. 2007; Scheinerman and Tucker 2010; Young and Scheinerman 2009; Marchette and Priebe 2008a), illustrated in Fig. 3. The left panel of Fig. 3 depicts a realization of $n = 100$ random vectors X_i in the simplex at time t_1; $n_1 = 9$ of the vectors are approximately normally distributed with mean $\overline{\pi}_1$ and covariance $\widehat{\Sigma}_1$, while $n - n_1$ are approximately normally distributed with mean $\overline{\pi}_0$ and covariance $\widehat{\Sigma}_0$. Edge probabilities between vertices are governed (but *not* specified deterministically) by the vertex processes.

3 Convergence results

3.1 Sequence of random dot process models

In this section, consider a sequence $\{(\mathscr{V}^r, W^r) : r > 0\}$ of random dot process models with vertices $[n]$ and attributes $\mathscr{K}$ where, for each $r > 0$, $(\mathscr{V}^r, W^r)$ has the change parameter $(t_1, n_1, \pi_0, \pi_1, rQ_0, rQ_1)$. The parameter r is identified as the *vertex process rate parameter*. Throughout this section we make the assumption that $t_1 \in \mathbb{N}$, $n_1 \in \{1, \ldots, n\}$, and $(K + 1)$-dimensional probability vectors π_0 and π_1 and irreducible aperiodic $(K + 1) \times (K + 1)$-dimensional matrices Q_0 and Q_1 are such that $\pi_0^T Q_0 = 0$ and $\pi_1^T Q_1 = 0$.

For each $r > 0$, we define our second-order approximation $\widetilde{\mathscr{V}}^r$ of $\mathscr{V}^r$ to be the second-order approximation of $\mathscr{V}^r$ with $\Sigma_0^r = \Sigma_0/r$ and $\Sigma_1^r = \Sigma_1/r$. Also, note that for each $r > 0$,
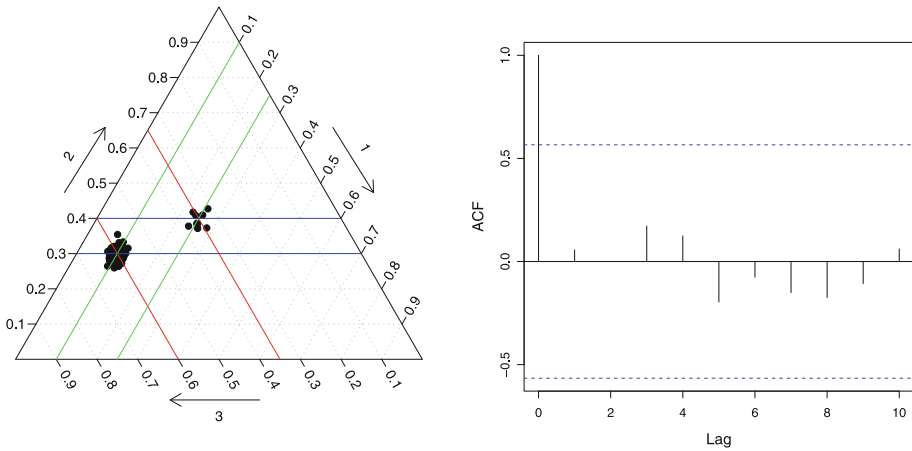
Fig. 3 *Left*: example of the vectors X_i in the simplex at time t_1 for the random dot process model, with $n = 100, n_1 = 9, K = 2, \pi_0 = (0.10, 0.30, 0.60)^T$ and $\pi_1 = (0.25, 0.40, 0.35)^T$. (The matrices Q_0 and Q_1 are given in Sect. 4.2.2.) For this case (with $r = 1024$—see Sect. 3.1) the second approximation (truncated multivariate normals) provides a good model. *Right*: autocorrelation as a function of lag for the edge attributes, showing that (for this case) the graphs are approximately uncorrelated in time

we have the first-order approximation $\overline{\mathcal{V}}^r$ of $\mathcal{V}^r$, and in fact, for any $r_1, r_2 > 0$, we have $\overline{\mathcal{V}}^{r_1} = \overline{\mathcal{V}}^{r_2}$. Hence, we will simply write $\overline{\mathcal{V}}$ for any $\overline{\mathcal{V}}^r$ whenever it is convenient.

3.2 Asymptotic independence

We first derive the asymptotic distribution of $D_i^r(t)$, where for each vertex i, time t and $r > 0$,

$$D_i^r(t) = \sqrt{r}(X_i^r(t) - \mathbf{E}[\overline{X}_i(t)]). \tag{24}$$

Lemma 1 *For each time t and vertex i, as $r \to \infty$, the sequence $\{D_i^r(t) : r > 0\}$ converges in distribution to a multivariate normal with mean zero and covariance matrix $\widehat{\Sigma}$, where*

$$\widehat{\Sigma} = \begin{cases} \widehat{\Sigma}_1, & \text{if } i \leq n_1 \text{ and } t \geq t_1, \\ \widehat{\Sigma}_0, & \text{otherwise.} \end{cases} \tag{25}$$

Proof To simplify our exposition, we will give a detailed proof only for the case where $n_1 = 0$ and $t_1 = \infty$, leaving the other cases to the reader; the techniques used here can be easily adapted.

First, we make some preliminary observations. Let Z be a continuous-time Markov process with state space $\mathcal{K}_+$ and its infinitesimal generator Q_0, and assume that Z has the initial distribution π_0. In particular, Z is ergodic, so by Functional Central Limit Theorem (FCLT) we have that as $r \to \infty$, the random vector $\widehat{Z}^r = (\widehat{Z}_1^r, \ldots, \widehat{Z}_K^r)$, where each $\widehat{Z}_k^r \equiv \sqrt{r}\left(\frac{1}{r}\int_0^r 1_{\{k\}}(Z(s))ds - \pi_{0,k}\right)$, converges in distribution to a multivariate normal vector with mean zero and covariance matrix given by the following formula (cf. Example 4.4 in Glynn and Whitt 1992):

$$\int_0^\infty \mathbf{E}[(Z(0) - \pi_0)(Z(s) - \pi_0)^T]ds + \int_0^\infty \mathbf{E}[(Z(s) - \pi_0)(Z(0) - \pi_0)^T]ds. \tag{26}$$

Also, recall that the transition matrix of Z can be represented by a matrix-exponential as follows; for each k and $k' \in \mathscr{K}_+$, and $t \in [0, \infty)$,

$$\mathbf{P}\left[Z(t) = k' | Z(0) = k\right] = [\exp(Q_0 t)]_{kk'}, \tag{27}$$

where $\exp(Q_0 t) = \sum_{n=0}^{\infty} \frac{(Q_0 t)^n}{n!}$. Now, we will also show that

$$\Sigma_0 = \int_0^{\infty} \mathbf{E}[(Z(0) - \pi_0)^T (Z(s) - \pi_0)]ds + \int_0^{\infty} \mathbf{E}[(Z(s) - \pi_0)^T (Z(0) - \pi_0)]ds.$$

To do this, observe that for each $t \in [0, \infty)$, and $k, k' \in \mathscr{K}_+$,

$$\mathbf{E}[Z_k(0)Z_{k'}(t)] = \mathbf{E}\left[Z_k(0)\mathbf{E}\left[Z_{k'}(t) | Z(0)\right]\right] \tag{28}$$
$$= \mathbf{E}\left[Z_k(0)[\exp(Q_0 t)]_{kk'}\right] \tag{29}$$
$$= \pi_{0,k}[\exp(Q_0 t)]_{kk'}, \tag{30}$$

where for convenience we denote $1_k(Z(\cdot))$ by $Z_k(\cdot)$. Moreover, appealing to (II.4.5) in Asmussen (2003), we have that for each $t \in [0, \infty)$, and k and $k' \in \mathscr{K}_+$,

$$\xi_{0,kk'} = \int_0^{\infty} [\exp(Q_0 s) - \mathbf{1}\pi_0^T]_{kk'}ds = \int_0^{\infty} [\exp(Q_0 s)]_{kk'} - \pi_{0,k'}ds. \tag{31}$$

