A latent process model for time series of attributed random graphs

Nam H. Lee
The Johns Hopkins University

joint work with
Carey E. Priebe

Accepted for publication, Nov 2010

Statistical Inference for Stochastic Processes
Overview

- Comparative Power Analysis Experiment (using approximations)
- Parameter estimation through "diffusion" approximation

Thanks to Collaborators

Youngser Park
Lucy Robinson
Glen Coppersmith
John Grothendick
Change point detection problem

N actors' usual (individual) flow of daily interest pattern

Some of N actors may break out of the usual pattern while the rest continue with the usual pattern

Goal: at time T+1, using the data collected from the social interaction observed for [T, T+1), determine if "break out" has occurred.
Time series of attributed random graphs

\[ \{t-2, t-1\} \quad \{t-1, t\} \quad \{t, t+1\} \quad \{t+1, t+2\} \]

\( n = \# \text{ of vertices} \)

Edge color = the nature of the connection (seen)

Vertex State = the nature of the actor (unseen)

\[ [A_{ij}(t)] = \text{Symmetric hollow matrix} \]

such that \( A_{ij}(t) \neq 0 \)

iff actor \( i \) and actor \( j \) had a "live" social connection during \( [t-1, t) \)
Modeling graphs using dot product representations

Edward R. Scheinerman · Kimberly Tucker

Abstract  Given a simple (weighted) graph, or a collection of graphs on a common vertex set, we seek an assignment of vectors to the vertices such that the dot products of these vectors approximate the weight/frequency of the edges. By transforming vertices into (low dimensional) vectors, one can bring geometric methods to bear in the analysis of the graph(s). We illustrate our approach on the Mathematicians Collaboration Graph [Grossman (1996) The Erdős number project, http://www.oakland.edu/enp] and the times series of Interstate Alliance Graphs (Gibler and Sarkees in J Peace Res 41(2):211–222, 2004).

Keywords  Social networks · Dimension reduction · Vector representations of graphs
$P_{ij}(t) =$ probability that actor $i$ and
actor $j$ being connected for $[t-1,t)$

\[ = \sum_{k=1}^{K-1} X_{i,k}(t) \cdot X_{j,k}(t) \]

\[ = \| x_i^*(t) \| \cdot \| x_j^*(t) \| \cdot \cos(\theta(t)) \]
Vertex (latent position) process

$V_i(t) = \text{The color (i.e. the latent position) of actor } i \text{ (i.e. vertex } i \text{) at continuous time } t \in [0, T+1]$  

$X_i(t) = \int_{t-1}^{t} g(V_i(s)) \, ds,$  

where $g : E \to \{ (1,0,\ldots,0)^T, (0,1,0,\ldots,0)^T, \ldots, (0,0,\ldots,0,1)^T \}$  

$k$-coordinates
Change point detection problem

Suppose that vertex \( i \) changes its transition rule at time \( T \). Now, detect the change from the social activities (i.e., the signal) observed.

\( Q_i(t) \) = the infinitesimal transition rate matrix for interval \([t-1, t]\)

\[
Q_0 = \begin{bmatrix}
-1 & \frac{1}{2} & \frac{1}{2} \\
\frac{1}{2} & -1 & \frac{1}{2} \\
\frac{1}{2} & \frac{1}{2} & -1
\end{bmatrix} \quad \text{normal (null)}
\]

\[
Q_1 = \begin{bmatrix}
-10 & 0 & 10 \\
0 & -10 & 10 \\
\frac{1}{20} & \frac{1}{20} & -\frac{1}{10}
\end{bmatrix} \quad \text{abnormal (att)}
\]
Change point detection problem

Suppose that

for $i = 1, \ldots, m$, $Q_i(t) = \begin{cases} Q_0 & \forall t = 1, 2, \ldots, T \\ Q_1 & @ t = T+1 \end{cases}$

for $i = m+1, \ldots, n$, $Q_i(t) = \begin{cases} Q_0 & \forall t = 1, 2, \ldots, T \\ Q_0 & @ t = T+1 \end{cases}$

