Proceedings of the Ninth International AAAI Conference on Web and Social Media

Modeling Response Time In Digital Human Communication

Nicholas Navaroli*
Department of Computer Science
University of California, Irvine
nnavarol@uci.edu

Padhraic Smyth
Department of Computer Science
University of California, Irvine
smyth@ics.uci.edu

Abstract

Our daily lives increasingly involve interactions with other individuals via different communication channels, such as email, text messaging, and social media. In this paper we focus on the problem of modeling and predicting how long it takes an individual to respond to an incoming communication event, such as receiving an email or a text. In particular, we explore the effect on response times of an individual's temporal pattern of activity, such as circadian and weekly patterns which are typically present in individual data. A probabilistic time-warping approach is used, considering linear time to be a transformation of "effective time," where the transformation is a function of an individual's activity rate. We apply this transformation of time to two different types of temporal event models, the first for modeling response times directly, and the second for modeling event times via a Hawkes process. We apply our approach to two different sets of real-world email histories. The experimental results clearly indicate that the transformation-based approach produces systematically better models and predictions compared to simpler methods that ignore circadian and weekly patterns.

Current technology allows us to collect large quantities of time-stamped individual-level event data characterizing our "digital behavior" in contexts such as texting, email activity, microblogging, social media interactions, and more — and the volume and variety of this type of data is continually increasing. The resulting time-series of events are rich in behavioral information about our daily lives. Tools for obtaining and visualizing such information are becoming increasingly popular, such as the ability to download your entire email history for mail applications such as Gmail, and various software packages for tracking personal fitness using data from devices such as Fitbit.

This paper is focused on modeling the temporal aspects of how an individual (also referred to as the "ego") responds to others, given a sequence of timestamped events (e.g. communication messages via email or text). What can we learn from the way we respond to others? Are there systematic patterns that can be extracted and used predictively? How do our daily sleep and work patterns factor in to our response patterns? These are common issues that arise when modeling communication events, as the ego's behavior typically changes significantly over the course of a day or week. Examples of such patterns are shown in Figure 1.

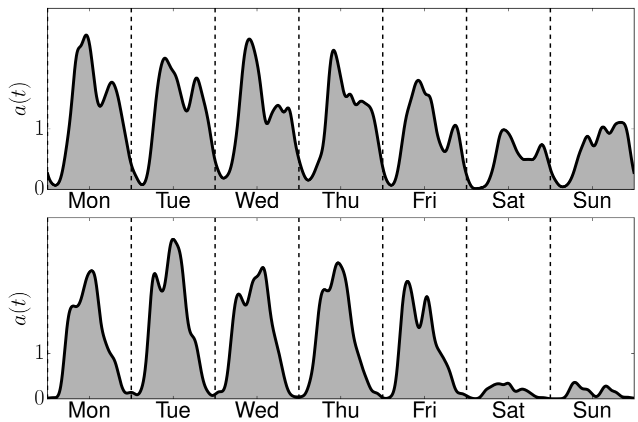

Figure 1: Example of smoothed daily and weekly patterns over two separate individuals' email histories. $a(t)$ on the y-axis represents an individual's typical activity as a function of time.

We propose a novel method for parameterizing these circadian (daily) and weekly patterns, allowing the time dimension to be transformed according to such patterns. This transformation of time will allow models to describe and predict behavioral patterns which are invariant to the ego's routine patterns. We apply this approach to predicting the ego's response time to an event. Learning such response patterns is useful not only for identifying relationships between pairs of individuals (Halpin and De Boeck 2013), but also as features for prioritization of events, e.g., the priority inbox implemented in Gmail (Aberdeen, Pacovsky, and Slater 2010). Our experimental results show clear advantages in terms of predictive power when modeling response times in the transformed time domain.

*Currently employed at Google. The research described in this paper was conducted while the author was a graduate student at UC Irvine.

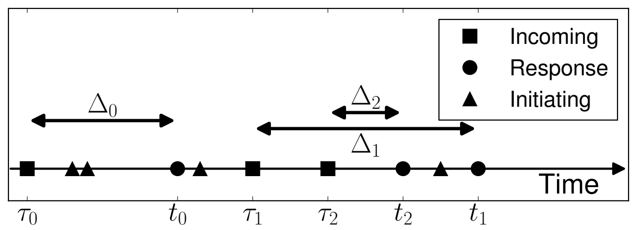

Figure 2: Illustration of communication event data and response times. The x-axis represents a timeline of events for an individual (e.g., for their email inbox).

Problem Definition and Notation

We define an event $e \equiv \{s, \tau, m\}$ to be an observable action performed by an actor $s \in \{1, 2, \cdots, S\}$ with a timestamp $\tau \in \mathbb{R}^+$ denoting the time the event occurred, and metadata m describing the event. As an example, consider a text message being sent at time τ; the actor s would be the message sender, and metadata could include the message's content, recipients, and GPS coordinates.

Here we consider communication events from an egocentric perspective, where all events involve a single individual or ego. For example, the ego may be owner of the phone receiving and sending text messages, or the owner of an email inbox. We distinguish between two different types of events in such data: *incoming* events directed towards the ego, and *response* events where the ego takes some action in response to an incoming event. The distinction between the event types is assumed to be contained in the metadata.

In this paper, we consider communication datasets (e.g. all emails in an inbox) where $\boldsymbol{\tau} = \{\tau_i : 1 \leq i \leq N\}$ denotes the timestamps of all incoming events (e.g. emails the ego receives) and $\boldsymbol{t} = \{t_i > \tau_i : 1 \leq i \leq N\}$ the timestamps of all response events from the ego. An ego's response pattern can be considered to be a mixture of two models; first, a binary model indicating whether the ego responds or not, and second the amount of time the ego takes to respond (conditioned on the fact that the ego responds). The problem of whether or not an ego will respond has been well-studied (Dredze, Blitzer, and Pereira 2005; Aberdeen, Pacovsky, and Slater 2010; Navaroli 2014). In this paper we focus on the second problem, that of modeling the time to respond given that a response occurs.

Let $\Delta_i \equiv t_i - \tau_i$ represent the individual's *response time* to event i. In the event that multiple responses to the same event occur (e.g. the ego sends follow-up replies to an email), Δ_i represents the response time of the *first* response. This paper focuses on probability distributions of the form $p(\Delta_i|\Theta)$, where Θ represents the model's parameters. Figure 2 illustrates what these different quantities may look like. Note that some events are labeled as *initiating* events — these events occur when the ego initiates a conversation (instead of replying). Although we will explore a model which is able to describe initiating events, the focus of this paper is on the timing of response events.