Hence, for each k and $k' \in \mathscr{K}_+$, it follows that

$$\int_0^{\infty} \mathbf{E}[(Z_k(0) - \pi_{0,k})(Z_{k'}(s) - \pi_{0,k'})]ds \tag{32}$$

$$= \int_0^{\infty} \mathbf{E}[Z_k(0)Z_{k'}(s)] - \pi_{0,k}\pi_{0,k'}ds \tag{33}$$

$$= \int_0^{\infty} \pi_{0,k}[\exp(Q_0 s)]_{kk'} - \pi_{0,k}\pi_{0,k'}ds \tag{34}$$

$$= \pi_{0,k}\int_0^{\infty} [\exp(Q_0 s)]_{kk'} - \pi_{0,k'}ds \tag{35}$$

$$= \pi_{0,k}\xi_{0,kk'}, \tag{36}$$

and therefore, we have that the k-th row and k'-th column of the covariance matrix is

$$\pi_{0,k}\xi_{0,kk'} + \pi_{0,k'}\xi_{0,k'k} = [diag(\pi_0)\xi_0 + \xi_0^T diag(\pi_0)]_{kk'} = \Sigma_{0,kk'}. \tag{37}$$

Now, for each $k \in \mathcal{K}_+$, we see that

$$D_{i,k}^r(t) = \sqrt{r} \left(X_{ik}^r(t) - \pi_{0,k} \right) \tag{38}$$

$$= \sqrt{r} \left(\int_{t-1}^{t} 1_{\{k\}}(W_i^r(s))ds - \pi_{0,k} \right) \tag{39}$$

$$\stackrel{d}{=} \sqrt{r} \left(\int_{t-1}^{t} 1_{\{k\}}(Z(rs))ds - \pi_{0,k} \right) \tag{40}$$

$$= \sqrt{r} \left(\frac{1}{r} \int_{0}^{r} 1_{\{k\}}(Z(t-1+s))ds - \pi_{0,k} \right) \tag{41}$$

$$\stackrel{d}{=} \sqrt{r} \left(\frac{1}{r} \int_{0}^{r} 1_{\{k\}}(Z(s))ds - \pi_{0,k} \right). \tag{42}$$

The Lemma follows from this equivalence in distribution together with our preliminary observations.   □

Next, we establish an asymptotic independence property.

Theorem 1 *Fix t and $h \geq 1$. For each vertex i, and $k, k' \in \mathcal{K}_+$, as $r \to \infty$ the sequence $\{(D_{i,k}^r(t), D_{i,k'}^r(t+h)) : r > 0\}$ converges in distribution to a bivariate normal random vector whose covariance matrix is the identity matrix.*

Proof To simplify our exposition, we will give a detailed proof only for the case where $n_1 = 0$ and $t_1 = \infty$. First, fix vertex i and times t and $t+h \in \mathbb{N}$. Let $(D_{i,k}(t), D_{i,k}(t+h))$ be a weak limit point of $\{(D_{i,k}^r(t), D_{i,k}^r(t+h)) : r > 0\}$, and note that by Lemma 1 such a weak limit exists, and moreover $D_{i,k}(t)$ and $D_{i,k}(t+h)$ are normal random variables. In fact, appealing to the FCLT used in our proof of Lemma 1, it can be shown that $(D_{i,k}(t), D_{i,k}(t+h))$ is indeed a bivariate normal vector. Hence, to show independence, it is enough to show that two random variables are uncorrelated.

Let Z be as in the proof of Lemma 1. Note that for each $k, k' \in \mathcal{K}_+$ and $r > 0$,

$$\mathbf{E}[D_{ik}^r(t)D_{ik'}^r(t+h)] = r\mathbf{E}[(X_{i,k}^r(t) - \pi_{0,k})(X_{i,k'}^r(t+h) - \pi_{0,k'})] \tag{43}$$

$$= rCov(X_{i,k}^r(t), X_{i,k'}^r(t+h)), \tag{44}$$

and by (3.12) in Parzen (1999) together with Theorem B in Keilson (1998), for some $c_0 \geq 0$ and $\lambda_0 > 0$, we have

$$r \left| Cov(X_{i,k}^r(t), X_{k'}^r(t+h)) \right| \tag{45}$$

$$= r^{-1} \left| Cov \left(\int_{r(t-1)}^{rt} Z_k(s)ds, \int_{r(t+h-1)}^{r(t+h)} Z_{k'}(u)du \right) \right| \tag{46}$$

$$= r^{-1} \left| \int_{r(t-1)}^{rt} \int_{r(t+h-1)}^{r(t+h)} Cov\left(Z_k(u), Z_{k'}(s)\right) dsdu \right| \tag{47}$$

$$\leq r^{-1} \int\limits_{r(t-1)}^{rt} \int\limits_{r(t+h-1)}^{r(t+h)} \left| \pi_{0,k} [\exp(Q_0(s-u))]_{kk'} - \pi_{0,k}\pi_{0,k'} \right| ds\,du \qquad (48)$$

$$= r^{-1} \pi_{0,k} \int\limits_{r(t-1)}^{rt} \int\limits_{r(t+h-1)}^{r(t+h)} \left| [\exp(Q_0(s-u))]_{kk'} - \pi_{0,k'} \right| ds\,du \qquad (49)$$

$$\leq r^{-1} \pi_{0,k} c_0 \int\limits_{r(t-1)}^{rt} \int\limits_{r(t+h-1)}^{r(t+h)} \exp(-\lambda_0(s-u))ds\,du \qquad (50)$$

$$= r^{-1} \pi_{0,k} c_0 \int\limits_{r(t-1)}^{rt} \exp(\lambda_0 u)du \int\limits_{r(t+h-1)}^{r(t+h)} \exp(-\lambda_0 s)ds \qquad (51)$$

$$= r^{-1} \pi_{0,k} c_0 \frac{-1}{\lambda_0^2} (e^{\lambda_0 rt} - e^{\lambda_0(rt-r)})(e^{-\lambda_0(rt+rh)} - e^{-\lambda_0(rt+rh-r)}) \qquad (52)$$

$$= r^{-1} \pi_{0,k} c_0 \frac{-1}{\lambda_0^2} (1 - e^{-\lambda_0 r})(e^{-\lambda_0 rh} - e^{-\lambda_0(rh-r)}), \qquad (53)$$

which converges to zero as $r \to \infty$. Hence, by Fatou's lemma (through a Skorohod representation), we have

$$|\mathbf{E}[D_{ik}(t)D_{ik'}(t+h)]| \leq \lim_{r\to\infty} \left| \mathbf{E}[D_{ik}^r(t)D_{ik'}^r(t+h)] \right| = 0. \qquad (54)$$

Since $D_i(t)$ and $D_i(t+h)$ are mean-zero random variables, it follows that

$$Cov(D_i(t), D_i(t+h)) = 0. \qquad (55)$$

$\square$

3.3 Total variation distance

In this section, we will show that when r is sufficiently large there is very little statistical difference amongst $\mathcal{V}$, $\widehat{\mathcal{V}}^r$ and $\mathcal{V}^r$, and this will be the basis for our approximate inference procedures.

Theorem 2 *For each $t \in [0, \infty)$, when r is sufficiently large,*

$$d_{TV}(G^r(t), \overline{G}(t)) = O(1/r), \qquad (56)$$
$$d_{TV}(G^r(t), \widehat{G}^r(t)) = o(1/r), \qquad (57)$$

where $d_{TV}(\cdot, \cdot)$ is the total-variation distance between two random elements.

