

An implied latent position process for doubly stochastic messaging activities

Nam H. Lee* and Carey E. Priebe† and Minh Tang‡

Department of Applied Mathematics and Statistics

Johns Hopkins University

Baltimore, MD 21218, USA

*nhlee@jhu.edu and †cep@jhu.edu and ‡mtang10@jhu.edu

Abstract—This paper studies the problem of identifying an inhomogeneous interaction structure amongst social agents. We construct the social network by a random graph and model the messaging activities via a multi-channel self-exciting point process. We design a methodology that divides the agents into two disjoint groups so that members within each group are considered to be of similar attributes. Our methodology and algorithm are useful for investigating and detecting abnormal activities within a network. We provide numerical illustrations based on a large email dataset from Enron.

Keywords- Social network; Multiple self-exciting point processes; Hypothesis testing; Risk mitigation.

In this paper, we propose a model to estimate and analyze the structure of messaging activities in a social network. This is motivated by the recent proliferation of mobile technology, along with spread of blogs, social networking site, and media-sharing technology. For classification, detection, tracking and other practical purposes, robust statistical analysis as well as a good understanding of the data structure are essential. In this paper, we consider a collection of messaging data, made public by the Federal Energy Regulatory Commission in 2003, that contains highly accurate information about the time at which each message was exchanged. We introduce a meaningful way to reduce messaging data to a random graph and explore its possible application to a community detection problem.

A simple and popular existing method to achieve this is to “pairwise threshold”, where for each pair of agents, an edge between vertex i and vertex j is formed if the number of messaging events between them exceeds a certain threshold. Such graphs are often thought to reveal a structure of an underlying social dynamic, motivating several successful models for a social network with sub-communities, and many tools for detecting a community with a particular graph theoretic and statistical properties have been proposed (Goldenberg et al. 2010, Kolaczyk 2009).

On the other hand, some recent research (De Choudhury et al. 2010), has documented that changing the thresholds in the reduction procedure can produce dramatically different graphs, resulting in vastly dissimilar communities. This issue has motivated the work such as Heard et al. (2010) and Perry and Wolfe (2010). In both studies, as a remedy, the messaging events are modeled by way of point processes.

In Heard et al. (2010), a piecewise-constant interaction rate is considered, while in Perry and Wolfe (2010) a Cox multiplicative intensity model is used with covariates that depend on the history of the process.

Our approach is to model the dynamic network via n low-dimensional latent processes. One challenge to overcome is to estimate the intensity processes. Given an estimate of the intensity processes, we propose a novel embedding methodology to facilitate inference. Based on our methodology, we can produce a random graph with a particular probabilistic structure that could be very useful for the purpose of community detection.

I. MESSAGING EVENTS LABELED WITH LINEARLY-ORDERED RISK LEVELS

We consider a network of n vertices, and denote $\mathbb{V} = \{1, \dots, n\}$. For $\ell \in \mathbb{N}$, we denote by τ_ℓ , $\{i_\ell, j_\ell\}$ and k_ℓ , respectively, the occurrence time, the messaging pair and the risk level of the ℓ -th message. Collectively, $d_\ell = (\tau_\ell, \{i_\ell, j_\ell\}, k_\ell)$ represents the ℓ -th messaging event. We require that

$$\tau_\ell < \tau_{\ell+1} \quad \text{and} \quad 1 \leq i_\ell \neq j_\ell \leq n.$$

The risk levels $\mathbb{K} = \{1, 2, \dots, K\}$ is assumed to be linearly ordered, and we will say that topic label $k_1 \in \mathbb{K}$ is more riskier than topic label $k_2 \in \mathbb{K}$ if $k_1 > k_2$.

For each (undirected) pair ij of the vertices and $t \in [0, T]$, we denote by $N_{ij,k}(t)$ the number of (undirected) messaging events on the topic k , between vertex i and vertex j during $[0, t]$. For each $t \in [0, T]$, let $\mathcal{D}_k(t)$ be the collection of all communication messaging events by time t on topic label k , i.e.,

$$\mathcal{D}_k(t) = \left\{ (\tau_{k,\ell}, \{i_{k,\ell}, j_{k,\ell}\}, k) : \ell = 1, \dots, \sum_{i < j} N_{ij,k}(t) \right\},$$