Related Work

When considering the time it takes for an individual to respond to an event, large variations are often seen — responses are known to be "bursty", usually sent quickly or after long periods of inactivity (Barabási 2005; Malmgren et al. 2009). To capture the large variance in response times, models such as the exponential distribution has been proposed (Mahmud, Chen, and Nichols 2013), along with longer-tailed distributions such as the power-law (Eckmann, Moses, and Sergi 2004; Barabási 2005) or lognormal (Stouffer, Malmgren, and Amaral 2006; Kaltenbrunner et al. 2008; Zaman, Fox, and Bradlow 2014). While these long-tailed distributional forms are able to model large variances in response times, they do not take into consideration the typical circadian or weekly behavior of the individual.

A more complex model which can incorporate circadian and weekly patterns is the non-homogeneous Poisson process, which models the rate at which responses occur (as opposed to the time difference between the incoming event and response). Periodic response patterns can be modeled by assuming the response rate $\lambda(t)$ is piecewise-constant and decomposed into a product of terms based on differently-scaled time intervals (e.g. daily, hourly). This decomposition of the response rate has been proven successful in describing circadian patterns of email communication (Malmgren et al. 2009), instant messaging (Pozdnoukhov and Walsh 2010), and telephone calls (Scott 2000).

A related model of communication response rates is the Hawkes process (Hawkes 1971), where the overall rate of a response is the superposition of many independent response rates, one for each event the ego receives. These "self-exciting" models have been effective in capturing the burstiness of communication behavior (Malmgren et al. 2008; Simma and Jordan 2010), however they do not directly model circadian patterns. Hawkes processes have also been used in other contexts of communication behavior, such as inferring latent relationships between individuals in social networks (Blundell, Beck, and Heller 2012; Fox et al. 2013; Zipkin et al. 2014) and modeling dyadic interactions (Halpin and De Boeck 2013; Masuda et al. 2013).

Our proposed approach for modeling circadian rhythms is most similar to the work of Jo et al. (2012), who investigated the rescaling of time via histograms to remove circadian and weekly patterns in exploratory data analysis of individual communication events. Our work is similarly motivated but extends this earlier work in several significant respects. First, we use smooth non-parametric kernels (rather than histograms) to model temporal patterns. Second, we demonstrate how the transformation of time can be effectively embedded in different statistical models such as the Hawkes process. Lastly, we quantify the improvements gained by transforming time with a series of systematic prediction experiments on out-of-sample data.

Temporal Patterns in Communication Data

In communication data, strong circadian and weekly patterns are commonly found in an individual's behavior. For example, in professional email communication, people are

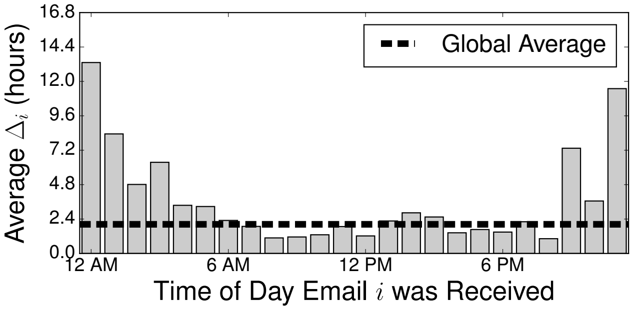

Figure 3: Average response time Δ_i as a function of hour-of-day, for a single individual's email data.

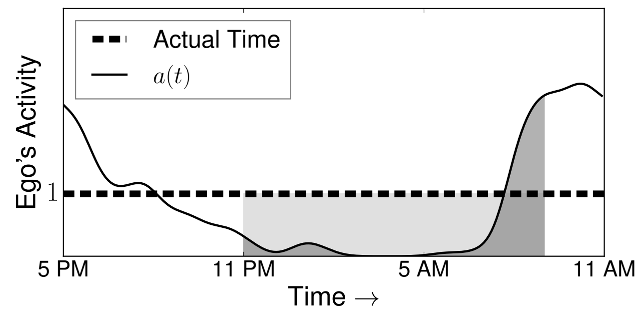

Figure 4: Effective response time (dark shaded region) for an ego responding, at 9 AM, to an incoming event from 11 PM the previous night.

generally more active and likely to write an email during the daytime and on weekdays (e.g. see Figure 1). These strong daily and weekly patterns are present not only for email data, but for many types of communication data, such as mobile text messaging (Jo et al. 2012) and "re-tweets" in Twitter (Zaman, Fox, and Bradlow 2014).

As a result, the ego's response time Δ_i to event i can fluctuate with the time of day or week. Suppose an incoming event occurs at night while the ego is asleep. The response time Δ_i for that event is likely to be large as the individual may not have a chance to be aware of the event (let alone reply to it) until they are awake. In contrast, an event received in the beginning of the ego's daily routine may be responded to after a shorter amount of time. An example of this phenomenon (for one particular individual) is illustrated in Figure 3, where the average response time is clearly dependent on when the email was received.

The circadian rhythms which cause Δ_i to fluctuate over time are not accounted for when modeling Δ_i using standard probability distributions. Thus, parameters inferred as a function over $\{\Delta_i\}$ are likely to misrepresent the individual's dynamic activity. Additionally, circadian rhythms are known to contribute to the large variances and heavy tails in the estimated distributions (Barabási 2005; Fox et al. 2013). Being able to effectively remove such patterns (in distributions over response time) will both 1) allow the distribution to change as a function of time of day and week, and 2) reduce the variance in the resulting distribution.

We approach this problem by using an inhomogeneous model where the distribution over Δ_i depends on and changes with τ_i, the time of the received event. There are many ways to accomplish this; for example, by using a conditional density model allowing the model parameters Θ to be a function of time, or using separate models for different time intervals (e.g. every x hours or days). While such approaches are potentially useful, both have their drawbacks. The temporal dependence of parameters in the conditional density approach may be quite non-linear and difficult to capture, while the binning approach will result in partitioning the data across time intervals, which are likely to be sparse. With this in mind, the temporal transformation approach we discuss in the following sections has the advantage of being straightforward to implement and, as will be shown in the experimental section, very effective.

Modeling Effective Response Times

We address the problems described in the previous section by considering the ego's *effective response time* $\tilde{\Delta}_i$ to event i. By placing a probability distribution $p(\tilde{\Delta}_i|\Theta)$ over effective response time and defining Δ_i to be a function $f(\tilde{\Delta}_i)$ of $\tilde{\Delta}_i$, the following distribution is induced over Δ_i:

$$q(\Delta_i|\tau_i, \Theta) = p\big(\tilde{\Delta}_i\big|\Theta\big)\frac{\partial}{\partial \Delta_i} f^{-1}(\Delta_i) \tag{1}$$

where $\tilde{\Delta}_i = f^{-1}(\Delta_i)$, and $\frac{\partial}{\partial \Delta_i} f^{-1}(\Delta_i)$ the Jacobian of the inverse transformation. In principle we could use any transformation here, but a natural choice is to transform with respect to the activity of the user so that time is dilated at times of high activity and contracted at times of low activity (Jo et al. 2012). Thus, we define the transformation $\Delta_i = f(\tilde{\Delta}_i)$ such that

$$\tilde{\Delta}_i = f^{-1}(\Delta_i|\tau) \equiv \int_{\tau_i}^{\tau_i+\Delta_i} a(u)du \tag{2}$$

where τ_i is the incoming event's timestamp, and $a : \mathbb{R}^+ \to \mathbb{R}^+$ a positive function, referred to as the *activity function*. Note that this transformation from $\tilde{\Delta}_i$ to Δ_i is dependent on τ_i. When $a(u) = 1$ for all u (i.e., a user's activity is constant over time), $\tilde{\Delta}_i = \Delta_i$ and $q(\Delta_i|\tau_i, \Theta) = p(\Delta_i|\Theta)$.