Proof Fix $r > 0$, $t \in [0, \infty)$ and graph g. First, we observe that there exists a finite collection $\{f_{i\ell}\}$ of monomials with K-variables such that the following holds:

$$\mathbf{P}[G^r(t) = g] = \mathbf{E}[\mathbf{P}[G^r(t) = g \,|\, X^r(t)]] \qquad (58)$$

$$= \mathbf{E}\left[\sum_\ell \prod_{i=1}^n f_{i\ell}(X_i^r(t)) \right] \qquad (59)$$

$$= \sum_\ell \prod_{i=1}^n \mathbf{E}[f_{i\ell}(X_i^r(t))]. \qquad (60)$$

For each i, let

$$\dot{X}_i^r(t) = X_i^r(t) - \mathbf{E}[X_i^r(t)]. \tag{61}$$

Now, consider any monomial $f_{i\ell}$. Let $\nabla f_{i\ell}$ be the gradient of $f_{i\ell}$, and let $\Delta f_{i\ell}$ be the Hessian of $f_{i\ell}$, where the gradient and the Hessian are computed at $\mathbf{E}[X_i^r(t)]$. Recall that any symmetric matrix, by the spectral theorem, may be regarded as a real diagonal matrix that has been re-expressed in some new coordinate system. In other words, when $\Lambda = diag(\lambda_1, \ldots, \lambda_K)$ is the matrix whose diagonal entries are eigenvalues of $\Delta f_{i\ell}$, we have $\Delta f_{i\ell} = M^T \Lambda M$ for some unitary matrix M whose rows are orthonormal eigenvectors of Λ, forming a basis. Note that M may depend on the monomial. Now, for each x,

$$x^T \Delta f_{i\ell} x = x^T (M^T \Lambda M) x = \sum_{k=1}^{K} \lambda_k ([Mx]_k)^2. \tag{62}$$

We claim that

$$\mathbf{E}[f_{i\ell}(X_i^r(t))] = f_{i\ell}(\mathbf{E}[X_i^r(t)]) + \frac{1}{2r}\left(\sum_{k=1}^{K}\sum_{a=1}^{K}\sum_{b=1}^{K}\lambda_k M_{ka} M_{kb} \Sigma_{ab}\right) + o(1/r), \tag{63}$$

$$\mathbf{E}[f_{i\ell}(\widehat{X}_i^r(t))] = f_{i\ell}(\mathbf{E}[\widehat{X}_i^r(t)]) + \frac{1}{2r}\left(\sum_{k=1}^{K}\sum_{a=1}^{K}\sum_{b=1}^{K}\lambda_k M_{ka} M_{kb} \Sigma_{ab}\right) + o(1/r). \tag{64}$$

To prove (63), using the Taylor expansion of $f_{i\ell}$ at $\mathbf{E}[X_i^r(t)]$ together with the fact that

$$\mathbf{E}[\nabla f_{i\ell} \cdot (\dot{X}_i^r(t))] = \nabla f_{i\ell} \cdot \mathbf{E}[\dot{X}_i^r(t)] = 0,$$

we observe that

$$\mathbf{E}[f_{i\ell}(X_i^r(t))] = f_{i\ell}(\mathbf{E}[X_i^r(t)]) + \mathbf{E}[\nabla f_{i\ell} \cdot (\dot{X}_i^r(t))] + \mathbf{E}\left[\frac{(\dot{X}_i^r(t))^T \Delta f_{i\ell} \dot{X}_i^r(t)}{2}\right] + o(1/r)$$

$$= f_{i\ell}(\mathbf{E}[X_i^r(t)]) + \mathbf{E}\left[\frac{(\dot{X}_i^r(t))^T \Delta f_{i\ell} \dot{X}_i^r(t)}{2}\right] + o(1/r).$$

Now, note that

$$\frac{1}{2}\mathbf{E}[(\dot{X}_i^r(t))^T \Delta f_{i\ell} \dot{X}_i^r(t)] + o(1/r)$$

$$= \frac{1}{2}\sum_{k=1}^{K}\lambda_k \mathbf{E}\left[\left(\sum_{a=1}^{K} M_{ka}\dot{X}_{ia}^r(t)\right)^2\right] + o(1/r)$$

$$= \frac{1}{2}\sum_{k=1}^{K}\lambda_k \mathbf{E}\left[\sum_{a=1}^{K}\sum_{b=1}^{K} M_{ka}M_{kb}\dot{X}_{ia}^r(t)\dot{X}_{ib}^r(t)\right] + o(1/r)$$

$$= \frac{1}{2}\sum_{k=1}^{K}\sum_{a=1}^{K}\sum_{b=1}^{K}\lambda_k M_{ka}M_{kb}\mathbf{E}[\dot{X}_{ia}^r(t)\dot{X}_{ib}^r(t)] + o(1/r)$$

$$= \frac{1}{2r}\sum_{k=1}^{K}\sum_{a=1}^{K}\sum_{b=1}^{K}\lambda_k M_{ka}M_{kb}Cov(D_{ia}^r(t), D_{ib}^r(t)) + o(1/r)$$

$$= \frac{1}{2r} \sum_{k=1}^{K} \sum_{a=1}^{K} \sum_{b=1}^{K} \lambda_k M_{ka} M_{kb} (\Sigma_{ab} + o(1)) + o(1/r)$$

$$= \frac{1}{2r} \left(\sum_{k=1}^{K} \sum_{a=1}^{K} \sum_{b=1}^{K} \lambda_k M_{ka} M_{kb} \Sigma_{ab} \right) + o(1/r).$$

This completes our proof of the claim in (63); the claim in (64) can also be proven similarly. Since $\mathbf{E}[\widehat{X}_i^r(t)] = \mathbf{E}[X_i^r(t)]$, it follows from (64) and (63) that

$$\mathbf{E}[f_{i\ell}(X_i^r(t))] = \mathbf{E}[f_{i\ell}(\widehat{X}_i^r(t))] + o(1/r), \tag{65}$$

and thus

$$\mathbf{P}[G^r(t) = g] = \sum_{\ell} \prod_{i=1}^{n} \mathbf{E}[f_{i\ell}(X_i^r(t))] \tag{66}$$

$$= \sum_{\ell} \prod_{i=1}^{n} \left(\mathbf{E}[f_{i\ell}(\mathbf{E}[\widehat{X}_i^r(t)])] + o(1/r) \right) \tag{67}$$

$$= \left(\sum_{\ell} \prod_{i=1}^{n} \mathbf{E}[f_{i\ell}(\mathbf{E}[\widehat{X}_i^r(t)])] \right) + o(1/r) \tag{68}$$

$$= \mathbf{P}[\widehat{G}^r(t) = g] + o(1/r). \tag{69}$$

It follows that for each $t \in [0, \infty)$ we have

$$d_{TV}(G^r(t), \widehat{G}^r(t)) = \sum_{g} \left| \mathbf{P}[G^r(t) = g] - \mathbf{P}[\widehat{G}^r(t) = g] \right| = o(1/r). \tag{70}$$

Also, using (63), one can prove, in a similar manner, that

$$d_{TV}(G^r(t), \overline{G}(t)) = \sum_{g} \left| \mathbf{P}[G^r(t) = g] - \mathbf{P}[\overline{G}(t) = g] \right| = O(1/r). \tag{71}$$

□

4 Application to a change-point detection problem

Our change-point detection problem is investigated via a comparative power analysis for the following test of hypotheses, given time $t \in \mathbb{N}$ and level of significance $\alpha \in (0, 1)$. For a random dot process model $(\mathcal{V}, W)$ with change parameter $(t_1, n_1, \pi_0, \pi_1, Q_0, Q_1)$,

$$\mathscr{H}_0 : t_1 > t,$$
$$\mathscr{H}_1 : t_1 = t.$$

In this section, we demonstrate how the first-order and second-order approximations of $\mathcal{V}$ can help us compare the power of certain test statistics for our change-point detection problem.