where $\tau_{k,\ell}$ is the time at which the ℓ -th messaging event labeled with k occurred and $i_{k,\ell}$ and $j_{k,\ell}$ have a similar interpretation. Then, collectively, we have the entire corpus:

$$\mathcal{D}(T) = \cup_{k=1}^K \mathcal{D}_k(T).$$

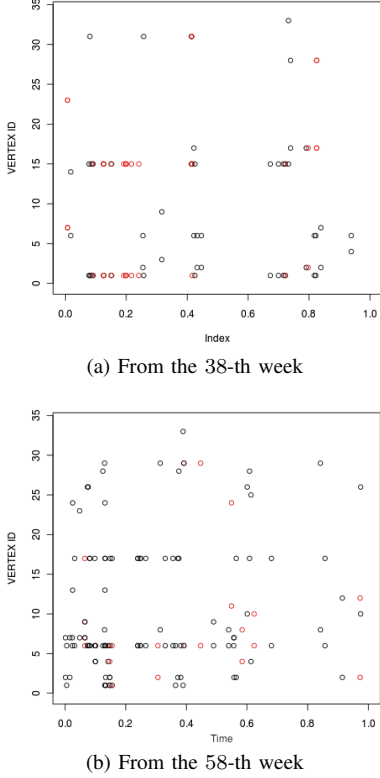


Figure 1: Timeplots of a subset of Ernon E-mail Corpus Data. The entire corpus involves 184 individuals, but for simplicity, only 34 individuals are used to make these plots. The vertical coordinates indicate the labels that identify the thirty four vertices. If a vertical line was to be drawn, then either the vertical line passes no circle or exactly two circles, which represent the sender-receiver pair of a particular e-mail, and the (common) color of the circles represents the topic of the e-mail. For each graph, time is scaled so that 0 denotes the beginning of a week and 1 denotes the end of the week.

For future reference, for each pair ij and topic k , let

$$\mathcal{D}_{ij,k}(t) = \left\{ (\tau_{ij,k}(\ell), i, j, k) : \ell = 1, \dots, \sum_{i < j} N_{ij,k}(t) \right\},$$

where $\tau_{ij,k}(\ell)$ is the time at which the ℓ -th messaging event labeled with k occurred between i and j .

II. MATHEMATICAL MODEL AND PROBLEM STATEMENT

Let $\mathcal{M}(n, d)$ be the set of all $n \times d$ matrices for some fixed d . Let $X = (X_1, \dots, X_K)$ be the solution of SDE:

$$\begin{aligned} \text{vec}(X(t)) &= \text{vec}(X(0)) \\ &+ \int_0^t b(\text{vec}(X(s))) ds + \int_0^t \sigma(\text{vec}(X(s))) dB_s, \end{aligned}$$

where B_t denotes a multivariate standard Brownian motion and for $M = (M_1, \dots, M_K) \in \mathcal{M}(n, d)^K$, $\text{vec}(M)$ is the column vector obtained by stacking the rows of M from the top row to the bottom row, starting from M_1 and ending with M_K . Moreover, we assume that each $X_{i,k}(t)$ has finite second moment and a density.

Let $N(t) = (N_{ij,k}(t))$ be a collection of doubly stochastic point processes for which, the intensity process for pair ij and topic k is:

$$\lambda_{ij,k}(t) = \lambda_0 \exp(-\|X_{k,i}(t) - X_{k,j}(t)\|),$$

where $\|\cdot\|$ is the usual Euclidean norm. In particular, we have

$$\mathbf{P}(N_{ij,k}(t+h) = N_{ij,k}(t) + 1 | \mathcal{F}_t) = \lambda_{ij,k}(t)h + o(h),$$

where $o(h)$ is a term that becomes negligible in comparison to $\lambda_{ij,k}(t)h$ as $h \rightarrow 0$. Note that each

$$\mathcal{N}_k(T) = \{N_{ij,k}(t) : 1 \leq i < j \leq n, t \in [0, T]\}$$

generates $\mathcal{D}_k(T)$. In summary, the collection of $n(n-1)/2$ point processes

$$\mathcal{N}(T) = \{N_{ij,k}(\cdot) : 1 \leq i < j \leq n, k = 1, \dots, K\}$$

generates the entire corpus $\mathcal{D}(T)$. Finally, note that $\mathcal{N}(T)$ is generated by $X = (X_1, \dots, X_n)$, where X_k is a stochastic process taking values in \mathbb{R}^d . Henceforth, we will say that the $n \times d$ dimensional process X generates the corpus $\mathcal{D}(T)$.