Goal: Devise a statistical procedure that detects this using $[A_{ij}(1), \ldots, A_{ij}(T), A_{ij}(T+1)]$ w/ a small type II error
Comparative power analysis of fusions

\[ J^x(t) = \sum_{k=1}^{K} \omega_k |A(t)|_k \]

\[ \bar{J}^x(t) = \frac{1}{L} \sum_{l=1}^{L} J^x(t-l) \]

\[ \psi^x(t) = \frac{J^x(t) - \bar{J}^x(t)}{\sqrt{\frac{1}{L-1} \sum_{l=1}^{L} (J^x(t-l) - \bar{J}^x(t))^2}} \]
Comparative power analysis of fusions

\[ x = (\omega_1, \omega_2) \]

\[ J^{c,x}(t) = c \cdot J^x(t) \]

\[ J^{x+y}(t) = J^x(t) + J^y(t) \]

\[ \psi^{c,x}(t) = \psi^x(t) \]

Finding optimal \( x \in \mathbb{R}^k \)

= Finding optimal \( x \in \mathbb{R}^k \) s.t. \( \|x\|_2^2 = 1 \),

where “optimality” is with respect to the power of the fusion statistic \( \psi^x \)

How much did “temporal dependence” affect our comparative power analysis of the fusions?
Models for approximating a sequence of exact models: "LLN" & "CLT" approx

Fix \( n, m, T, Q_0, Q_A \), and let

\[
M^r = (n, m, T, rQ_0, rQ_1).
\]

The \( r \)-th exact model

\[
\{ V_1^r(t) \} \perp \perp \{ V_2^r(t) \} \perp \ldots \perp \{ V_n^r(t) \} \\
\downarrow \\
\{ X_1^r(t) \} \perp \perp \{ X_2^r(t) \} \perp \ldots \perp \{ X_n^r(t) \} \\
\downarrow \\
[A_{ij}^r(0)], [A_{ij}^r(1)], \ldots, [A_{ij}^r(T)], [A_{ij}^r(T+1)]
\]

1st approximation (LLN approx.)

\( \Pi_0 = \) invariant probability vector of \( Q_0 \)

\( \Pi_A = \) invariant probability vector of \( Q_1 \)

\( \bar{A}_{ij}(t) = \begin{cases} 1 & \text{w/ prob. } \pi_{i,1}(t) \pi_{j,1}(t) \\ 2 & \text{w/ prob. } \pi_{i,2}(t) \pi_{j,2}(t) \\ \vdots \\ K-1 & \text{w/ prob. } \pi_{i,K-1}(t) \pi_{j,K-1}(t) \\ 0 & \text{otherwise} \end{cases} \)
Models for approximating a sequence of exact models: LLN & CLT approx

\[ r\text{-th second approximation} \]

\[
\xi_0 = (1\pi_0^T - Q_0)^{-1}(\text{diag}(1) - 1\pi_0^T),
\]

\[
\xi_1 = (1\pi_1^T - Q_1)^{-1}(\text{diag}(1) - 1\pi_1^T).
\]

\[
\Sigma_0 = \text{diag}(\pi_0)\xi_0 + \xi_0^T \text{diag}(\pi_0),
\]

\[
\Sigma_1 = \text{diag}(\pi_1)\xi_1 + \xi_1^T \text{diag}(\pi_1).
\]

\[ \tilde{X}_{i_r}(t) = \text{The truncation of a multivariate normal random vector} \]

\[ \text{with mean vector } \pi_{i_r}(t) \text{ and covariance matrix } \Sigma_{i_t}/r \]

\[
\{ \tilde{X}_{i_1}(t) \} \uparrow \{ \tilde{X}_{i_2}(t) \} \uparrow \ldots \uparrow \{ \tilde{X}_{i_m}(t) \}
\]

\[
[\tilde{A}_{ij}(1)] \uparrow [\tilde{A}_{ij}(2)] \uparrow \ldots \uparrow [\tilde{A}_{ij}(r)] \uparrow [\tilde{A}_{ij}(r+1)]
\]
As $r \to \infty$, for each $i$,

$$\sqrt{r} \left( X_i^r(t) - \pi_i^r(t) \right) \Rightarrow D_i(t),$$

where

(i) $D_i(t)$ is a $\text{Normal}(0, \Sigma_i(t))$

(ii) $D_i(t) \perp D_i(s)$ for $t < s$. 

![Graph showing various approximations and exact results.](image)
Observed quantities are records about who talks to whom about what at what time

\[ M = \{ M_e : e \in \mathbb{N} \} , \]

where

\[ M_e = ( \tau_e, \{ U_e, V_e \}, K_e ) \]

\[ \tau_e < \tau_{e+1} \]

\[ U_e, V_e \in \{ 1, 2, 3, \ldots, n \} \]

\[ K_e \in \{ 1, 2, \ldots, K \} \]
Model Estimation Problem

The unseen process is the superposition of $n(n-1)/2$ independent homog. Poisson processes with unit rate.