4.1 Fusion statistics

Definition 6 For each $x \in \mathbb{R}^K$, the linear size fusion of $G(t)$ with fusion parameter x is the random variable $J^x(t)$ defined by

$$J^x(t) = \sum_{k=1}^{K} x_k |G(t)|_k, \tag{72}$$

where $|G(t)|_k$ denotes the number of edges in $G(t)$ with attribute k.

The class of statistics J^x allows us to investigate power for different attribute weighting schemes. For instance, if $x = (1, 1, \ldots, 1) \in \mathbb{R}^K$, then $J^x(t)$ is the number of edges in $G(t)$, while if x is the unit vector e_k with its k-th coordinate being 1, then $J^x(t)$ is the number of edges in $G(t)$ with attribute k.

For our next definition, for each $x \in \mathbb{R}^K$, $m \geq 1$, and $t \geq m + 1$, let

$$\overline{J}^x_m(t) = \frac{1}{m} \sum_{s=1}^{m} J^x(t - s), \tag{73}$$

providing for normalization based on recent past (assumed to be evolving according to the null dynamics), and note that for x and $y \in \mathbb{R}^K$ and $c \in \mathbb{R}$,

$$J^{cx}(t) = c J^x(t), \tag{74}$$

$$J^{x+y}(t) = J^x(t) + J^y(t). \tag{75}$$

Definition 7 For each $x \in \mathbb{R}^K$, $m \geq 1$, and $t \geq m + 1$, the normalized fusion statistic with parameter x is the random variable $T^x_m(t)$ defined by

$$T^x_m(t) = \begin{cases} J^x(t) - \overline{J}^x_m(t), & \text{if } m = 1, \\ \dfrac{J^x(t) - \overline{J}^x_m(t)}{\sqrt{\frac{1}{m-1} \sum_{s=1}^{m} \left(J^x(t-s) - \overline{J}^x_m(t) \right)^2}}, & \text{if } m \geq 2. \end{cases} \tag{76}$$

where we take $0/0 = 0$ and $x/0 = \infty$ for each $x \neq 0$.

For each $x \in \mathbb{R}^K$, $m \geq 1$ and $t \in \{m + 1, \ldots, t_1\}$, we define a function $u^x_{m,t} : \mathbb{R} \to (0, 1)$ by letting, for each $c \in \mathbb{R}$,

$$u^x_{m,t}(c) = \mathbf{P}[T^x_m(t) > c]. \tag{77}$$

(We consider a one-sided test, for *increased* activity, for concreteness; the following development can be replicated, *mutatis mutandis*, for a two-sided test.)

Note that the function $u^x_{m,t}$ is only right-continuous with finite left-limits and is non-increasing with strictly decreasing only at finitely many points. Nevertheless, we will assume that each $u^x_{m,t}$ is indeed continuous and strictly decreasing; our rationale for this is provided in Remark 2.

Definition 8 Let $x \in \mathbb{R}^K$ and $m \geq 1$. The power function of T^x_m is the function $\beta^x_m : (0, 1) \to [0, 1]$ defined by letting, for each $\alpha \in (0, 1)$,

$$\beta^x = \beta^x_m(\alpha) = u^x_{m,t_1}(u^{x,-1}_{m,t_1-1}(\alpha)), \tag{78}$$

where $u^{x,-1}_{m,t_1-1}$ denotes the inverse function of u^x_{m,t_1-1}.

Remark 2 Smoothing and linear interpolation can be done for each pair of two consecutive discontinuity points of the function $u_{m,t}^{x,-1}$. So, we have identified the function $u_{m,t}^{x,-1}$ with its smoothed version for the purpose of this section. More specifically, for each $m \in \mathbb{N}$, $t \geq m+1$ and $x \in \mathbb{R}^K$, we identify $u_{m,t}^{x,-1}$ with a smoothed version of the function $v_{m,t}^{x,-1} : (0, 1) \to \mathbb{R}$ given by letting, for each pair of consecutive discontinuity points α_1 and α_2 with $\alpha_1 < \alpha_2$,

$$v_{m,t}^{x,-1}(\alpha) = \frac{\alpha - \alpha_1}{\alpha_2 - \alpha_1} u_{m,t}^{x,-1}(\alpha_2) + \frac{\alpha_2 - \alpha}{\alpha_2 - \alpha_1} u_{m,t}^{x,-1}(\alpha_1). \tag{79}$$

Remark 3 Let $m \in \mathbb{N}$ and $\alpha \in (0, 1)$. Then, there exists $x \in \mathbb{R}^K$ with $\|x\| = 1$ such that for each $y \in \mathbb{R}^K$ such that $\|y\| = 1$,

$$\beta^x \geq \beta^y. \tag{80}$$

We will briefly outline a proof. For this, fix $m \in \mathbb{N}$ and $\alpha \in (0, 1)$, and let $D = \{x \in \mathbb{R}^K : \|x\| = 1\}$. Then, define a function $H_{m,\alpha} : D \to [0, 1]$ by letting, for each $x \in D$,

$$H_{m,\alpha}(x) = \beta^x. \tag{81}$$

β^x can be constructed so that $H_{m,\alpha}$ is a continuous function. Then, since D is compact, it follows that there exists $y_0 \in D$ such that for each $y \in D$,

$$\beta^{y_0} = H_{m,\alpha}(y_0) \geq H_{m,\alpha}(y) = \beta^y. \tag{82}$$

Then it follows (see Lemma 2 below) that for each $y \in \mathbb{R}^K$ with $y \neq 0$,

$$\beta^{y_0} \geq \beta^{y/|y|} = \beta^y. \tag{83}$$

On the other hand, if $y = 0$, then $\beta^0 = \alpha$, where the equality is due to the fact that the test is conducted simply by "tossing a biased coin".

Our observation made in Remark 3 can be in fact extended to the following; there exists $x \in \mathbb{R}^K$ with $\|x\| = 1$ such that for each $y \in \mathbb{R}^K$ the inequality (80) is satisfied. For this, it is sufficient to note the following "scaling" property which is a consequence of (74).

Lemma 2 *Let $c \in \mathbb{R} \setminus \{0\}$ and $m \geq 2$. For each $x \in \mathbb{R}^K$ and $t \geq m + 1$,*

$$T_m^{cx}(t) = \frac{c}{|c|} T_m^x(t). \tag{84}$$

Proof By the scaling property in (74),

$$T_m^{cx}(t) = \frac{J^{cx}(t) - \overline{J}_m^{cx}(t)}{\sqrt{\frac{1}{m-1} \sum_{k=1}^m \left(J^{cx}(t-k) - \overline{J}_m^{cx}(t) \right)^2}}$$

$$= \frac{cJ^x(t) - c\overline{J}_m^x(t)}{\sqrt{\frac{1}{m-1} \sum_{k=1}^m \left(cJ^x(t-k) - c\overline{J}_m^x(t) \right)^2}}$$

$$= \frac{c(J^x(t) - \overline{J}_m^x(t))}{\sqrt{\frac{c^2}{m-1} \sum_{k=1}^m \left(J^x(t-k) - \overline{J}_m^x(t) \right)^2}}$$

$$= \frac{c}{|c|} T_m^x(t).$$

$\square$

4.2 Power estimates

4.2.1 Asymptotic

In this section, by taking the number of vertices n large and assuming that the first-order approximation is appropriate, we will derive an approximation of the power function of our fusion statistics. To be specific, we fix $t_1, n_1, \pi_0, \pi_1, x, m$ and the level of significance α, and consider a sequence $\{\overline{\mathcal{V}}_n : n \geq n_1 + 1\}$ of the first-order approximation. For large values of n, we will derive an approximation of β^x.