Recovering the original X based on N is challenging if not impossible. In this paper, we develop an alternative characterization of N_t in terms of another latent process that is amenable to conducting estimation, inference and model checking.

III. IMPLIED LATENT POSITION BASED ON MULTIPLE POINT PROCESSES

Let \mathcal{S} be the set of all $n \times n$ symmetric matrices such that:

- (i) the diagonal elements are zero,
- (ii) the off-diagonal elements are non-negative.

For each t and topic k , let $D_k(t)$ be the symmetric matrix such that for each pair of $i, j \in \mathbb{V}$ with $i < j$,

$$D_{k,ij}(t) = -\log(\lambda_{ij,k}(t)/\lambda_0).$$

Note that, when $\Lambda_{ij,k}(t)$ is large, $D_{ij,k}(t)$ is small. This incorporates the idea that the more intensely vertex i and vertex j talk about topic k , the more closely vertex i and vertex j are aligned in topic k space.

In addition, it is well known that there exists $\mathbf{X}_k(t)$ such that, when $\mathbf{D}_k^{(2)}(t)$ is the matrix obtained by squaring each entry of $\mathbf{D}_k(t)$,

$$\begin{aligned}\mathbf{X}_k(t)\mathbf{X}_k^t(t) &= -\frac{1}{2}(\mathbf{I} - \mathbf{1}\mathbf{1}^t/n)\mathbf{D}_k^{(2)}(t)(\mathbf{I} - \mathbf{1}\mathbf{1}^t/n), \\ D_{k,ij}(t) &= \|\mathbf{X}_{k,i}(t) - \mathbf{X}_{k,j}(t)\|.\end{aligned}$$

Note that such a decomposition is not unique because one can always use an orthogonal transformation to produce another one. In other words, depending on a particular choice of decomposition rule, $\mathbf{X}_k(t)$ and $X_k(t)$ may be different from each other by an orthogonal transformation and a rigid motion translation.

One important main observation to note here is that, given the collection

$$\{\lambda_{i,j,k}(t) : 1 \leq i < j \leq n, 1 \leq k \leq K\},$$

we can always construct $\mathbf{X}_1(t), \dots, \mathbf{X}_K(t)$. This is true even when there is no explicit underlying processes $X_1(t), \dots, X_K(t)$ as we have posited in the last section.

We now define a function ψ^* that maps each element of \mathcal{S} to an element of $\mathcal{M}(n, d)$, which is the set of all $n \times d$ matrices for some fixed d .

To get around the aforementioned non-uniqueness issue due to the equivalent orthogonal rotations, we fix $Z^* \in \mathcal{M}(n, d)$ to be a matrix such that for each pair $i < j$, $\|Z_i^* - Z_j^*\| > 0$. Then, for each $M \in \mathcal{S}$, let

$$\psi(M) = \arg \min_{X \in \mathcal{M}(n, d)} \|\tau(M) - XX^T\|_F^2,$$

where $\|\cdot\|_F$ denotes the Frobenius norm and

$$\tau(M) = -\frac{1}{2}(\mathbf{I} - \mathbf{1}\mathbf{1}^t/n)M^{(2)}(\mathbf{I} - \mathbf{1}\mathbf{1}^t/n).$$

For each $M \in \mathcal{S}$, being a real symmetric matrix, the geometric multiplicity of M is the same as the algebraic multiplicity of M . In particular, the column vectors should be all linearly independent, and each $X \in \psi(M)$ has the full rank. Note also that for each $M \in \mathcal{S}$, $\psi(M)$ is closed and bounded, whence it is compact.

For each $M \in \mathcal{S}$, let

$$\psi^*(M) = \arg \min_{X \in \psi(M)} \|X - Z^*\|_F^2.$$

We show now that our definition of ψ^* yields a well-defined continuous embedding of $M \in \mathcal{S}$ as an element of $\mathcal{M}(n, d)$.

Theorem 1: The mapping $M \rightarrow \psi^*(M)$ defines a continuous function on \mathcal{S} .