The value of the activity function $a(t)$ at any point in time can be interpreted as a relative "rate of activity" for the individual. We illustrate this interpretation with an example in Figure 4. Suppose an incoming event occurs at $\tau_i = 11$ PM and the ego responds to it at $t_i = 9$ AM the following day. The actual response time is $\Delta_i = 10$ hours (the light shaded area). With $a(t)$ defined as the solid curve in Figure 4, the effective response time $\tilde{\Delta}_i = \int_{11\text{ PM}}^{9\text{ AM}} a(u)du \approx 4$ hours, much less than the actual response time. Thus, values of $a(u) < 1$ throughout this time interval suggests that the ego's activity is lowered (e.g. asleep during the night). Similarly, values of $a(u) > 1$ represent times where the ego is more active than average (e.g. throughout the ego's work day). Figure 1 showed examples of what $a(t)$ may look like over a week.

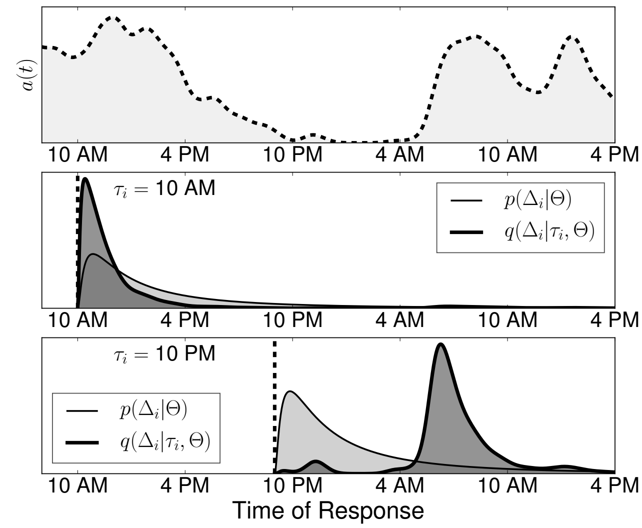

Figure 5: Example of a probability distribution $q(\Delta_i|\tau_i,\Theta)$ induced by a distribution over effective response time $\tilde{\Delta}_i$. Unlike $p(\Delta_i|\Theta)$, the shape of $q(\Delta_i|\tau,\Theta)$ depends on τ_i.

As an example of what the transformed distribution $q(\Delta_i|\tau_i,\Theta)$ defined by Equation 1 may look like, suppose the distribution p is lognormal with a mean time to respond of 4 hours and variance 3 hours. Figure 5 shows what $q(\Delta_i|\tau_i,\Theta)$ (dark solid curve) would look like, compared to a lognormal distribution over actual response time $p(\Delta_i|\Theta)$ (light solid curve), for responding to events received at $\tau_i =$ 10 AM (middle plot) and $\tau_i = 10$ PM (bottom plot).

In this example, we assume the activity function used to transform time is defined by the top plot in Figure 5. The lognormal density $p(\Delta_i|\Theta)$ is invariant to τ_i, thus has the same shape for both events. While this distribution may be sensible for the event received at $\tau_i = 10$ AM, it does not accurately describe the individual's behavior in responding to the event received at $\tau_i = 10$ PM. In this case, much of the probability mass is "wasted" in the late hours of the evening, where the individual's activity is low.

In contrast, the shape of the distribution $q(\Delta_i|\tau_i,\Theta)$ in Figure 5 significantly changes between the two events. For the event received at $\tau_i = 10$ PM, most of the probability mass is shifted away from the night and to the early hours of the morning where the ego becomes active. For the event received at $\tau_i = 10$ AM, the increased activity function (indicating the ego is active) causes $q(\Delta_i|\tau_i,\Theta)$ to become more peaked than $p(\Delta_i|\Theta)$, thus the event is likely to be responded to more quickly. As these examples show, the specification of the activity function $a(t)$ allows the distribution $q(\Delta_i|\tau_i,\Theta)$ over response time to "warp" significantly in shape as a function of τ_i, minimizing the effects of the individual's circadian and weekly rhythms.

In the following sections, we describe how to apply the time transformation using $\tilde{\Delta}_i$ to two different models of response behavior. The first directly models the response time Δ_i, while the second models the timestamp $t_i > \tau_i$ as a stochastic process. We then describe how to estimate the activity function $a(t)$ given historical response data.

Note that our overall approach can be viewed from two equivalent perspectives: (1) a "transformed time" approach where we first transform the time dimension with respect to $a(t)$, then use a simple distributional form (such as a lognormal) to model $\tilde{\Delta}_i|\Theta$, and (2) a relatively complex model over actual response time $q(\Delta_i|\tau_i,\Theta)$ induced by the simpler model $\tilde{\Delta}_i|\Theta$ in the transformed time domain. For example, while $p(\Delta_i|\Theta)$ in Figure 5 is unimodal (e.g. lognormal), $q(\Delta_i|\tau_i,\Theta)$ in actual time can be multimodal, allowing the model to capture complex aspects of the ego's behavior relative to the activity function $a(t)$.

Direct Modeling of Response Times

In this section, we show how to use the transformation of time to directly model the ego's response time Δ_i to event i.

Under this model, Δ_i is assumed to be generated from $q(\Delta_i|\tau_i,\Theta)$ induced by the distribution over effective response time $p(\tilde{\Delta}_i|\Theta)$. Any distribution can be used to model $p(\tilde{\Delta}_i|\Theta)$; common examples used for response times include the exponential (Malmgren et al. 2008; Fox et al. 2013), Gamma (Halpin and De Boeck 2013), and lognormal (Stouffer, Malmgren, and Amaral 2006; Zaman, Fox, and Bradlow 2014) distributions. The log-likelihood of all response times $\{\Delta_i\}$ (assuming response times are conditionally independent given the model) is

$$\log p(\{\Delta_i\}|\{\tau_i\},\Theta) = \sum_{i=1}^{N} \log q(\Delta_i|\tau_i,\Theta) \qquad (3)$$

Using the form of $q(\Delta_i|\tau_i,\Theta)$ and the transformation using $\tilde{\Delta}_i$ defined by Equations 1 and 2 respectively, the log-likelihood is equal to

$$\sum_{i=1}^{N} \log p(\tilde{\Delta}_i|\Theta) + \sum_{i=1}^{N} \log a(\tau_i + \Delta_i) \qquad (4)$$

where the first term is the log-likelihood over effective response times $\{\tilde{\Delta}_i\}$, and the second term the sum of log-activity rates over the timestamps of all the ego's responses. Note that, conditioned on the activity function $a(t)$, the second term is constant. Thus, estimating the model parameters Θ can be done using traditional methods (e.g. maximum likelihood) over effective response times $\{\tilde{\Delta}_i\}$.

