Strategic Timing of Content in Online Social Networks

Sina Modaresi
Department of Industrial Engineering, University of Pittsburgh, Pittsburgh PA 15261, sim23@pitt.edu

Juan Pablo Vielma
Sloan School of Management, Massachusetts Institute of Technology, Cambridge MA 02139, jvielma@mit.edu

Tauhid Zaman
Sloan School of Management, Massachusetts Institute of Technology, Cambridge MA 02139, zlisto@mit.edu

Today, having a strong social media presence is an important issue for large and small companies. A key social media challenge faced by these companies' marketing teams is how to maximize the impressions or views of the content they post in a social network. Optimizing the *posting time* of content to increase impressions is an approach not considered before because it was not clear how to systematically select the optimal posting time and what would be the potential gain in impressions. In this work we show how to select posting times to maximize impressions and the potential gains of this strategic timing. We use data from several Twitter users to build a model for how users view content in a social network. With this model we are able to provide a simple equation that gives the impression probability for a piece of content as a function of its posting time. We show that for several real social media users strategic timing can significantly increase impressions. Furthermore, this increase in impressions comes at no cost because choosing the time to post is free. In addition, all calculations use publicly available data, so this approach can be implemented very easily. Finally, we consider the situation where strategic timing becomes widely adopted and posting times are scheduled by an online application. This situation leads to potentially intractable optimization problems and a natural trade-off between aggregate performance and fairness objectives. However, we show that surprisingly, increasing fairness actually seems to improve aggregate performance in this setting. In addition, we show that solutions that are nearly optimal for both objectives can be easily constructed.

Key words: marketing, advertising and media, social media, social networks, statistics, optimization

1. Introduction

Having a strong social media presence is becoming more and more important for a wide range of companies (Pozin 2010). Modern social networks provide the opportunity for companies to easily reach a massive audience. The social media strategy of a company involves designing and posting content to grow their customer base and engage with existing customers. The marketing team of a company will typically be responsible for posting important information on social networks, such as new product releases or promotional material, with the hope of reaching as large an audience as possible. Social networks have an incredibly large potential audience for these posts: the social network Twitter has over 302 million users (Twitter 2015) and the

image based social network Instagram has 300 million users (Instagram 2015). While this potentially large audience is attractive, the challenge is that today the number of competing posts from other social network users is huge, with over 500 million tweets posted per day on Twitter (Twitter 2015) and 70 million photos and videos posted per day on Instagram (Instagram 2015). This large volume of posts means that it can be difficult for a given post to reach the intended audience.

Social network users who post content wish to have their audience *engage* with their content. This is because engagement is an active interaction with the post and represents a better reflection of users who liked the post versus users who saw the post and did not care for it. Engagement comes in different varieties depending on the specific social network. On Twitter engagement includes *favoriting* a tweet or *retweeting* a tweet (forwarding a tweet to others). On Instagram engagement is done by *liking* a post. All these different forms of engagement involve clicking on the post and creating an easily measured signal. Therefore, engagement allows one to estimate how many people actually viewed the post and also deemed it interesting enough to interact with it in some manner.

Engagement can only occur if a user actually sees the content, which is referred to as an *impression*. Therefore, increasing the number of impressions can potentially increase the overall engagement. Many factors determine the number of impressions content receives in social networks. One factor impacting the number of impressions is how engaged users are on the social network. If no one ever checks for content on the social network, then there will be no impressions. If the overall quality of content improves on the social network, this can make users check it more frequently. However, this factor cannot be affected by an individual user trying to maximize his own impressions. Another factor impacting the number of impressions is follower count. Having a larger number of followers means there is a larger potential audience for the posts. Therefore, increasing follower count is one way to gain more impressions.

There is another important, but very often overlooked factor impacting the impression count that has not received much attention: *the timing of the post*. A very natural way to generate impressions is to post content when there are many users checking the social network. If there is a large audience at a certain time, then that is a "good" time to post. However, in a social network, one's post must compete with other users' posts for the attention of followers. Therefore, it is logical to also avoid times when there are many others posting. Combining these two effects, there is a clear trade-off between the number of users checking the social network and the number of users competing to attract their attention. Therefore one can see that the best time to post is when a large number of followers are checking the social network and when not many others are posting. In other words, if you want someone to hear what you have to say, say it when they are around to listen and also when no one else is talking to them.