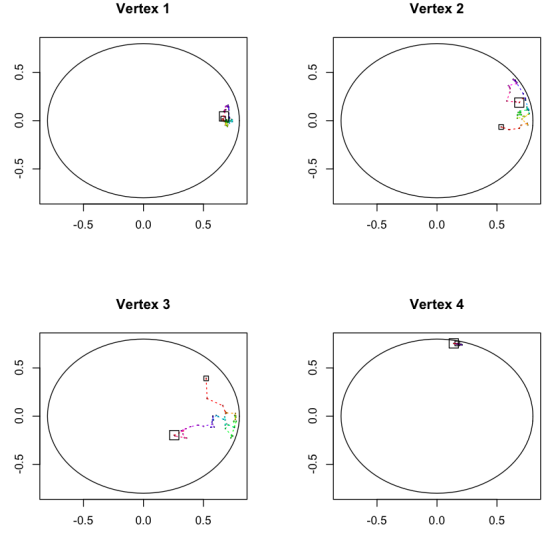


Figure 2: Time-plots for the implied latent position of Vertex 1 through Vertex 4, where the latent position of each vertex is computed from the estimated (pan-topic) intensity process via a 2-dimensional embedding. The smaller box locates the initial location of vertices, and the bigger box locates the final location of vertices

Proof: Assume that as $M_\varepsilon \rightarrow M_0$ as $\varepsilon \rightarrow 0$. We shall show that

$$\psi^*(M_0) = \lim_{\varepsilon \rightarrow 0} \psi^*(M_\varepsilon).$$

First, we establish a preliminary fact. For this, for each ε , fix an arbitrary $Y_\varepsilon \in \psi(M_\varepsilon) \neq \emptyset$. Now, by taking a subsequence necessary, we may assume that as $\varepsilon \rightarrow 0$, we have

$$Y^{(\varepsilon)} \rightarrow Y^{(0)} \in \psi(M_0).$$

Let $X \in \psi(M_0)$. We claim that there exists an orthogonal matrix $Q^{(0)}$ such that

$$X = Y^{(0)}Q^{(0)}. \quad (1)$$

First, because $XX^T = Y^{(0)}(Y^{(0)})^T$ and $X^T X$ is invertible (X is of full column rank), we have

$$X = Y^{(0)}(Y^{(0)})^T X (X^T X)^{-1}. \quad (2)$$

Because $Y^{(0)}$ is also of full column rank and $(Y^{(0)})^T X (X^T X)^{-1}$ is a $d \times d$ matrix, it follows that $(Y^{(0)})^T X (X^T X)^{-1}$ is of rank d and hence invertible.

Thus, $X = Y^{(0)}Q$ for some invertible matrix Q . From $XX^T = Y^{(0)}(Y^{(0)})^T$ we have

$$\text{tr}(Y^{(0)}(QQ^T - I)(Y^{(0)})^T) = 0. \quad (3)$$

Once again by the fact that $Y^{(0)}$ is of full-column rank, implies $QQ^T = I$ and thus Q is orthogonal. This proves our claim in (1).

We now show that

$$\lim_{\varepsilon \rightarrow \infty} \psi^*(M_\varepsilon) = \psi^*(M_0). \quad (4)$$

For this, we take an arbitrary convergent subsequence of $\{\psi^*(M_\varepsilon)\}$ but for simplicity, we suppress this aspect from our notation with understanding that the limit is to be taken along the (arbitrary chosen) convergent subsequence. Now, it can be shown that $\lim_{\varepsilon \rightarrow 0} \psi^*(M_\varepsilon) \in \psi(M_0)$. Hence, from our earlier calculation, we see that there exists an orthogonal matrix Q such that

$$\left(\lim_{\varepsilon \rightarrow 0} \psi^*(M_\varepsilon) \right) Q = \psi^*(M_0). \quad (5)$$

Now,

$$\begin{aligned} \left\| \lim_{\varepsilon \rightarrow 0} \psi^*(M_\varepsilon) - Z^* \right\|_F^2 &\geq \left\| \psi^*(M_0) - Z^* \right\|_F^2 \\ &= \left\| \lim_{\varepsilon \rightarrow 0} \psi^*(M_\varepsilon) Q - Z^* \right\|_F^2 \\ &= \lim_{\varepsilon \rightarrow 0} \left\| \psi^*(M_\varepsilon) Q - Z^* \right\|_F^2 \\ &\geq \lim_{\varepsilon \rightarrow 0} \left\| \psi^*(M_\varepsilon) - Z^* \right\|_F^2 \\ &\geq \left\| \lim_{\varepsilon \rightarrow 0} \psi^*(M_\varepsilon) - Z^* \right\|_F^2. \end{aligned}$$

In other words, we have:

$$\left\| \lim_{\varepsilon \rightarrow 0} \psi^*(M_\varepsilon) - Z^* \right\|_F^2 = \left\| \psi^*(M) - Z^* \right\|_F^2.$$

Then by strict convexity of $\|\cdot\|_F^2$ along with the facts that (i) all of the convergent subsequences share the common limit, and (ii) each subsequence has a convergent subsequence, we have our claim in (4).