First, let π_{00}, π_{01} and π_{11} be the K-dimensional sub-probability vectors defined by letting for each $k \in \mathcal{K}$,

$$\pi_{00,k} = \pi_{0,k}\pi_{0,k}, \tag{85}$$

$$\pi_{01,k} = \pi_{0,k}\pi_{1,k}, \tag{86}$$

$$\pi_{11,k} = \pi_{1,k}\pi_{1,k}, \tag{87}$$

and then let η_{00}, η_{01} and η_{11} be the $K \times K$-dimensional matrices defined by letting for k and $k' \in \mathcal{K}$,

$$[\eta_{00}]_{k,k'} = \begin{cases} \pi_{00,k}(1 - \pi_{00,k}), & \text{if } k = k', \\ -\pi_{00,k}\pi_{00,k'}, & \text{otherwise}, \end{cases} \tag{88}$$

$$[\eta_{01}]_{k,k'} = \begin{cases} \pi_{01,k}(1 - \pi_{01,k}), & \text{if } k = k', \\ -\pi_{01,k}\pi_{01,k'}, & \text{otherwise}, \end{cases} \tag{89}$$

$$[\eta_{11}]_{k,k'} = \begin{cases} \pi_{11,k}(1 - \pi_{11,k}), & \text{if } k = k', \\ -\pi_{11,k}\pi_{11,k'}, & \text{otherwise}. \end{cases} \tag{90}$$

We assume that the following conditions hold for the remainder of our analysis.
Assumption 1 The symmetric matrices η_{00}, η_{01} and η_{11} are positive definite, and

$$\binom{n_1}{2}(\pi_{11} - \pi_{00}) + (n - n_1)n_1(\pi_{01} - \pi_{00}) \neq 0. \tag{91}$$

Lemma 3 *Let $t \in \{m + 1, \ldots, t_1\}$ and $x \in \mathbb{R}^K$. For sufficiently large n, the random variable $\psi_m^x(t)$ has approximately Student t-distribution with degrees of freedom $m - 1$ and non-centrality parameter μ^x, where*

$$\mu^x = \begin{cases} \dfrac{\binom{n_1}{2}\langle x, \pi_{11} - \pi_{00} \rangle + (n - n_1)n_1\langle x, \pi_{01} - \pi_{00} \rangle}{\sqrt{(1 + 1/m)\binom{n}{2}\langle x, \eta_{00}x \rangle + \binom{n_1}{2}\langle x, \eta_{11}x - \eta_{00}x \rangle + (n - n_1)n_1\langle x, \eta_{01}x - \eta_{00}x \rangle}}, & \text{if } t = t_1, \\ 0, & \text{if } t < t_1, \end{cases} \tag{92}$$

$$\psi_m^x(t) = \begin{cases} \dfrac{1}{\sqrt{(1 + 1/m) + \dfrac{\binom{n_1}{2}}{\binom{n}{2}}\left(\dfrac{\langle x, \eta_{11}x \rangle}{\langle x, \eta_{00}x \rangle} - 1\right) + \dfrac{(n - n_1)n_1}{\binom{n}{2}}\left(\dfrac{\langle x, \eta_{01}x \rangle}{\langle x, \eta_{00}x \rangle} - 1\right)}} T_m^x(t), & \text{if } t = t_1, \\ \dfrac{1}{\sqrt{1 + 1/m}} T_m^x(t), & \text{if } t < t_1. \end{cases} \tag{93}$$

Proof Fix $t \in \{m + 1, \ldots, t_1\}$ and $x \in \mathbb{R}^K$. For each $t - \ell \in \mathbb{N}$, we have that the random variable $J^x(t - \ell)$ has mean

$$\mathbf{E}[J^x(t - \ell)] = \left(\binom{n}{2} - \binom{n_1(t - \ell)}{2} - (n - n_1(t - \ell)) \cdot n_1(t - \ell)\right)\langle x, \pi_{00} \rangle$$

$$+ \binom{n_1(t - \ell)}{2}\langle x, \pi_{11} \rangle + (n - n_1(t - \ell)) \cdot n_1(t - \ell)\langle x, \pi_{01} \rangle,$$

and variance

$$Var[J^x(t-\ell)] = \left(\binom{n}{2} - \binom{n_1(t-\ell)}{2} - (n - n_1(t-\ell)) \cdot n_1(t-\ell)\right)\langle x, \eta_{00}x\rangle$$
$$+ \binom{n_1(t-\ell)}{2}\langle x, \eta_{11}x\rangle + (n - n_1(t-\ell)) \cdot n_1(t-\ell)\langle x, \eta_{01}x\rangle,$$

where $n_1(s)$ denotes the number of abnormal vertices at time s. So, if $t < t_1$,

$$\psi_m^x(t) = \frac{(J_m^x(t) - \overline{J}_m^x(t))/\sqrt{(1 + 1/m)\binom{n}{2}\langle x, \eta_{00}x\rangle}}{\sqrt{\frac{1}{m-1}\sum_{k=1}^m \left(J_m^x(t-k) - \overline{J}_m^x(t)\right)^2}/\sqrt{\binom{n}{2}\langle x, \eta_{00}x\rangle}}$$

and if $t = t_1$,

$$\psi_m^x(t) = \frac{\dfrac{J_m^x(t) - \overline{J}_m^x(t)}{\sqrt{(1+1/m)\binom{n}{2}\langle x, \eta_{00}x\rangle + \binom{n_1}{2}(\langle x, \eta_{11}x\rangle - \langle x, \eta_{00}x\rangle) + (n-n_1)n_1(\langle x, \eta_{01}x\rangle - \langle x, \eta_{00}x\rangle)}}}{\sqrt{\frac{1}{m-1}\sum_{k=1}^m \left(J_m^x(t-k) - \overline{J}_m^x(t)\right)^2}/\sqrt{\binom{n}{2}\langle x, \eta_{00}x\rangle}}.$$

Therefore, when $\binom{n}{2} = n(n-1)/2$ is sufficiently large, $\psi_m^x(t)$ follows the non-central Student t-distribution with degrees of freedom $m-1$ and non-centrality parameter μ^x as defined in (92). □

Let $c_0(\alpha)$ be such that $\mathbf{P}[\psi_m^x(t_1 - 1) > c_0(\alpha)] = \alpha$, and let

$$c(\alpha) = c_0(\alpha)\sqrt{1 + 1/m}, \qquad (94)$$

$$c_1(\alpha) = \frac{c_0(\alpha)}{\sqrt{1 + \frac{m}{m+1}\left(\frac{\binom{n_1}{2}}{\binom{n}{2}}\left(\frac{\langle x, \eta_{11}x\rangle}{\langle x, \eta_{00}x\rangle} - 1\right) + \frac{(n-n_1)n_1}{\binom{n}{2}}\left(\frac{\langle x, \eta_{01}x\rangle}{\langle x, \eta_{00}x\rangle} - 1\right)\right)}}. \qquad (95)$$

As a consequence of Lemma 3, after some brief calculations, we have the following approximations:

$$\mathbf{P}[T_m^x(t_1 - 1) > c(\alpha)] \approx \mathbf{P}[\psi_m^x(t_1 - 1) > c_0(\alpha)] = \alpha, \qquad (96)$$
$$\widetilde{\beta}^x = \widetilde{\beta}_m^x(\alpha) = \mathbf{P}[T_m^x(t_1) > c(\alpha)] \approx \mathbf{P}[\psi_m^x(t_1) > c_1(\alpha)]. \qquad (97)$$

Therefore, $\widetilde{\beta}^x$ is a function of x (explicitly) and α, m, n, n_1, π_0, and π_1 (implicitly).

We are interested in the maximizer of $\widetilde{\beta}^x$ as x ranges over $\mathbb{R}^K \setminus \{0\}$ while all the other parameters being fixed. By the scaling property (cf. (102)), maximizing $\widetilde{\beta}^x$ over $\mathbb{R}^K \setminus \{0\}$ is equivalent to maximizing $\widetilde{\beta}^x$ over the unit circle

$$D = \{x \in \mathbb{R}^K : \|x\| = 1\}, \qquad (98)$$

where $\|\cdot\|$ denotes Euclidean norm. Thus, we wish to identify the set $\arg\max_{x \in D} \widetilde{\beta}^x$. The power approximation $\widetilde{\beta}^x$ of the test as a function of x may be determined by various factors, but the dominating factor is likely to be μ^x. Thus, it is of interest to find the optimizer of μ^x as x ranges over D.