1.1. Our Contribution

The above discussion suggests that strategic timing can be used to increase the number of impressions received by a piece of content in a social network. In this work we show that this is indeed the case and

that the gains from strategic timing can be quite large. We consider the situation where a user is trying to maximize the probability that a target follower sees his post. We begin by proposing a model for how the follower creates an impression for the user's post in a social network. This model is fairly general and applies to several major social networks such as Twitter and Instagram. The main components of the model are 1) the *arrival process* of the follower to the social network, 2) the posting process from all users the follower follows, which we call the *timeline process*, and 3) how many posts the follower views each time he arrives to the social network, which we call the *cutoff window*. With this model we are able to calculate the optimal times for the user to post content in order to maximize the probability that it is seen by the target follower. We also provide simple, tractable approximations that allow for easy calculation of the impression probability.

In order to calculate the impression probability using our model, we need to know the rates for the arrival and timeline processes. To do this, we develop a parametric Bayesian model for the rate functions. The rate functions capture the key temporal patterns seen in typical user behavior on social networks, such as daily and weekly oscillations in the arrival and timeline rates. Using data from actual users on Twitter, we estimate these model parameters and obtain values for the rate functions. We then use the estimated rate functions to calculate impression probabilities.

We find that strategic timing of content can significantly increase the number of impressions received. This is a huge gain given that strategic timing has no additional cost and requires virtually no extra effort on the part of the user. In addition, we show both theoretically and empirically that our impression probability calculations are robust to uncertainty in model parameters. This is an important result since some model parameters, such as the arrival rate and cutoff window, are difficult to measure. Therefore, we show that strategic timing can be achieved using publicly available data which is easily accessible by anyone.

We then study the impact of wide adoption of strategic timing of content. In particular, we consider a hypothetical application that optimizes the posting times of several users at the same time. This application would naturally have to balance the usual trade-off between the aggregate performance and fairness objectiveness. Furthermore, both objectives can lead to a potentially intractable global optimization problem. However, we will show the somewhat surprising result that improving fairness tends to improve average performance and that solutions that are nearly optimal for both objectives can easily be constructed.

This work has immediate applications to firms or individual users who utilize social media for promotional purposes or to gain influence. Strategic timing is an easily implementable and cost effective tactic for increasing the impressions for content in social networks and can seamlessly be integrated with a broader social media campaign or strategy. It requires no modification to the type or volume of content created. The method presented here can be directly applied by any social network user because the required data is publicly available and selecting the time of a post does not incur any direct costs.

The rest of the paper is structured as follows. We begin with a review of previous relevant work in Section 1.2. In Section 2 we present our model for creating impressions in social networks and our calculation of the impression probability. In Section 3 we present our Bayesian models for the arrival and timeline processes and their estimation using data from sevaral Twitter users. We calculate impression probabilities for these users and show the robustness of the calculations to uncertainty in model parameters in Section 4. We show how real Twitter users can modify their posting times to increase their impressions and the corresponding impression gains in Section 5. The optimization of posting times is presented in Section 6. We then summarize the main insights regarding strategic timing in Section 7. All proofs are located in the Appendix.

1.2. Previous Work

Our work belongs to the body of research concerning the spread of information in social networks. In this area the questions of interest are how to model and predict the diffusion of the information and determine what features are conducive to rapid spreading of the information. We now review the major work in these areas.

Several researchers have looked at how features of individuals or the content affect the spread of information. Aral and Walker (2012) conduct a randomized experiment with 1.3 million Facebook users and find that features such as gender, age, and marital status are predictive of influence and susceptibilty. Katona et al. (2011) study adoption data for an online social network and find that features of the local network structure of an individual such as their degree and edge density impact their adoption probability. A similar study done in Ugander et al. (2012) on Facebook adoption data shows that the number of connected components of a user's local network impacts the adoption probability. In Twitter, studies of the information spreading problem have focused on retweets. Several works have studied what features of the users and content of the tweet cause users to retweet (Peng et al. 2011), (Suh et al. 2010) (Naveed et al. 2011) (Petrovic et al. 2011). Through various prediction techniques they show that user features such as degree, and tweet features such as the presence of hashtags or URLs impact the likelihood of a retweet.