In summary, this proves the (pointwise) continuity of ψ^* . ■

IV. RANDOM GRAPHS BASED ON A SELF-EXCITING REPRESENTATION OF A DOUBLY STOCHASTIC PROCESS

A. The classical Zakai equation

In this section, we propose an alternative to estimating $\lambda_{ij,k}(t)$ which is often a challenging if not impossible. For this, we will use a known fact that with respect to the internal history

$$\mathcal{F}_t = \sigma(N(s) : s \in [0, t]),$$

each $N_{ij,k}$ is also a self-exciting process whose intensity function is

$$\widehat{\lambda}_{ij,k}(t) \equiv \mathbf{E}[\lambda_{ij,k}(t) | \mathcal{F}_t].$$

For reference, see [Snyder \(1975, Chapter 5\)](#). Now, note also that:

$$\begin{aligned} \widehat{\lambda}_{ij,k}(t) &= \mathbf{E}[\lambda_{ij,k}(t) | \mathcal{F}_t] \\ &= \lambda_0 \mathbf{E}[\exp(-D_{ij,k}(t)) | \mathcal{F}_t] \\ &= \lambda_0 \mathbf{E} \left[\exp \left(-\sqrt{\sum_{\ell=1}^d |X_{i,k,\ell}(t) - X_{j,k,\ell}(t)|^2} \right) | \mathcal{F}_t \right]. \end{aligned}$$

Hence, $\widehat{\lambda}_{ij,k}(t)$ can be written in terms of the conditional distribution π_t of X_t given \mathcal{F}_t , where \mathcal{F}_t denotes the observation until time t .

The conditional distribution π_t can be computed/estimated in various ways. We briefly summarize the main idea of these methods. For this, let \mathcal{C}_b^2 denote the set of all bounded and twice continuously differentiable functions. Also, let \mathcal{L} is the second order differential operator:

$$\mathcal{L} = \sum_i b_i(x) \partial_i + \frac{1}{2} \sum_{i,j} a_{ij}(x) \partial_i \partial_j,$$

and $a(x) = \sigma(x)\sigma(x)^t$, and denote by $a_{ij}(x)$ and $b_i(x)$, the entry of $a(x)$ and $b(x)$, respectively. Under some regularity conditions, for each $f \in \mathcal{C}_b^2$,

$$\pi_t(f) = \frac{\rho_t(f)}{\rho_t(1)},$$

where and $\rho_t(\cdot)$ is a measure-valued process that solves the so-called *classical Zakai equation*:

$$\begin{aligned} \rho_t(f) &= \rho_0(f) + \int_0^t \rho_s(\mathcal{L}f - \sum_{i<j} \sum_k f(\lambda_{ij,k} - 1)) ds \\ &\quad + \sum_{i<j} \sum_k \int_0^t \rho_{s-}(f \cdot (\lambda_{ij,k} - 1)) d(N_{ij,k}(t) - t). \end{aligned}$$

Then, the method in [Scott and Zeng \(2008\)](#) uses a Bayesian filtering approach. Also, one may find an alternative approach in [Frey et al. \(2011\)](#), which is based on a finite element solution to a univariate classical Zakai equation. Also, we can also appeal to a suboptimal filtering procedure studied in [Snyder \(1975\)](#).

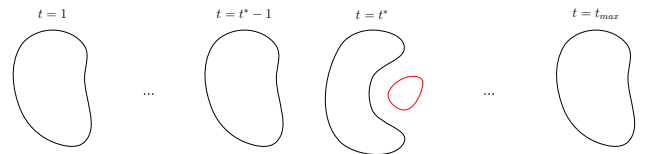


Figure 3: Notational depiction of the change-point in a time-series of graphs. Given a time-series of graphs, interested is the problem of testing, for a time $t \in \{1, 2, \dots\}$, the hypotheses that t is the change-point.