This decoupling of the log-likelihood in Equation 4 emphasizes the modularity of our time transformation approach. For example, suppose $p(\tilde{\Delta}_i|\Theta)$ is modeled as an exponential distribution, where λ is the exponential mean (in units of days). Conditioned on $a(t)$, a maximum-likelihood estimate of λ can be obtained by computing $\frac{1}{N}\sum_{i=1}^{N}\tilde{\Delta}_i$, the mean effective response time. Given any (potentially complex) distributional form p, the introduction of and conditioning on $a(t)$ does not increase the complexity of parameter inference — the only added complexity to the model is first estimating $a(t)$, which is discussed later.

As a side note, the process of applying the time transformation to an exponential distribution of response time has a strong connection to survival modeling. In survival models, the response time Δ_i is modeled with a survival function

$$S(x) \equiv P(\Delta_i > x) = \text{Exp}\left(-\int_0^x h(u)du\right)$$

where $h(u)$ is referred to as a *hazard function* (Aalen, Borgan, and Gjessing 2008). One possibility is to set $h(u) = \lambda a(u)$ for $\lambda > 0$. Under this parameterization, probabilities under the survival model become equivalent to an exponential distribution with parameter λ warped according to the activity function. This can be shown using the cumulative density function $Q(\Delta_i|\tau_i, \lambda)$ of an exponential distribution transformed according to Equation 1 (assuming the incoming event was received at time $\tau_i = 0$):

$$\begin{aligned} q(\Delta_i > x|\tau_i = 0, \lambda) &= 1 - Q(x|\tau_i = 0, \lambda) \\ &= \text{Exp}\left(-\lambda\int_0^x a(u)du\right) = S(x) \end{aligned}$$

The relationship between the two models is useful in that it allows the activity function $a(t)$ to further be interpreted as a rate at which an event (e.g. a response) will occur. Additionally, it shows that the proposed model can be viewed as a variant of survival modeling where alternative distributional forms other an exponential can be used.

Stochastic Process Models of Response Times

To illustrate the modularity of the time transformation approach, we now show how to embed it within a stochastic process model. Stochastic processes do not model response times, but rather the timestamps at which the ego sends a communication event (e.g. the timing of the response and initiating events from Figure 2). Additionally, repeated responses to the same event could potentially be modeled — for more details, see Navaroli (2014).

In this paper, we focus on Hawkes processes (Hawkes 1971), which have successfully been used to model email communication patterns (Simma and Jordan 2010; Fox et al. 2013; Halpin and De Boeck 2013). In a Hawkes process, the ego's rate $\lambda(t)$ of sending a message at time t is

$$\lambda(t) \equiv \lambda_0(t) + \sum_{i:\tau_i<t} g(t|\tau_i) \tag{5}$$

where $\lambda_0(t)$ is the rate at which the ego initiates events (e.g. sending an email that is not a reply; the initiating events in Figure 2) and $g(t|\tau_i)$ are the *triggering functions* describing the rate at which the ego responds to event i. Typically, $g(t|\tau_i) = \nu p(\Delta_i|\Theta)$, where $p(\Delta_i|\Theta)$ is defined as before (e.g. exponential, Gamma) and ν is interpreted as the expected number of replies to a single event (Fox et al. 2013).

The Hawkes process is useful in that the ego's response rate is a function of $\{\tau_i\}$ — as incoming events are received, the ego's rate of reply increases. However, the triggering function $g(t|\tau_i)$ is a function of the actual response time Δ_i and is prone to the same problems caused by circadian and weekly patterns when modeling $p(\Delta_i|\Theta)$. We can instead model the rate of a response using the transformed time domain, where $g(t|\tau_i) = \nu q(\Delta_i|\tau_i, \Theta)$ and $q(\Delta_i|\tau_i, \Theta)$ defined as in Equation 1.

The log-likelihood of response times according to the Hawkes process is thus,

$$\log p(\{t_i\}|\{\tau_i\}, \Theta) = \sum_{i=1}^{N} \log \lambda(t_i) - \int_0^T \lambda(u)du \tag{6}$$

where it is assumed that the observed dataset is over the time interval $[0, T]$ (Daley and Vere-Jones 2003).

The log-likelihood contains a log function over summations of terms (with $\lambda(t)$ defined by Equation 5), which can make parameter inference intractable. If the response structure (e.g. which incoming event the ego replies to) is unknown, the log-likelihood must marginalize over the response rates corresponding to all incoming events (Halpin and De Boeck 2013; Olson and Carley 2013). However, when the response structure is known (as is the case for the email datasets used in the experiments), the response rate $\lambda(t)$ reduces to

$$\lambda(t) = \begin{cases} g(t|\tau_i) & \text{if the ego replied to event } i \\ \lambda_0(t) & \text{if the ego initiated the event} \end{cases}$$

With this parameterization of $\lambda(t)$, maximum-likelihood estimates of model parameters can be numerically calculated efficiently (no closed form exists due to the integral term in Equation 6). As discussed with the direct model of response time, applying the time transformation approach (via activity function $a(t)$) to the Hawkes process does not increase the complexity of parameter inference.

The modeling choice for $\lambda_0(t)$ is independent of the modeling of response times, as it describes the rate at which the ego initiates new events (e.g. the initiating events from Figure 2). As the activity function $a(t)$ from the previous section can be interpreted as a relative activity rate of the ego, an appropriate modeling choice is $\lambda_0(t) \propto a(t)$, learning the proportionality factor via maximum-likelihood.

Estimation of the Activity Function

In order to apply the time transformation approach to the direct response time model and the Hawkes process, an estimate of $a(t)$ is required. A kernel-based approach is used to estimate $a(t)$ nonparametrically, and can be applied in either a batch or online setting. This allows for efficient updates to the estimation of $a(t)$ in the presence of new activity.