Other work on the retweet problem aim not to predict the likelihood of an individual retweet, but the total number of retweets received by a tweet. Hong et al. (2011) and Bandari et al. (2012) use a variety of algorithms to predict not the exact number of retweets, but rather a coarse interval for the number of retweets of a tweet. In particular, Hong et al. (2011) investigates the factors influencing information propagation in Twitter including message content, temporal information, and users' social graph. Bandari et al. (2012) use regression and classification algorithms to show that it is possible to predict ranges of popularity of tweets with reasonable accuracy. Zaman et al. (2014) predict the final retweet count of a tweet using the time-series path of its retweets. They use a Bayesian approach to develop a probabilistic model for the evolution of the retweets using the retweet times and the social network structure. Their approach predicts the exact number of retweets of a tweet within minutes of its posting time with a very low error.

The problem of maximizing the spread of information in a social network was framed as an optimization problem in Kempe et al. (2003). Here the goal is to select the best set of seed users to initialize with the information in order to reach the maximum number of users through natural diffusion on the social network. A greedy algorithm was proposed to solve this combinatorial optimization problem. Variants of this problem and algorithmic solutions were subsequently studied in Kempe et al. (2005), Chen et al. (2009), Chen et al. (2010).

Our work is also related to the research on competition for attention in social networks. Several authors have looked at the problem of competing for attention in a crowded social network. Van Zandt (2004) first proposed the idea that receiver attention will become a bottleneck as it becomes easier to send messages or share information. Anderson and De Palma (2012) and Anderson and De Palma (2013) study a similar model and show that higher costs for posting content or messages can increase the average utility of a message and increase the overall viewing of messages. Iyer and Katona (2015) propose a model for the incentives for entering a social network and posting content. They investigate the choice of users to post or just view content and also how the network structure affects entry to the network and competition for attention. They show that this model predicts the so called participation inequality phenomenon where as the communication span of the social network increases, a smaller fraction of the users actually post content, but do so at higher frequency.

Our work differs from the approach of this extant work because we do not explicitly consider costs or incentives in theoretical models of social networks. Rather, we use data from real social networks to model the dynamics of user behavior and show how this model can be used to select the timing of a post in order to maximize the chance of it gaining someone's attention. Since our work is based upon real user data, it can be directly applied to many popular social networks such as Twitter or Instagram.

2. Impression Probability Model

We now present a model of how a user views content, or generates an impression, in a social network. Our model assumes that content is displayed to the user in chronological order, with the most recently posted content being seen first. This method of displaying content is used by social networks such as Twitter and Instagram. We assume that users access the social network through an application on their mobile phones, which is the predominant way to access many social networks. For instance, over 80% of Twitter users access the social network through a mobile application (Twitter 2015). The main implication of this assumption is that because of the size of a mobile phone screen, additional user action is required in order to see older content which is not immediately displayed on the screen. As we will show, the assumptions on the chronological ordering of content and the user action required to view older content, plus the temporal dynamics in the behavior of users of the social network (both content creators and consumers) lead to significant differences in the impression probability of content as a function of the time it is posted. We

next discuss the elements of our model in detail and show how they can be used to calculate the impression probability.

2.1. Model Components

Our model is for a specific user seeing a piece of content (creating an impression). We refer to this user viewing the content as the *follower*. We want to calculate the probability that a piece of content posted at time t by another user is seen by the follower. To do this, we must model three components of the follower: his arrival process to the social network, his timeline process, and the user interface for the social network application. The arrival process models the times when the follower checks the social network application for new content. Throughout the paper we refer to the content as a post, and use both terms interchangeably. In a social network a follower can choose to follow other people. We will use the Twitter terminology and refer to these users as the follower's *friends*. When the follower checks the social network application, he will look at his *timeline*, which contains posts from all of his friends. The timeline process models the arrivals of posts to the follower's timeline. The user interface characterizes how the follower views the content. The timeline posts are arranged in chronological order, with the most recent post located at the top of the timeline. As time passes, older posts are pushed down on the timeline. Typically, a follower only looks at a certain number of new posts when he checks his timeline. We refer to this number of posts as the follower's *cutoff window*.