B. Stress graphs

We are interested in a (random) graph $G_T^{(k)}$ that could be used to access the goodness of fit of our estimated π_t . More to the point, considering the mapping $T \rightarrow G_T^{(k)}$ as a stochastic process, we would like to define $G_T^{(k)}$ so that one can find an interesting graph-theoretic pattern in $G_T^{(k)}$ as anomalous messaging activities are emerging in our network.

A notation of the change-point in a time-series of graphs is illustrated in Figure 3.

For such exploitation tasks, we propose to use graph-statistics based on “stress graphs”, which are constructed using the random time-change mapping:

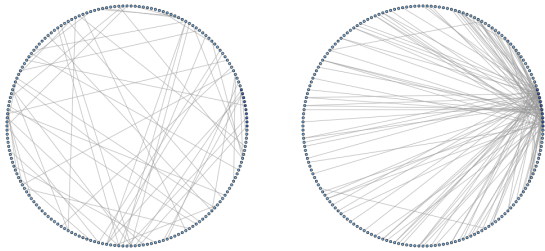
$$t \rightarrow \int_0^t \hat{\lambda}_{ij,k}(s) ds. \quad (6)$$

The theory of point processes (Daley and Vere-Jones, 2008, Theorem 7.4.V) shows that the sequences:

$$\{\xi_{ij,k}(\ell) : i \neq j \in \mathbb{V} \text{ and } 1 \leq \ell \leq N_{ij,k}(T)\}$$

form a collection of independent exponential random variables with a common mean one, where for each $\ell = 1, \dots, N_{ij,k}(T)$,

$$\xi_{ij,k}(\ell) = \int_{\tau_{ij,k}(\ell-1)}^{\tau_{ij,k}(\ell)} \hat{\lambda}_{ij,k}(s) ds. \quad (7)$$



(a) Erdos-Renyi, signaling absence of misfit. (b) Non Erdos-Renyi patterns, signaling a presence of misfit.

Figure 4: Stress graph. Each gray edge represents an event that the test statistic for the pair has the p -value lower than the pre-specified level of significance α_0 . On the right, the ten vertices with the most edges are also the vertices whose deviation from the normal behavior happens to be the largest.

Now, we outline a procedure for constructing the (random) graph $G_T^{(k)}$ where each f can be any Borel measurable function on \mathbb{R} . To simplify our notation, we will suppress the dependence of $G_T^{(k)}$ on f unless needed for clarification.

First, for each ij and k , we let $Q_{ij,k}(T)$ be a statistic based on

$$\{\xi_{ij,k}(\ell) : \ell = 1, \dots, N_{ij,k}(T)\} \quad (8)$$

where as T gets large, the asymptotic (null) distribution of $Q_{ij,k}(T)$ can be computed or estimated. For example, $Q_{ij,k}(T)$ can be defined to be:

$$Q_{ij,k}(T) = \left(\frac{\sum_{\ell=1}^{N_{ij,k}(T)} (\xi_{ij,k}(\ell) - 1)}{\sqrt{N_{ij,k}(T)}} \right), \quad (9)$$

or alternatively, one can take $Q_{ij,k}(T)$ be the Kolmogorov-Smirnov statistics. For the rest of our discussion in this

section, we will fix $Q_{ij,k}(T)$ to be of the form given by (9). In this case, the asymptotic null distribution is simply the standard normal distribution.

Up to this point, we need not specify the function f . Now fix a Borel measurable function f on \mathbb{R} . Then, let $G_T^{(k)}(f)$ be the weighted graph such that for each pair of vertices i and j , the weight of the edge between i and j is $f(Q_{ij,k}(T))$. For the statistic in (9), we can take, for a fixed $\alpha_0 \in (0, 1)$,

$$f_0(x) = \mathbf{1}\{\Phi(x) > 1 - \alpha_0\},$$

where $\mathbf{1}\{\cdot\}$ denotes the indicator function and Φ denotes the cumulative distribution of a standard normal random variable. Note that when T is sufficiently large, one has that $\Phi(Q_{ij,k}(T))$ is (approximately) uniformly distributed on $(0, 1)$, and hence, $f_0(Q_{ij,k}(T))$ is a Bernoulli random variable with its success probability α_0 . Some notional examples of such a graph are illustrated in Figure 4.