Given the historical set of times of responses $\boldsymbol{t} = \{t_i : 1 \leq i \leq N\}$ by the ego to incoming events (note that $t_i \equiv \tau_i + \Delta_i$), the activity function $a(t)$ is estimated as

$$a(t) \equiv \frac{Z}{W}\sum_{i=1}^{N} w_i \exp\left(-\frac{\Delta_w(t, t_i)^2}{2h^2}\right) \tag{7}$$

where

- $\Delta_w(t_a, t_b) \in [0, 3.5]$ days is the minimum weekly time difference between two timestamps t_a and t_b
- $h > 0$ is referred to as the *bandwidth parameter*
- $w_i > 0$ is the weight or "influence" of t_i
- $W = \sum_{i=1}^{N} w_i$ is the sum of weights

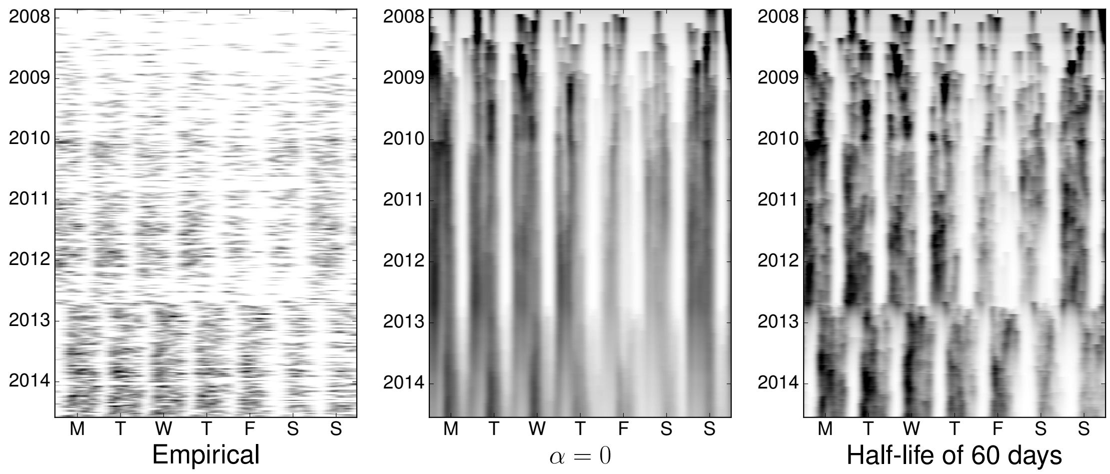

Figure 6: Example where exponential decay of weighting datapoints becomes necessary. Left: smoothed histogram counts over number of emails sent throughout each week; a clear change in behavior occurs in mid-2012. Middle: estimated $a(t)$ over time with $\alpha = 0$ (no decay). Right: estimated $a(t)$ over time with α set such that datapoints lose half their influence after 60 days.

- $Z \equiv 7(2\pi h^2)^{-\frac{1}{2}}$ a normalization constant such that $\int_t^{t+7} a(u)du = 7$, ensuring that model parameters are on the scale of days for interpretability

The structure of this estimate is similar to weighted kernel density estimation; the majority of activity "mass" in $a(t)$ is placed during previously-active times for the ego. Examples of estimated activity functions using Equation 7 for two different egos were shown in Figure 1, using a bandwidth parameter of $h = 90$ minutes. Clear circadian and weekly patterns can be seen for both individuals.

When $w_i = 1$ for all i, each historical timestamp is given an equal weight in the estimation of $a(t)$. This can cause problems when a significant change in the ego's behavior occurs. For example, the ego represented in Figure 6 experiences a significant change in behavior (left plot) in the middle of 2012 — the ego moved from one country to another. Not only did the timezone shift for the ego, but their non-work days shifted from Friday/Saturday to Saturday/Sunday. When $w_i = 1$, the resulting estimate of $a(t)$ averages the patterns from both countries together, resulting in an activity function that is not representative of the individual's actual behavior (middle plot of Figure 6).

This problem can be avoided by allowing the weights of datapoints in the estimation of $a(t)$ to decay over time. Here we use exponential weighting:

$$w_i \equiv \exp\big(-\alpha(t_N - t_i)\big)$$

where $\alpha \geq 0$ is a decay parameter determining the "half-life" of weights, and t_N the most recent time point. When $\alpha = 0$, $w_i = 1$ for all k. By weighting the datapoints the estimated activity function is able to "forget" older events, allowing the estimate to adapt to behavioral changes. The right plot in Figure 6 shows that, by setting $\alpha > 0$, the change in behavior due to the ego's moving between countries is quickly reflected in estimates of $a(t)$ after the changepoint.

Experiments

Email Datasets

We evaluated our time transformation approach using two different email corpora. The first is Gmail data from three volunteers affiliated with our research group. The email timestamps, anonymized email ids, and response structure were downloaded from the email servers using a Python script. The inbox activity for each individual spans several years with thousands of responses, and is shown in Table 1.

Ego	First Email	Last Email	# Responses
Gmail A	Nov 2007	Jul 2014	4268
Gmail B	Nov 2007	Jun 2014	1613
Gmail C	Aug 2006	May 2014	23950

Table 1: Summary statistics of Gmail datasets.

Second, we use data from the publicly-available Enron email corpus (Klimt and Yang 2004). Unlike the Gmail data, the response structure was not available — the structure was inferred based on recipient and normalized subject matching (e.g. removing instances of "fwd" and "re"). Enron employees that sent 200 or more responses were used in these experiments. The number of responses and activity timespan across the resulting 23 employees are shown in Figure 7.

Thus, in total we use email histories from 26 different individuals (3 Gmail, 23 Enron), varying significantly not only in volume and timespan, but also in temporal behavior and the number of email recipients responded to. A more thorough analysis of the datasets is provided in Navaroli (2014). Each email dataset was independently analyzed in the following experiments (e.g. model parameters are not shared across individuals)[1].

[1]Python code available at http://www.datalab.uci.edu/resources and tested using Python 2.7, Numpy 1.8.1, Scipy 0.14.0, and Matplotlib 1.3.1.

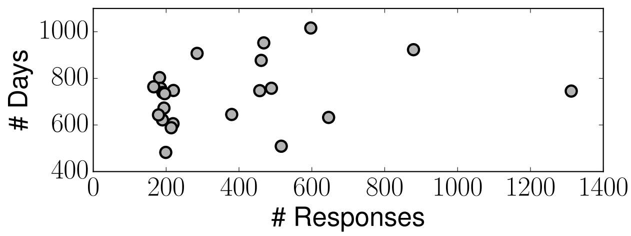

Figure 7: Summary statistics of Enron datasets.

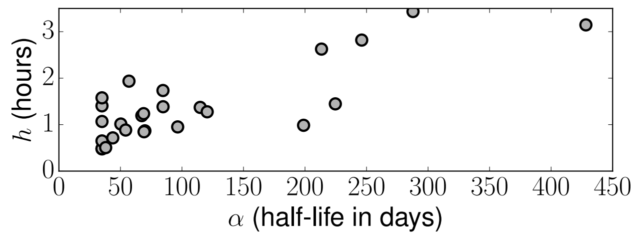

Figure 8: Estimated values of hyperparameters h and α, per email dataset.

Experimental Setup

For each email dataset, the following models are considered:

- Direct models of response times (e.g. Equation 3) using exponential, Gamma, and lognormal distributions,
- Hawkes processes of response rates (e.g. Equation 6) using exponential, Gamma and lognormal triggering functions.