We make some preliminary observations about μ^x as a function of $x \in \mathbb{R}^K \setminus \{0\}$ and to facilitate our discussion, for each $x \in \mathbb{R}^K \setminus \{0\}$, let

$$p(x) = \binom{n_1}{2} \langle x, \pi_{11} - \pi_{00} \rangle + (n - n_1)n_1 \langle x, \pi_{01} - \pi_{00} \rangle,$$

$$q(x) = (1 + 1/m)\binom{n}{2} \langle x, \eta_{00}x \rangle + \binom{n_1}{2} \langle x, \eta_{11}x - \eta_{00}x \rangle + (n - n_1)n_1 \langle x, \eta_{01}x - \eta_{00}x \rangle.$$

By Assumption 1, $q(x) > 0$ for each $x \in \mathbb{R}^K \setminus \{0\}$ and for some $x \in \mathbb{R}^K \setminus \{0\}$,

$$p(x) = \left\langle x, \binom{n_1}{2}(\pi_{11} - \pi_{00}) + (n - n_1)n_1(\pi_{01} - \pi_{00}) \right\rangle > 0. \tag{99}$$

Hence, μ^x is not identically zero on $\mathbb{R}^K$. Next, for each $x \in \mathbb{R}^K \setminus \{0\}$,

$$\mu^x = \frac{p(x)}{\sqrt{q(x)}}, \tag{100}$$

$$\mu^{-x} = \frac{p(-x)}{\sqrt{q(-x)}} = \frac{-p(x)}{\sqrt{q(x)}} = -\mu^x, \tag{101}$$

and analogous to Lemma 2, it can be shown that for each $x \in \mathbb{R}^K \setminus \{0\}$ and $c > 0$, μ^x has the scaling property

$$\mu^{cx} = \mu^x. \tag{102}$$

One can check that the optimizer of μ^x as x ranges over D must satisfy the following set of equations:

$$2\partial_k p(x)q(x) - p(x)\partial_k q(x) = 0, \quad \text{for } k = 1, \ldots, K. \tag{103}$$

To further illustrate this idea, for $K = 2$, we will derive an optimizer of μ^x as x ranges over D. (In Sect. 4.2.2 we investigate the optimizer(s) via simulation and numerical procedures.)
Let

$$a = \left(1 + \frac{1}{m}\right)\binom{n}{2} - \binom{n_1}{2} - (n - n_1)n_1, \tag{104}$$

$$b = \binom{n_1}{2}, \tag{105}$$

$$c = (n - n_1)n_1, \tag{106}$$

and

$$A = a\pi_{00,1}(1 - \pi_{00,1}) + b\pi_{11,1}(1 - \pi_{11,1}) + c\pi_{01,1}(1 - \pi_{01,1}), \tag{107}$$

$$B = a\pi_{00,1}\pi_{00,2} + b\pi_{11,1}\pi_{11,2} + c\pi_{01,1}\pi_{01,2}, \tag{108}$$

$$C = a\pi_{00,2}(1 - \pi_{00,2}) + b\pi_{11,2}(1 - \pi_{11,2}) + c\pi_{01,2}(1 - \pi_{01,2}), \tag{109}$$

$$E = b(\pi_{11,2} - \pi_{00,2}) + c(\pi_{10,2} - \pi_{00,2}), \tag{110}$$

$$F = b(\pi_{11,1} - \pi_{00,1}) + c(\pi_{10,1} - \pi_{00,1}). \tag{111}$$

Theorem 3 *There exists $x^* \in D$ such that $\mu^{x^*} \neq 0$, $\arg\max_{x \in D} \mu^x \in \{x^*, -x^*\}$ and $\arg\min_{x \in D} \mu^x \in \{x^*, -x^*\}$. Moreover, if x^* is neither $(1, 0)$ nor $(-1, 0)$, then*

$$x^* = \frac{(u^*, 1)}{\|(u^*, 1)\|}, \tag{112}$$

where

$$u^* = \frac{BE + FC}{FB + AE}.$$

Proof To begin, define $f : \mathbb{R} \to \mathbb{R}$ by letting, for each $u \in \mathbb{R}$,

$$f(u) = \frac{p_1(u)}{\sqrt{q_1(u)}},$$

where $x(u) = (u, 1)$, $p_1(u) = p(x(u))$ and $q_1(u) = q(x(u))$. Note that $f(u) = \mu^{x(u)}$ and μ^x is continuous on $\mathbb{R}^2 \setminus \{0\}$. Therefore, it follows that f is continuous on $\mathbb{R}$, and moreover, if f does not achieve its optimum on $\mathbb{R}$, then by continuity of f and the odd symmetry of the function μ^x,

$$\sup_{u \in \mathbb{R}} |f(u)| = \lim_{|u| \to \infty} |f(u)| = |\mu^{(1,0)}| = |\mu^{(-1,0)}|. \tag{113}$$

We will first show that if a local optimizer (i.e., maximizer or minimizer) of f exists, then the local optimizer is u^* and hence, unique. Note that both $p_1(\cdot)$ and $q_1(\cdot)$ are polynomials of one variable and hence are continuously differentiable. Moreover, since the unit circle D is compact and $q(x) > 0$ for each $x \in D$, it follows that $\inf_{u \in \mathbb{R}} q_1(u) \geq \inf_{x \in D} q(x) > 0$. For each $u \in \mathbb{R}$, taking the first derivative of f at u, we see that

$$f'(u) = \frac{p_1'(u)\sqrt{q_1(u)} - (1/2)p_1(u)q_1'(u)/\sqrt{q_1(u)}}{q_1(u)} \tag{114}$$

$$= \frac{1/2}{\sqrt{q_1(u)}} \frac{2p_1'(u)q_1(u) - p_1(u)q_1'(u)}{q_1(u)}. \tag{115}$$

In particular, f is differentiable for each $u \in \mathbb{R}$, and f' is continuous on $\mathbb{R}$. Now, suppose that f has a local optimizer on $\mathbb{R}$. To find the local optimizer of f, we find the roots of f'. This is equivalent to solving for $u \in \mathbb{R}$ such that

$$0 = 2p_1'(u)q_1(u) - p_1(u)q_1'(u). \tag{116}$$

Now, let $e_1 = (1, 0)$ and note that for each $u \in \mathbb{R}$,

$$p_1'(u) = \langle e_1, b(\pi_{11} - \pi_{00}) + c(\pi_{01} - \pi_{00}) \rangle, \tag{117}$$

$$q_1'(u) = 2\langle x(u), e_1^T(a\eta_{00} + b\eta_{11} + c\eta_{01}) \rangle. \tag{118}$$

Then, we see that (116) is a quadratic equation, but a routine simplification of (116) will show that (116) is indeed linear in u. In particular, the only root of (116) is u^*. This proves our preliminary claim.