The seen (i.e., observed) process is the superposition of the $n(n-1)/2$ count processes, each of which is a thinned & marked version of an unseen homogeneous Poisson process.
\( N_{ij,k}(t) = \# \text{ of records between actor } i \text{ and actor } j \text{ about topic } k \text{ prior to time } t \)

\[ N_{ij}(t) = \sum_{k=1}^{K} N_{ij,k}(t) \]

\( = \text{ inhomogeneous Poisson process} \)

\( \text{with its (non-deterministic) rate function } \dot{N}_{ij}(t) \)

\[ \dot{N}_{ij}(t) = \left< X_i(t), X_j(t) \right>, \text{ where} \]

\[ X_i(t) = \int_{t-1}^{t} g(Y_i(s)) \, ds \]

\[ Y_i(t) = \text{c.t.f.s. Markov process w/ its infin. trans. rate matrix} \]

\[ Q_i \]
Model Estimation Problem

The rate function for actor \( i \) & actor \( j \)

- the more actor \( i \) and actor \( j \) are similar, the less likely to be thinned.

\[ \Lambda_{ij}^r(t) = \langle X_i^r(t), X_j^r(t) \rangle \]

The likelihood \( f_{[0,T]}^{c_{ij}}(t_1, t_2, \ldots, t_m) \) of observing events at time \( t_1, t_2, \ldots, t_m \)

\[ \propto \Lambda_{ij}(t_1) \cdots \Lambda_{ij}(t_m) \cdot \exp\left(-\int_0^{t_m} \Lambda_{ij}^r(s) \, ds \right) \cdot \mathcal{S}_{[0,T]}^{c_{ij}}(X_i^r, X_j^r) \]
Model Estimation Problem

A $K$-dimensional Jacobi process is an Itô process $J = (J_1, J_2, ..., J_K)$ s.t.

$$
\begin{align*}
dJ_k(t) &= \beta (J_k(t) - \Pi_k) \, dt \\
&\quad + \sigma \sqrt{J_k(t)} \, dW_k(t) \\
&\quad - \sigma J_k(t) \sum_{l=1}^{K} \sqrt{J_l(t)} \, dW_l(t),
\end{align*}
$$

where $\beta \in (-\infty, 0)$, $\sigma \in (0, \infty)$, $\Pi_k \in (0, 1)$, and $\sum_{k=1}^{K} \Pi_k = 1$,

and $W_1, W_2, ..., W_K$ are independent standard Brownian motion.

For sufficiently large $r$, we approximate each $x_i^r$ as a Jacobi process with

$$
\phi^r = (\beta_i^r, \Pi_{i,1}^r, ..., \Pi_{i,k}^r, \sigma_i^r).
$$
Model Estimation Problem

$r$ small

$r$ large
Model Estimation Problem

\[ dX_i(t) = \beta_i \cdot (X_i(t) - \pi_i) \, dt \]
\[ + \sigma_i \sqrt{X_i(t)(1-X_i(t))} \, dB_i(t) \]

The likelihood of observing events at time \( t_1, t_2, \ldots, t_m \) between actor \( i \) and actor \( j \) is proportional to

\[ \Lambda_{i,j}(t_1) \cdots \Lambda_{i,j}(t_m) \cdot \exp\left(-\int_0^{t_m} \Lambda_{i,j}(s) \, ds\right) \cdot f_{[0, t_1]}(X_i^r) \cdot f_{[0, t_m]}(X_j^r) \]

where \( f_{[0, t]} \) denotes the Radon-Nikodym density of \( X_i^r \) with respect to the standard Brownian motion.

21/24
Model Estimation Problem

EM algorithm

$\phi_n^r = \text{the } n\text{-th estimate of } \phi^r$

$\phi_{n+1}^r = \arg\max \ E_{\phi_n^r, t_1, \ldots, t_m} \left[ \log L_{\hat{\phi}}^r \right]$

Need to know

how to sample $X_i^r, X_j^r$ conditioning on $t_1, t_2, \ldots, t_m$ and $\psi_n^r$

- Rejection sampling
- Avoid high rejection rate
Model Estimation Problem
Thanks!