For each model, two separate variants are trained, each using a different estimator of $a(t)$:

1. **Baseline**: $a(t) = 1$, e.g. no transformation of time. This variant does not take into account circadian rhythms when estimating model parameters.
2. **Proposed model**: $a(t)$ is estimated nonparametrically from historical data, allowing for normalization of response times according to circadian rhythms.

All models are trained and evaluated in an online setting where new data is continually being observed. The first 20% (chronologically) of each individual's data is used for initialization, e.g., to estimate the hyperparameters h and α (described in the following section). For the remaining 80% of the data, the model is iteratively trained on data up to and including day $d - 1$ and then probabilities are computed for the (test) response times observed from day d.

As each model moves forward in time, their parameters are updated with each day's worth of new data, resulting in new estimates of model parameters and activity function $a(t)$. We use this sequential approach as it mimics closely how models like this could be used with real event data; similar experimental results were obtained when using more traditional "batch" training/test datasets.

Model evaluation is done using the probability assigned by the model to new data; more specifically the test log-likelihood which is the sum of log probability over test datapoints unseen by the model. The test log-likelihood is widely used for evaluating the quality of probabilistic models (Murphy 2012), where models yielding higher test log-likelihood are better in the sense that they assign higher probabilities to future data that actually occurred.

Priors and Hyperparameters

Appropriate priors are placed over each model parameter (e.g. a Gaussian prior over the lognormal mean, a Gamma prior over the exponential mean), and are set to be relatively weak — only the first few days of processed data are significantly affected by the prior. For parameters with non-conjugate priors (e.g. a Gamma prior over each parameter in the Gamma distribution), maximum a posteriori estimates are obtained numerically.

The hyperparameters h and α used to estimate $a(t)$ are derived from the 20% initial data. 10-fold cross-validation is used to set h, maximizing the cumulative value of $\log a(t)$ over validation points. The value of α is heuristically set such that, on average:

$$W = \sum_{i=1}^{N} w_i \approx \tilde{n}(1 - \exp(-\alpha))^{-1} = 150$$

where $\tilde{n}$ is the average number of emails the ego sends per day, estimated from the initial data. Solving for α in the above equation allows the sum of weights in the estimation of $a(t)$ via Equation 7 at any point in time to be approximately 150 "effective" datapoints. In some cases, this heuristic setting will cause the estimation to "forget" datapoints too rapidly — the minimum value of α considered is one where the half-life a datapoint's influence is 35 days (e.g. $\alpha \geq \frac{\log 2}{35}$). The resulting estimates of h and α for each of the 26 email datasets are shown in Figure 8.

Lastly, the estimation of $a(t)$ is regularized in order to stabilize estimates in the presence of little data:

$$a(t) = \left(\frac{W}{W + \rho}\right)\hat{a}(t) + \left(\frac{\rho}{W + \rho}\right)$$

where $\rho \geq 0$ is a smoothing parameter (interpreted in units of pseudo-datapoints), and $\hat{a}(t)$ is the estimation of $a(t)$ via Equation 7. For these experiments, we set $\rho = 10$ pseudo-datapoints; the experimental results were largely unaffected for values of $5 \leq \rho \leq 20$.

Experimental Results: Log-Likelihood

Figures 9 and 10 compare the mean test log-likelihood between the proposed model where $a(t)$ is estimated nonparametrically and the baseline model where $a(t) = 1$, for the direct and Hawkes process models respectively. In each figure, a single email dataset is represented as three datapoints — one for each of the exponential, Gamma, and lognormal distributions. The x-axis represents the baseline modeling of response time, and the y-axis represents the proposed modeling of response time.

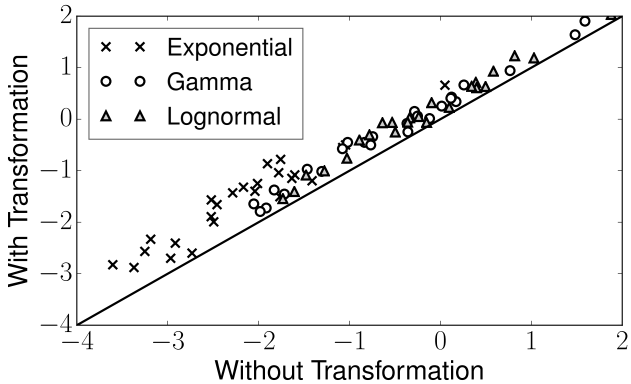

Figure 9: Test mean log-likelihood between direct models modeling $q(\Delta_i|\tau_i, \Theta)$ (y-axis) versus $p(\Delta_i|\Theta)$ (x-axis). Higher values indicate a better fit of the model.

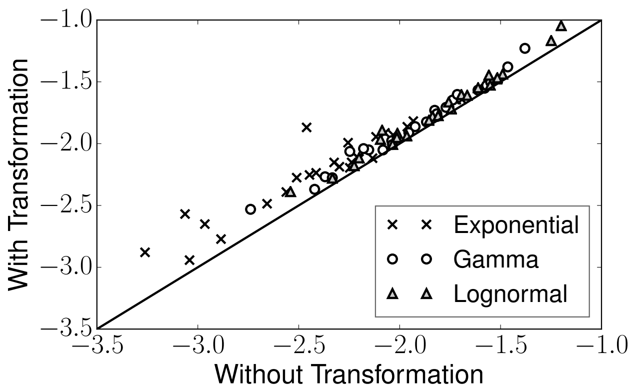

Figure 10: Test mean log-likelihood between Hawkes process models modeling $q(\Delta_i|\tau_i, \Theta)$ (y-axis) versus $p(\Delta_i|\Theta)$ (x-axis). Higher values indicate a better fit of the model.

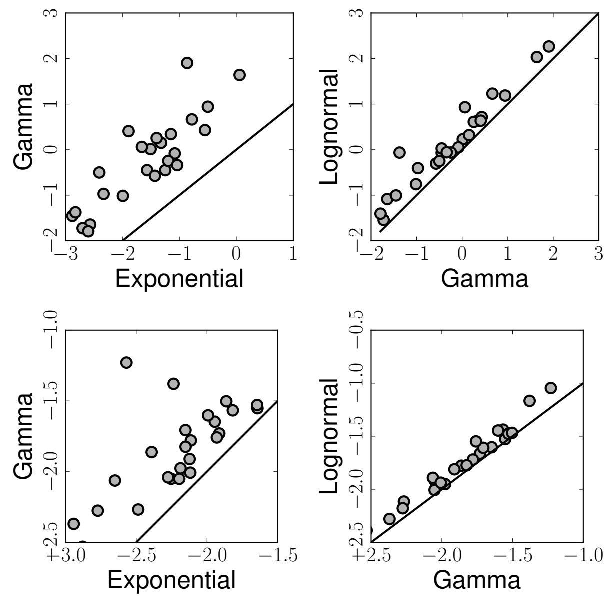

Figure 11: Test mean log-likelihood between different distributional forms of $q(\Delta_i|\tau_i, \Theta)$. Top row: direct response time models. Bottom row: Hawkes process models.