To prove our main claim, we consider two cases. For our first case, suppose that $|f|$ does not achieve its maximum on $\mathbb{R}$. Then, by continuity of μ^x as a function of x, we must have that

$$|\mu^{(1,0)}| = |\mu^{(-1,0)}| = \sup_{u \in \mathbb{R}} |f(u)|. \tag{119}$$

Recall that μ^x is not identically zero as x ranges over D, and $\mu^{-x} = -\mu^x$ for each $x \in D$. Hence, neither $\mu^{(-1,0)}$ nor $\mu^{(1,0)}$ can be 0. Therefore, this implies that one of $(-1, 0)$ and $(1, 0)$ must be the maximizer of μ^x over D and the other must be the minimizer of μ^x over D. Uniqueness follows naturally from our observations so far. For our second case, suppose that f does achieve its optimum on $\mathbb{R}$. Then, by our preliminary claim, we see

that f can have exactly one (local) optimizer over $\mathbb{R}$, and it is u^*. Therefore, we must have $0 \le |f(u)| < |f(u^*)|$ for each $u \in \mathbb{R} \setminus \{u^*\}$, and by the odd symmetry of μ^x together with the fact that f has exactly one local optimizer, it follows that $|\mu^{(1,0)}| = |\mu^{(-1,0)}| < |f(u^*)|$. Our main claim follows from these observations, and this completes our proof. $\qquad\square$

4.2.2 Simulation

We present here an illustrative Monte Carlo experiment using the following parameters:

$$K = 2, n = 100, n_1 = 9, m = 10, t_1 = 12, r = 1024, \tag{120}$$

$$\pi_0 = (0.10, 0.30, 0.60)^T, \tag{121}$$

$$\pi_1 = (0.25, 0.40, 0.35)^T, \tag{122}$$

and

$$Q_0 = \begin{pmatrix} -1 & 0 & 1 \\ 0 & -1 & 1 \\ 1/6 & 1/2 & -2/3 \end{pmatrix}, \tag{123}$$

$$Q_1 = \begin{pmatrix} -1 & 0 & 1 \\ 0 & -1 & 1 \\ 5/7 & 8/7 & -13/7 \end{pmatrix}. \tag{124}$$

One Monte Carlo replicate of this model yields, at the time of the change $t = t_1 = 12$, the latent position vectors presented earlier in Fig. 3.

Power estimates at level $\alpha = 0.05$ for linear size fusion statistics T_m^x for this example are presented in Fig. 4 and Table 1. We consider first approximation asymptotics, developed in Sect. 4.2.1, and first approximation, second approximation, and exact model power estimates obtained via Monte Carlo simulation (with 10,000 Monte Carlo replicates). Fig. 4 shows power as a function of angle θ, where $x = (\cos(\theta), \sin(\theta))$. We consider for illustration four special cases for x, corresponding to $\theta \in \{0, \pi/2, \pi/4, \theta^* \approx 0.263\}$, which correspond to attribute weighting schemes of: consider only attribute 1, consider only attribute 2, consider both attributes equally, the attribute weight ratio of 3.719 : 1. (The asymptotic first-order approximation power analysis yields the analytic maximizer $\theta^* \approx 0.263$; this corresponds to an attribute weight ratio for optimal power of 3.719 : 1.) Power estimates for these four cases are given in Table 1. (The four vertical lines in Fig. 4 correspond to these four cases.)

Because we are considering a one-sided test, and because $\pi_{1,k} > \pi_{0,k}$ for both $k = 1, 2$, the power is maximized in the first quadrant, and $\beta^x < \alpha$ for θ in the third quadrant.

We observe that the power analysis obtained using the approximate models gives a useful indication of comparative power in the exact model. In particular, we see that the first-order

Table 1 Numerical power estimates $\widetilde{\beta}^x$ and $\widehat{\beta}^x$ for $x = (\cos(\theta), \sin(\theta))$

Method	$\theta = 0$	$\theta = \pi/2$	$\theta = \pi/4$	$\theta = \theta^*$
First approx. (asymptotic)	0.550	0.327	0.530	0.687
First approx. (Monte Carlo)	0.536	0.315	0.520	0.681
Second approx. (Monte Carlo)	0.537	0.303	0.511	0.682
Exact (Monte Carlo)	0.534	0.322	0.535	0.684

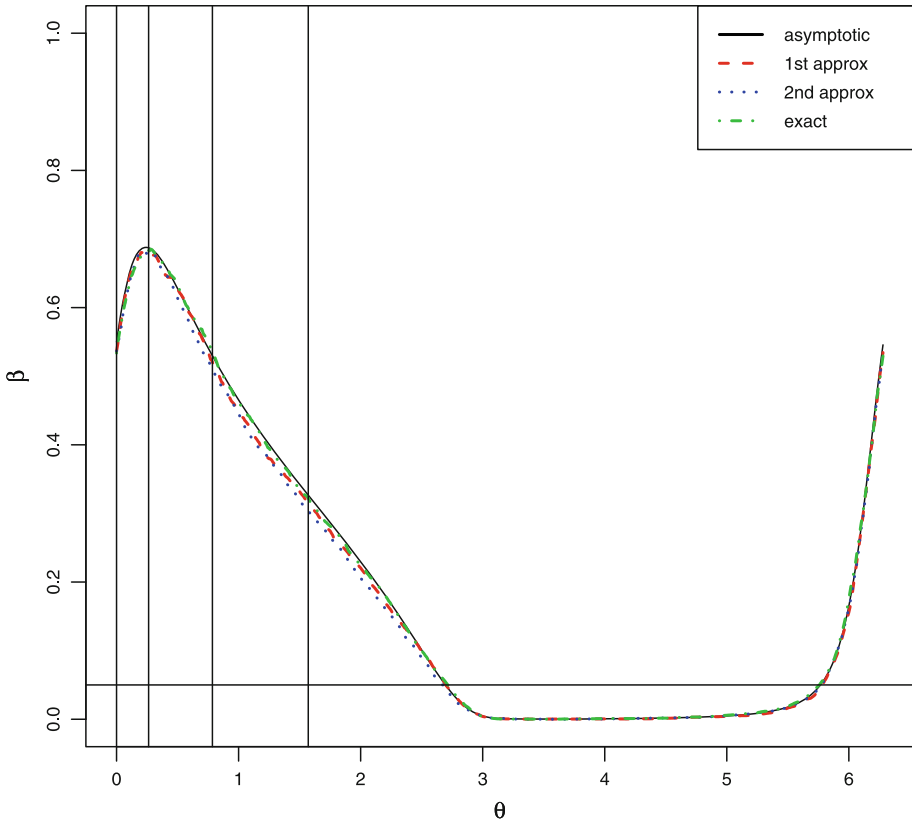10,000 Monte Carlo replicates yields $\sigma_{\widehat{\beta}^x} \le 0.005$

Fig. 4 Power as a function of angle θ for $x = (\cos(\theta), \sin(\theta))$. The plot shows asymptotic estimate $\widetilde{\beta}^x$ (*black*), and first approximation (*red*), second approximation (*blue*), and exact model (*green*) estimates $\widehat{\beta}^x$ via Monte Carlo. (10,000 Monte Carlo replicates yields $\sigma_{\widehat{\beta}^x} \leq 0.005$.) Analytic maximizer from Sect. 4.2.1 yields $\theta^* \approx 0.263$. Numerical power estimates corresponding to the four vertical lines ($\theta \in \{0, \pi/2, \pi/4, \theta^*\}$) are given in Table 1. (Horizontal line is $\alpha = 0.05$)

asymptotically optimal linear size fusion statistic provides statistically significantly greater power than simpler attribute weighting schemes, even for the exact model. Thus, we see that inferential analysis under the approximate models can, at least in some cases, provide useful insight into the behavior of the exact model. (An additional advantage of the two approximate models over the exact model is greatly reduced computational burden for estimating the power of test statistics via Monte Carlo.)