Our main assumption regarding the arrival and timeline processes is that they can be modeled as *non-homogeneous Poisson processes* with time varying rates. For any times t and $t+s$ ($t, s > 0$) and a follower u, we denote the number of arrivals in the arrival and timeline processes in the time interval $[t, t+s)$ as $M^u(t, t+s)$ and $N^u(t, t+s)$, respectively. In other words, $M^u(t, t+s)$ represents the number of times the follower checks the social network for new content in the time interval $[t, t+s)$ and $N^u(t, t+s)$ represents the number of new posts on the follower's timeline in the time interval $[t, t+s)$. The rate of the arrival process is $\psi^u(t)$ and the rate of the timeline process is $\lambda^u(t)$. Another useful definition is the mean number of arrivals in a time window $[t, t+s)$ for $s \geq 0$, which is given by

$$\mathbf{E}\left[M^u(t, t+s)\right] = \Psi^u(t, t+s) = \int_t^{t+s} \psi^u(\tau) d\tau \tag{1}$$

for the arrival process and

$$\mathbf{E}\left[N^u(t, t+s)\right] = \Lambda^u(t, t+s) = \int_t^{t+s} \lambda^u(\tau) d\tau \tag{2}$$

for the timeline process. We will specify the parametric form of these rates and their temporal dependence in Section 3 when we look at user data from Twitter. The cutoff window of a follower is denoted by C^u. For convenience, we list the model parameters in Table 2.1. Under these modeling assumptions, we will next show how to calculate the probability of an impression being created by the follower for a piece of content as a function of the time it was posted.

Parameter	Definition
u	Follower index
C^u	Cutoff window
$M^u(t, t+s)$	Arrival process count
$N^u(t, t+s)$	Timeline process count
$\psi^u(t)$	Arrival process rate
$\lambda^u(t)$	Timeline process rate
$\Psi^u(t)$	Average arrival process count
$\Lambda^u(t)$	Average timeline process count

Table 1 Description of parameters for components of the impression probability model.

2.2. Impression Probability

In order for a follower to see a piece of content, or create an impression, the following events must occur. First, the content is posted at a time that is chosen by the content producer. The follower must check his timeline at a time after the content is posted. Second, the time when the follower checks his timeline is determined by the arrival process. In the time interval between the posting of the content and the follower arrival, there will be a number of new posts on the follower's timeline, which is determined by the timeline process. If this number of posts is less than the follower's cutoff window, the follower will see the content and an impression is created, otherwise the content is not seen.

We define the probability that a follower sees a piece of content posted to his timeline at time t as $q(t)$. This will be referred to as the *impression probability*. We have the following result for the value of the impression probability in this model.

THEOREM 2.1. For a piece of content posted at time $t > 0$, let $q(t)$ be the impression probability of a follower with arrival rate $\psi(t)$, timeline rate $\lambda(t)$, and cutoff window C. For $s > 0$, let the mean number of arrivals in the arrival and timeline processes be given by equations (1) and (2), respectively. Then,

$$q(t) = \sum_{k=0}^{C} \frac{1}{k!} \int_0^\infty \psi(t+s)(\Lambda(t, t+s))^k e^{-(\Psi(t,t+s)+\Lambda(t,t+s))} ds. \tag{3}$$

For any follower we are able to obtain a closed form expression for their impression probability as a function of posting time given their arrival rate, timeline rate, and cutoff window. Equation (3) appears complicated upon first glance, but we will see that in practice, it can be greatly simplified without sacrificing very much in terms of operational performance.

2.3. Proportional Timeline and Arrival Rates

To gain insight into the impression probability function, we consider the situation where the timeline and arrival rates of a follower are proportional to each other. Assuming non-homogeneous Poisson processes for both the timeline and arrival process, we obtain a simple expression for the impression probability given by the following lemma.

LEMMA 1. *For a piece of content posted at time $t > 0$, let $q(t)$ be the impression probability of a follower with timeline rate $\lambda(t)$, arrival rate $\psi(t) = a\lambda(t)$ for some $a > 0$, and cutoff window C. Then,*

$$q(t) = 1 - (a+1)^{-(C+1)}. \tag{4}$$

According to this expression, when the arrival and timeline rates are proportional, then the impression probability is constant in time and there is no value for strategic timing for a content producer. Also, this expression clearly shows the impact of the cutoff window on the impression probability. Not surprisingly, we see that $q(t)$ approaches one as C increases. This means that if a follower has a large cutoff window and checks a large number of posts, then it is more likely he will see any post. In addition, for proportional arrival and timeline rates, the convergence to one is exponentially fast in C.