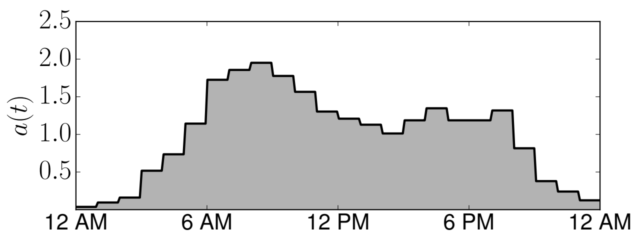

Figure 12: Example of a piecewise-constant estimate of $a(t)$.

A systematic increase in log-likelihood is seen across all datasets and distributional forms when using the proposed model of response times, compared to the baseline model. These results are significant on a 99% confidence interval, using a Wilcoxon signed test. This improvement results from the ability of the proposed model to rescale time, taking into account the typical temporal patterns of each individual. For example, the largest gain from transforming time in event-wise log-likelihood occurs during the morning hours (not shown), where the ego is presumably responding to emails from the previous night; this is precisely the type of scenario portrayed in the bottom plot in Figure 5.

In Figure 11, the log-likelihood of the different distributional forms of $q(\Delta_i|\tau_i, \Theta)$ are compared (similar patterns were found for $p(\Delta_i|\Theta)$), for both direct models of response time (top row) and Hawkes processes (bottom row). Across all email datasets, the log-likelihood using the Gamma distribution is higher than the exponential distribution, with the highest log-likelihood achieved using the lognormal distribution. While this has previously been shown for constant-time models (Stouffer, Malmgren, and Amaral 2006), Figure 11 shows that similar results are obtained 1) for models over transformed response time, and 2) for Hawkes processes.

In addition to the baseline, another model was applied where the activity function $a(t)$ is estimated as a histogram (piecewise-constant) using one-hour time intervals and a periodicity of 24 hours. An example of such an estimate is shown in Figure 12. Note that this parameterization allows the direct response time model using an exponential distribution to be equivalent to a non-homogeneous Poisson process with piecewise-constant response rate, similar to Malmgren et al. (2009). In general, the log-likelihood of this intermediate model was in-between the baseline and proposed models. Results are omitted due to space, but can be found in Navaroli (2014).

Experimental Results: Prediction of Response Within a Time Interval

In our second set of experiments, we explore the proposed model's power in terms of predicting whether or not an incoming event will be responded to within a certain interval of time (e.g. 2 hours), compared to the baseline model of the previous section. Conditioned on the ego responding to an incoming event, the task is to predict whether that event will be responded to sooner instead of later, e.g. for the task of email prioritization (Aberdeen, Pacovsky, and Slater 2010).

Let $Q(\Delta_i|\tau_i, \Theta)$ be the cumulative density function of $q(\Delta_i|\tau_i, \Theta)$. For a given time window ϵ, the probability that an event received at time τ_i will be responded to within ϵ time is

- $p_{\tau_i} = Q(\Delta_i = \epsilon|\tau_i, \Theta)$ for direct response time models,
- $p_{\tau_i} = 1 - \exp(-Q(\Delta_i = \epsilon|\tau_i, \Theta))$ for Hawkes process models (Daley and Vere-Jones 2003).

After calculating p_{τ_i} for each incoming event i in the sequential manner described earlier (for each individual and model), the set of resulting probabilities $\{p_{\tau_i}\}$ are sorted and an area-under-the-curve (AUC) metric is computed relative to ground truth (i.e. whether or not the event was replied to within ϵ time). Larger AUC values indicate that the model is able to better distinguish which incoming events will be responded to within ϵ time. Here results for $\epsilon = 2$ hours are reported — similar results were obtained for other windows of time ranging from 1 to 8 hours.

Figures 13 and 14 compare the AUC between the proposed model where $a(t)$ is estimated nonparametrically and the baseline model where $a(t) = 1$, for the direct and Hawkes process models respectively. Modeling response times using the proposed model results in a significant increase in AUC across the 26 email datasets. This is due to the ability of the proposed model to 1) adapt the predictions p_{τ_i} to the daily and weekly effects parameterized by $a(t)$, and 2) predict that the probability of the ego quickly responding is high (low) when their activity during the time of the received event is high (low) (e.g. see Figure 15).

We also compared to the performance of the intermediate model (where $a(t)$ is estimated using a histogram, e.g. Figure 12), however the results are not shown here. Similar to the log-likelihood results, the AUC of the intermediate model was generally higher than the baseline model, but lower than the proposed model.

Conclusions

In this paper we addressed the problem of modeling the time it takes an individual to respond to incoming communication events. Understanding such response patterns is important in terms of both improving our general understanding of human communication patterns and for designing better tools to assist us in managing our increasingly rich and complex communication channels (e.g., via automated prioritization). The primary novel contribution of our work is the explicit modeling of the effective time it takes an ego to respond based on their typical daily and weekly patterns (as determined by the activity function $a(t)$). We showed the flexibility of the proposed technique by applying it to two different approaches for modeling response times.

Experimental results across 26 different email histories showed a noticeable improvement in predictive performance when introducing the activity function $a(t)$ into both direct and stochastic process models of response times, suggesting

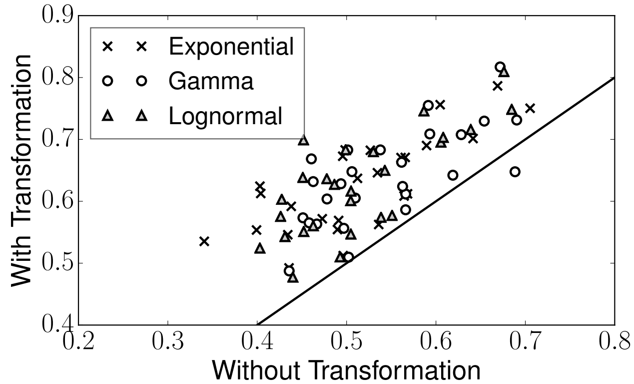

Figure 13: AUC between direct models modeling $q(\Delta_i|\tau_i, \Theta)$ (y-axis) versus $p(\Delta_i|\Theta)$ (x-axis). Higher values indicate more accurate predictions.

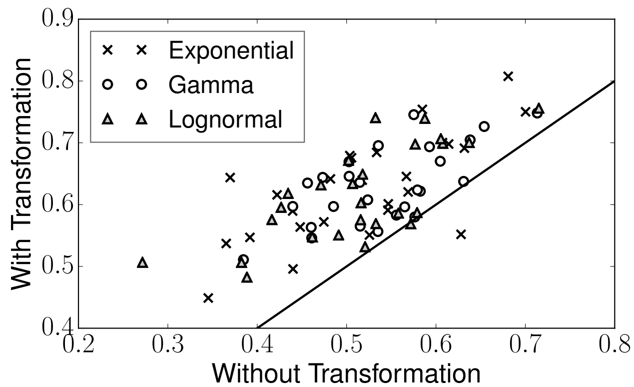

Figure 14: AUC between Hawkes process models modeling $q(\Delta_i|\tau_i, \Theta)$ (y-axis) versus $p(\Delta_i|\Theta)$ (x-axis). Higher values indicate more accurate predictions.