5 Conclusion

We have presented a model in which a collection of (latent) continuous-time finite-state stochastic processes induces a (discrete-time) time series of attributed random graphs. The model is designed for applicability to scenarios comprising multiple modes of association among a collection of actors over time. Our model is illustrated via application to change-point detection. Two mathematically simpler approximations to the model are available, and
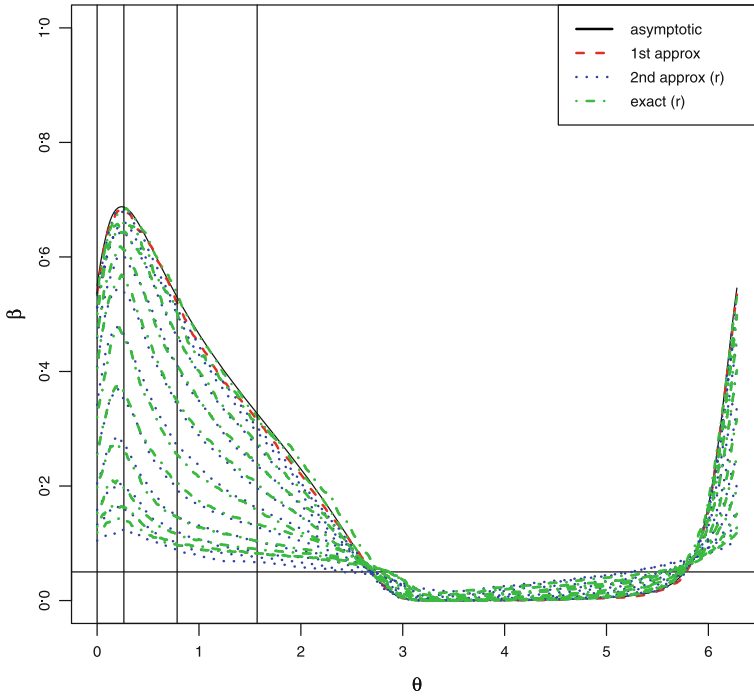
Fig. 5 Power as a function of angle θ for $x = (\cos(\theta), \sin(\theta))$, analogous to Fig. 4. In addition to asymptotic estimate (*black*) and first approximation (*red*), this plot shows second approximation (*blue*) and exact model (*green*) estimates as a function of the vertex process rate parameter $r \in \{2^{\ell}; \ell = 1, 2, \ldots, 10\}$. See text for discussion

Monte Carlo and asymptotic analysis indicates that the approximations provide valuable information for a comparative power analysis of various test statistics.

Various generalizations of the setting considered herein are of interest. First, it is noteworthy that the results presented herein can be immediately generalized to inhomogeneity, in which the "null process" is different for each vertex. The change model involves altering some subset of these vertex processes, and under certain assumptions (e.g., $\pi_{1,k} > \pi_{0,k}$ for all vertices), the linear size fusion statistics considered in Sect. 4 behave much as presented here. Generalization to statistics other than the linear size fusion statistics is obviously of interest. Scan statistics have been considered for anomaly detection in time series of graphs (Priebe et al. 2005; Marchette and Priebe 2008b) and for time series of attributed graphs (Priebe et al. 2010), and comparative power investigations for various statistics under our first approximation demonstrates that the choice of statistic is non-trivial (Pao et al. 2011; Rukhin and Priebe 2011, 2010; Grothendieck et al. 2010). Another generalization involves relaxing the *finite* attribute space assumption. For example, for communications analysis in which the attributes model complex properties of the communication such as topic, it may be useful to consider the attribute space $\mathscr{K} = \mathbb{R}^d$, which requires generalization of our vertex processes.

Analysis for the second approximation, analogous to that performed herein for the first approximation, suggests that a large value of the vertex process rate parameter r results in the second approximation being nearly identical to the first approximation and accurate for the exact model, as proven in Sect. 3 and depicted in Fig. 4. As r decreases, (1) the second

approximation deviates from the first approximation, (2) the second approximation remains accurate for the exact model for moderate r, and (3) eventually the temporal dependence induced by small r results in the second approximation deviating from the exact model. See Fig. 5. This analysis is ongoing.

Estimating null model parameters from data for the first- and second-order approximations, under a (short-time) stationarity assumption, is straightforward. For the exact model, we suggest that Sarkar and Moore (2005); Fu et al. (2009) may provide methods for addressing the temporal dependence induced by small r, and an approach similar to that employed in Liu et al. (2009), wherein estimation of autocorrelation structure of a sequence of truncated multivariate normal vectors is studied, may prove appropriate.

In conclusion, our latent process model for time series of attributed random graphs provides a useful framework for understanding the behavior of inference procedures applied to complex data sets consisting of multiple modes of association over time among a collection of actors.

References

Airoldi E, Blei D, Fienberg S, Xing E (2008) Mixed membership stochastic blockmodels. J Mach Learn Res 9:1981–2014

Asmussen S (2003) Applied probability and queues. Springer, New York

Bollobas B (2001) Random graphs. Cambridge University Press, Cambridge

Bollobas B, Janson S, Riordan O (2007) The phase transition in inhomogeneous random graphs. Random Struct Algorithm 31:3–122

Doreian P, Batagelj V, Ferligoj A (2005) Generalized blockmodeling. Cambridge University Press, Cambridge

Fu W, Song L, Xing EP (2009) Dynamic mixed membership blockmodel for evolving networks. In: ICML '09: proceedings of the 26th annual international conference on machine learning, New York, pp 1–8

Glynn PW, Whitt W (1992) The asymptotic efficiency of simulation estimators. Oper Res 40:505–520

Goldenberg A, Zheng AX, Fienberg SE, Airoldi EM (2010) A survey of statistical network models. Found Trends Mach Learn 2:129–233

Grothendieck J, Priebe CE, Gorin AL (2010) Statistical inference on attributed random graphs: fusion of graph features and content. Comput Stat Data Anal 54:1777–1790

Handcock MS, Raftery AE, Tantrum JM (2007) Model-based clustering for social networks. J Royal Stat Soc A 170(2):301–354

Hoff P, Raftery AE, Handcock MS (2002) Latent space approaches to social network analysis. J Am Stat Assoc 97:1090–1098

Keilson J (1998) Covariance and relaxation time in finite Markov chains. J Appl Math Stoch Anal 11(3): 391–396

Kolaczyk ED (2009) Statistical analysis of network data. Springer, New York

Kraetzl M, Nickel C, Scheinerman ER (submitted) Random dot product graphs: a model for social networks

Liu X, Daiels M, Marcus B (2009) Joint models for the association of longitudinal binary and continuous process with application to a smoking cessation trial. J Am Stat Assoc 104:429–438

Marchette DJ, Priebe CE (2008a) Predicting unobserved links in incompletely observed networks. Comput Stat Data Anal 52(3):1373–1386

Marchette DJ, Priebe CE (2008b) Scan statistics for interstate alliance graphs. Connections 28(2):43–68

Olding BP, Wolfe PJ (submitted) Inference for graphs and networks: Extending classical tools to modern data

Pao H, Coppersmith GA, Priebe CE (2011) Statistical inference on random graphs: comparative power analyses via monte carlo. J Comput Graph Stat 20(2):395–416

Parzen E (1999) Stochastic processes. SIAM, Philadelphia

Porter MA, Onnela JP, Mucha PJ (2009) Communities in networks. Notices Am Math Soc 56(9):1082–1097

Priebe CE, Conroy JM, Marchette DJ, Park Y (2005) Scan statistics on Enron graphs. Comput Math Organ Theory 11:229–247

Priebe CE, Park Y, Marchette DJ, Conroy JM, Grothendieck J, Gorin AL (2010) Statistical inference on attributed random graphs: fusion of graph features and content: An experiment on time series of enron graphs. Comput Stat Data Anal 54:1766–1776

Rukhin A, Priebe CE (2010) On the limiting distribution of a graph scan statistic. Commun Stat Theory Meth (forthcoming)

Rukhin A, Priebe CE (2011) A comparative power analysis of the maximum degree and size invariants. J Stat Plan Infer 141:1041–1046

Sarkar P, Moore AW (2005) Dynamic social network analysis using latent space models. SIGKDD Explor Newsl 7(2):31–40

Scheinerman ER, Tucker K (2010) Modeling graphs using dot product representations. Comput Stat 25(1): 1–16

Young SJ, Scheinerman ER (2009) Directed random dot product graphs. Internet Math 5:89–101