We next look to examine the meaning and impact of the constant a in the above expression. To do this, we assume that all friends of the follower have the same arrival rate ψ and that each time they arrive, they post a new piece of content. For a follower with F friends, this means that the timeline rate will be $\lambda = F\psi(t)$, or equivalently, $\psi(t) = a\lambda$ and $a = F^{-1}$. If we substitute this expression for a in equation 4, we find that

$$q(t) = 1 - \left(\frac{F}{F+1}\right)^{C+1}.$$

This shows that the impression probability is a decreasing function of the number of friends F. This is expected because having a larger number of friends will result in a larger timeline rate, which means a post is visible for a shorter amount of time. Therefore, it will be harder for a post to be seen and we expect $q(t)$ to be smaller.

Though this analysis made strong assumptions about follower behavior, it does align with our intuition about what increases the impression probability. Our result shows that followers with a higher impression probability do not have a large amount of competing posts on their timelines (small friend count) and check many older posts on their timelines (large cutoff window).

2.4. Approximating the Impression Probability

While simple expressions can be obtained for the impression probability when we assume proportional timeline and arrival rates, they are difficult to obtain in a more general setting. However, in certain parameter regimes which occur in practice, simple approximations which provide useful insight can be obtained. We now provide and analyze such approximations.

In equation (3) we have assumed that the timeline process is a non-homogeneous Poisson process. To simplify the expression for $q(t)$, we now assume the timeline process is deterministic with rate $\lambda(t)$, so that the number of arrivals in an interval $[t, t+s)$ is given by $N(t, t+s) = \Lambda(t, t+s)$. Define the residual time for an arrival given that there has not been an arrival at time t as S_t. For a non-homogeneous Poisson process with rate $\psi(\cdot)$, it can be shown that the random variable S_t has a density given by $f_{S_t}(s) =$

$\psi(t+s)e^{-\Psi(t,t+s)}$, where we have defined $\Psi(t,t+s)=\int_0^s \psi(t+\tau)d\tau$ (Cox and Isham 1980). With this assumption, an impression is created if $\Lambda(t,t+S_t)\leq C$. The probability of this is given by

$$\begin{aligned} q_1(t) &= \mathbf{E}\left[\mathbb{1}\left(\Lambda(t,t+s)\leq C\right)\right] \\ &= \int_0^\infty \mathbb{1}\left(\Lambda(t,t+s)\leq C\right) f_{S_t}(s)ds \\ &= \int_0^\Delta \psi\left(t+s\right) e^{-\Psi(t,t+s)}ds \\ &= 1-e^{-\Psi(t,t+\Delta)}, \end{aligned} \tag{5}$$

where we define

$$\Delta = \sup\left\{s : \Lambda(t,t+s)\leq C\right\}. \tag{6}$$

The time Δ is the amount of time the content posted at time t has before it is pushed out of the cutoff window and as a result is not seen by the user. We will refer to this as the *lifetime* of the content. We obtain the following bound on the error of approximation $q_1(t)$.

THEOREM 2.2. Let the arrival rate $\psi(t)$ and the timeline rate $\lambda(t)$ of a follower be such that $F_1 \leq \lambda(t)/\psi(t)$ for some constants $1<F_1$. Let $q(t)$ be the exact impression probability given by equation (3) and let $q_1(t)$ be the approximation of the impression probability given by equation (5). Then the error of the approximation is bounded by