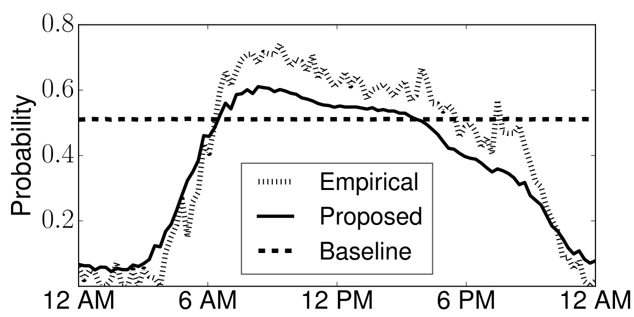

Figure 15: Probability of, for one individual, responding to an email within $\epsilon = 2$ hours according to 1) the empirical data, 2) the proposed model, and 3) the baseline model.

that the models were able to adapt to the strong circadian and weekly patterns experienced by the ego. A potential direction of further study would be to incorporate additional information into the distribution over response time — for example, how long the ego takes to respond may depend on who sent the incoming event. However, we found in preliminary experiments (not shown) that including such information did not improve predictions; this is likely to be due to data sparsity at the individual sender level.

While here we only explored the time transformation approach for direct and Hawkes process models, the approach is both modular and general and can be straightforwardly applied to other event response models, such as different forms of Poisson rate models.

Acknowledgments

This work was supported in part by an NDSEG fellowship for NN, a Google faculty award for PS, and NSF award IIS-1320527. Additionally, the authors thank Moshe Lichman for his participation in discussions and his helpful comments and suggestions.

References

Aalen, O.; Borgan, O.; and Gjessing, H. 2008. *Survival and Event History Analysis: A Process Point of View*. Springer.

Aberdeen, D.; Pacovsky, O.; and Slater, A. 2010. The learning behind Gmail priority inbox. In *NIPS 2010 Workshop on Learning on Cores, Clusters and Clouds*.

Barabási, A. 2005. The origin of bursts and heavy tails in human dynamics. *Nature* 435(7039):207–211.

Blundell, C.; Beck, J.; and Heller, K. 2012. Modelling reciprocating relationships with Hawkes processes. In *Advances in Neural Information Processing Systems 25*. Curran Associates, Inc. 2600–2608.

Daley, D., and Vere-Jones, D. 2003. *An Introduction to the Theory of Point Processes. Vol. I.* Springer-Verlag.

Dredze, M.; Blitzer, J.; and Pereira, F. 2005. Reply expectation prediction for email management. In *2nd Conference on Email and Anti-Spam*.

Eckmann, J.; Moses, E.; and Sergi, D. 2004. Entropy of dialogues creates coherent structures in e-mail traffic. *Proceedings of the National Academy of Sciences of the United States of America* 101(40):14333–14337.

Fox, E.; Short, M.; Schoenberg, F.; Coronges, K.; and Bertozzi, A. 2013. Modeling e-mail networks and inferring leadership using self-exciting point processes. Preprint.

Halpin, P., and De Boeck, P. 2013. Modelling dyadic interaction with Hawkes processes. *Psychometrika* 78(4):793–814.

Hawkes, A. 1971. Spectra of some self-exciting and mutually exciting point processes. *Biometrika* 58(1):83–90.

Jo, H.; Karsai, M.; Kertész, J.; and Kaski, K. 2012. Circadian pattern and burstiness in mobile phone communication. *New Journal of Physics* 14(1):13055–13071.

Kaltenbrunner, A.; Gómez, V.; Moghnieh, A.; Meza, R.; Blat, J.; and López, V. 2008. Homogeneous temporal activity patterns in a large online communication space. *IADIS International Journal on WWW/INTERNET* 6(1):61–76.

Klimt, B., and Yang, Y. 2004. The Enron corpus: a new dataset for email classification research. In *Proceedings of the European Conference on Machine Learning*, 217–226. Springer.

Mahmud, J.; Chen, J.; and Nichols, J. 2013. When will you answer this? Estimating response time in Twitter. In *Proceedings of the Seventh International Conference on Weblogs and Social Media*. AAAI Press.

Malmgren, R.; Stouffer, D.; Motter, A.; and Amaral, L. 2008. A Poissonian explanation for heavy tails in e-mail communication. *Proceedings of the National Academy of Sciences of the United States of America* 105(47):18153–18158.

Malmgren, R.; Hofman, J.; Amaral, L.; and Watts, D. 2009. Characterizing individual communication patterns. In *Proceedings of the 15th ACM SIGKDD International Conference on Knowledge Discovery and Data Mining*, 607–616. ACM.

Masuda, N.; Takaguchi, T.; Sato, N.; and Yano, K. 2013. Self-exciting point process modeling of conversation event sequences. In *Temporal Networks*, Understanding Complex Systems. Springer Berlin Heidelberg. 245–264.

Murphy, K. 2012. *Machine Learning: A Probabilistic Perspective*. MIT Press.

Navaroli, N. 2014. *Generative Probabilistic Models for Analysis of Communication Event Data with Applications to Email Behavior*. Ph.D. Dissertation, University of California, Irvine.

Olson, J., and Carley, K. 2013. Exact and approximate EM estimation of mutually exciting Hawkes processes. *Statistical Inference for Stochastic Processes* 16(1):63–80.

Pozdnoukhov, A., and Walsh, F. 2010. Exploratory novelty identification in human activity data streams. In *Proceedings of the ACM SIGSPATIAL International Workshop on GeoStreaming*, 59–62. ACM.

Scott, S. 2000. Detecting network intrusion using a Markov modulated nonhomogeneous Poisson process. *Journal of the American Statistical Association*. Submitted.

Simma, A., and Jordan, M. 2010. Modeling events with cascades of Poisson processes. In *Proceedings of the 26th Conference on Uncertainty in Artificial Intelligence*, 546–555. AUAI Press.

Stouffer, D.; Malmgren, R.; and Amaral, L. 2006. Lognormal statistics in e-mail communication patterns. *arXiv preprint physics/0605027*.

Zaman, T.; Fox, E.; and Bradlow, E. 2014. A Bayesian approach for predicting the popularity of tweets. *Annals of Applied Statistics* 8(3):1583–1611.

Zipkin, J.; Schoenberg, F.; Coronges, K.; and Bertozzi, A. 2014. Point-process models of social network interactions: parameter estimation and missing data recovery. *European Journal of Applied Mathematics*. Submitted.