Finally, the graphs $G_T^{(1)}(f_0), \dots, G_T^{(K)}(f_0)$ form a collection of independent Erdos-Reny random graphs. This is a fact that can be useful in developing a methodology for change point detection as one can search for a strong signal for deviation from $G_T^{(k)}(f_0)$ being an Erdos-Renyi graph.

V. CONCLUSION AND FUTURE WORK

The point process approach has the potential to yield many benefits. In this paper, we have explored the benefits of using a system of dependent point processes to generate networks and graphs of interest.

By considering the network as evolving in continuous time, the noise induced by constructing graphs from temporally aggregated can be reduced. In addition, the effect of temporal resolutions and summary schemes on the derived graph properties can be further understood. The benefits of a point process approach also carry over to anomaly, outlier, and change detection where better detection capabilities can be achieved.

As the relationship between entities will drive the structure of the resulting point processes, one can consider investigating how to construct, estimate, and make inferences on the mutual dependence structure between entities based on various histories of messaging activity. The concepts of point process thinning as well as rescaling can be used for goodness-of-fit testing for model validation.

Another component of our approach in this paper is to model the network dynamics via a set of low-dimensional latent processes based on the point process intensities. In addition, the creation of a latent graph space allows communities to be discovered and tracked easier by significantly reducing the dimensionality of the problem. In particular, we have

considered an embedding algorithm based on the messaging intensities that can scale over time and network size. By analyzing these embeddings as they change, we might discover communities that share a similar spatial trajectory over time.

Finally, for an exact model like in this paper, it is often useful to derive an approximate model allowing one to conduct more detailed mathematical analysis. For a time series of random graphs, this approach has been explored in Lee and Priebe (2011), and subsequently in Priebe et al. (2011), Tang et al. (2011). Following this strategy, another fruitful approach is to consider the asymptotic distribution of the following scaled processes when $r \rightarrow \infty$:

$$\widehat{N}_{ij,k}^r(t) = \frac{N_{ij,k}(rt) - C_{ij,k}(rt)}{\sqrt{r}}, \quad (10)$$

where

$$C_{ij,k}(s) = \int_0^s \widehat{\lambda}_{ij,k}(u) du.$$

Under some regularity conditions, the functional invariance principle (Anderson et al. 1995, Theorem II.5.4) suggests that as $r \rightarrow \infty$, the scaled process \widehat{N}^r converges in distribution to a (multivariate) Brownian motion.

REFERENCES

- P. K. Anderson, O. Borgan, R. Gill, and N. Keiding. *Statistical models based on counting processes*. Springer, 1995. 6
- D. J. Daley and D. Vere-Jones. *An introduction to the theory of point processes: Volume 1*. pringer-Verlag, 2008. 5
- M. De Choudhury, W. Mason, J. Hofman, and D Watts. Inferring relevant social networks from interpersonal communication. In *In Proc. 19th Intl Conf. World Wide Web, New York*, pages 301–310. Association for Computing Machinery, 2010. 1
- R. Frey, T. Schmidt, and L. Xu. On Galerkin approximations for the Zakai equation with diffusive and point process observations. *Working Paper*, 2011. 4
- A. Goldenberg, A. Zheng, S. Fienberg, and E. Airoldi. A survey of statistical network models. *Found Trends Mach Learn*, 2010. 1
- N. Heard, D. Weston, K. Platanioti, and D. Hand. Bayesian anomaly detection methods for social networks. *Ann. Appl. Statist.*, 4:645–662, 2010. 1
- E. Kolaczyk. *Statistical Analysis of Network Data*. Springer, New York, 2009. 1
- N. H. Lee and C. E. Priebe. A latent process model for time series of attributed random graphs. *Statistical Inference for Stochastic Processes*, 14(3):231–253, October 2011. 6
- P. Perry and P. 2010 Wolfe. Point process modeling for directed interaction networks. *Submitted for publication*, 2010. 1
- 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 Statistical Signal Processing Workshop*, July 2011. 6
- Laurie C. Scott and Yong Zeng. A class of multivariate micro-movement models of asset price and their Bayesian model selectin via filtering. In *Markov processes and related topics: a Festschrift for Thomas G. Kurtz*, volume 4, pages 123–136. Institute of Mathematical Statistics, 2008. 4
- D. Snyder. *Random point processes*. John Wiley & Sons Inc, 1975. 4
- M. Tang, Y. Park, N. H. Lee, and C. E. Priebe. Attribute fusion in a latent process model for time series of graphs. Submitted, 2011. 6