$$|q(t)-q_1(t)| \leq \frac{\sqrt{2\pi C}}{F_1}.$$

The approximation $q_1(t)$ will be accurate for followers with small cutoff windows and a large timeline rate relative to their arrival rate. This approximation is simpler than the exact expression, but still rather complex because it contains the content lifetime Δ and the mean arrival process Ψ. To provide a more intuitive expression, we make the following additional approximations. First we expand the exponential in equation (5) to first order in $\Psi(t,t+\Delta)$ to obtain $1-e^{-\Psi(t,t+\Delta)}\approx\Psi(t,t+\Delta)$. Next, we use a first order Taylor approximation about t for the arrival process mean to obtain $\Psi(t,t+\Delta)\approx\psi(t)\Delta$. Finally, we approximate the timeline process mean using a first order Taylor expansion about t to obtain $\Lambda(t,t+s)\approx\lambda(t)s$ for $s\leq\Delta$. Using this approximation for $\Lambda(t,t+s)$ along with equation (6) we obtain $\Lambda(t,t+\Delta)\approx C\approx\lambda(t)\Delta$, or more simply $\Delta\approx C/\lambda(t)$. These Taylor approximations will be accurate as long as Δ is small compared to the time scale over which the timeline and arrival rates vary. This will typically be true because Δ is on the order of minutes, whereas the arrival and timeline rate variations are on the order of hours or days (more on this in Section 3). Combining these approximations we obtain the following simple equation for the impression probability

$$q_2(t) = C\frac{\psi(t)}{\lambda(t)}. \tag{7}$$

The above approximation is a simple, intuitive expression for the impression probability which captures the main aspects of our model. We can easily see the following important facts regarding the impression probability from equation (7). First, if a follower checks many posts upon each visit to the social network (large cutoff window C), then the impression probability is higher. Second, the times when the follower is likely to be on the social network correspond to times when the arrival rate $\psi(t)$ is large, and this is also when the impression probability is high. Finally, the times when there are a large number of posts on the follower's timeline correspond to times when the timeline rate $\lambda(t)$ is large, and it is at these times when the impression probability will be lower. Furthermore, for followers whose arrival rate is much smaller than their timeline rate, equation (7) is a very accurate approximation for the impression probability. This is made precise in the following theorem.

THEOREM 2.3. Let the arrival rate $\psi(t)$ and the timeline rate $\lambda(t)$ of a follower be such that $F_1 \leq \lambda(t)/\psi(t) \leq F$ for some constants $1 < F_1 \leq F$. Let $q(t)$ be the exact impression probability given by equation (3) and let $q_2(t)$ be the approximation of the impression probability given by equation (7). Then the error of the approximation is bounded by

$$|q(t) - q_2(t)| \leq \frac{\sqrt{2\pi C}}{F_1} + \frac{C^2 + C(F - F_1)}{F_1^2}.$$

The approximation error for $q_2(t)$ is slightly larger than for $q_1(t)$ due to the extra approximations made. However, the regimes where both approximations are good are similar: large timeline rate and small cut-off window. We will see in Section 4 that many real Twitter users fall in this regime and therefore both approximations are highly accurate.

2.5. Impact of Arrival Rate versus Posting Rate

To calculate the impression probability in our model the follower's arrival rate is required. In practice this may be difficult to measure. However, measuring the posts of a follower can be done rather easily. For instance, this data can be obtained for a Twitter user from the Twitter API. The posting rate will be less than the arrival rate because in order to post, a follower must first arrive to the social network. We model the posting process as a random sample of the of the arrival process, with sampling probability $\gamma \in (0, 1]$. This means that each time the follower arrives, he posts with probability γ. The posting process is then a non-homogeneous Poisson process with rate $\gamma\psi(t)$. We show here that having uncertain knowledge of the value of γ does not impact key predictions of our model, namely the optimal time to post.

We establish the following result for the impact of using the posting rate instead of the arrival rate to find the optimal time to post. From equation (7) we define the following approximation for the impression probability when one uses the posting rate $\gamma\psi(t)$ instead of the arrival rate $\psi(t)$ as

$$\widehat{q}_2(t) = C\frac{\gamma\psi(t)}{\lambda(t)}, \tag{8}$$

and we define the time which maximizes this expression as

$$\widehat{t}_2 = \arg\max_t \widehat{q}_2(t). \tag{9}$$

We also define the time that maximizes the true impression probability as

$$t^* = \arg\max_t q(t) \tag{10}$$

We are concerned with the difference between the impression probability if one posts at $\widehat{t}_2$ versus t^*. We have the following result, which can be seen as a simple application of Theorem 1.

COROLLARY 1. *Let the arrival rate $\psi(t)$ and the timeline rate $\lambda(t)$ of a follower be such that $F_1 \leq \lambda(t)/\psi(t) \leq F$ for some constants $1 \leq F_1 \leq F$. Let $q(t)$ be the exact impression probability given by equation* (3)*. Let t^* be given by equation* (10) *and for $\gamma \in (0, 1]$, let $\widehat{t}_2$ be given by equation* (9)*. Then we have that*

$$\left|q(t^*) - q(\widehat{t}_2)\right| \leq 2\frac{\sqrt{2\pi C}}{F_1} + 2\frac{C^2 + C(F - F_1)}{F_1^2}.$$

This result shows that even if an approximation for the impression probability is used with the posting rate and not the arrival rate, the actual impression probability will not be changed by a very large amount for users with a small cutoff window relative to the ratio of their timeline rate to their arrival rate. Therefore, the posting rate can be used in place of the arrival rate and the approximation $q_2(t)$ can be used instead of $q(t)$ to determine the optimal time to post content. This is important because it is difficult to measure the arrival rate for a user, while the posting rate is readily available.

2.6. Impact of Cutoff Window

The cutoff window is another model parameter that may be difficult to measure. For companies running social networking applications this parameter can be measured by recording what content is displayed on a follower's screen. However, this information is not readily available to those not within the company. Due to this restriction, we would like to understand how sensitive the impression probability is to this parameter. We define $q_2(t, C')$ as the value of the approximation to the impression probability given equation (7) for a cutoff window C'. For an arbitrary value of the cutoff window $C' > 0$ we define the time which maximizes $q_2(t, C')$ as

$$\widehat{t}_{C'} = \arg\max_t q_2(t, C'). \tag{11}$$

The time that maximizes the true impression probability using the correct cutoff window C, arrival rate $\psi(t)$ and timeline rate $\lambda(t)$ is t^* as defined in equation (10). We have the following result, which again is simply an application of Theorem 1.

COROLLARY 2. *Let the arrival rate $\psi(t)$ and the timeline rate $\lambda(t)$ of a follower be such that $F_1 \leq \lambda(t)/\psi(t) \leq F$ for some constants $1 \leq F_1 \leq F$. Let $q(t)$ be the exact impression probability given by equation* (3) *using the true cutoff window C. Let t^* be given by equation* (10) *and for $C' > 0$, let $\widehat{t}_{C'}$ be given by equation* (11)*. Then we have that*

$$\left|q(t^*) - q(\widehat{t}_{C'})\right| \leq 2\frac{\sqrt{2\pi C}}{F_1} + 2\frac{C^2 + C(F - F_1)^2}{F_1^2}.$$

Here we see that as long as the arrival rate is much smaller than the timeline rate, then the impression probability is robust to errors in the cutoff window. This is an important result because of the difficulty in measuring this parameter.

3. Bayesian Model For Timeline and Posting Processes

We have seen how to calculate the impression probability once the follower arrival rate and timeline rate are both known. We also saw that we can get a good approximation to the impression probability if we use the posting rate instead of the arrival rate. In this section we present Bayesian models for the timeline and posting processes for users in Twitter. We focus on the posting process instead of the arrival process because we can obtain data on the posting process through the Twitter API. We take a Bayesian approach because it provides a very natural way to obtain credibility intervals for all model parameters and the model estimation procedure is fairly direct. This then allows for simple calculation of credibility intervals for the impression probabilities, which we present in Section 4.

3.1. Exploratory Data Analysis

To begin our model development, we perform an exploratory analysis of the posting and timeline processes for 1498 Twitter users. These users are random samples of the followers of prominent Twitter users such as Barack Obama (@BarackObama), Taylor Swift (@taylorswift12), and several others. To obtain the timeline of each user in our dataset we collected the posts of a maximum 100 of their friends. This allowed us to reconstruct a sampled version of the timeline process for each user. We only collected the posts of 100 friends per user because of rate limitations by the Twitter API. All post times are in GMT, which is the default for the Twitter API.

We begin by studying the temporal variation in the timeline processes. We define the smoothed timeline rate as the posts per unit time in a sliding six hour window. We chose this window size to produce a smooth curve which would show the main qualitative features of the timeline rate. We plot the smoothed timeline rate for different Twitter users in Figure 1. As can be seen, there are clear oscillations over a one day period. Also, the peak time of these oscillations are different for each user.

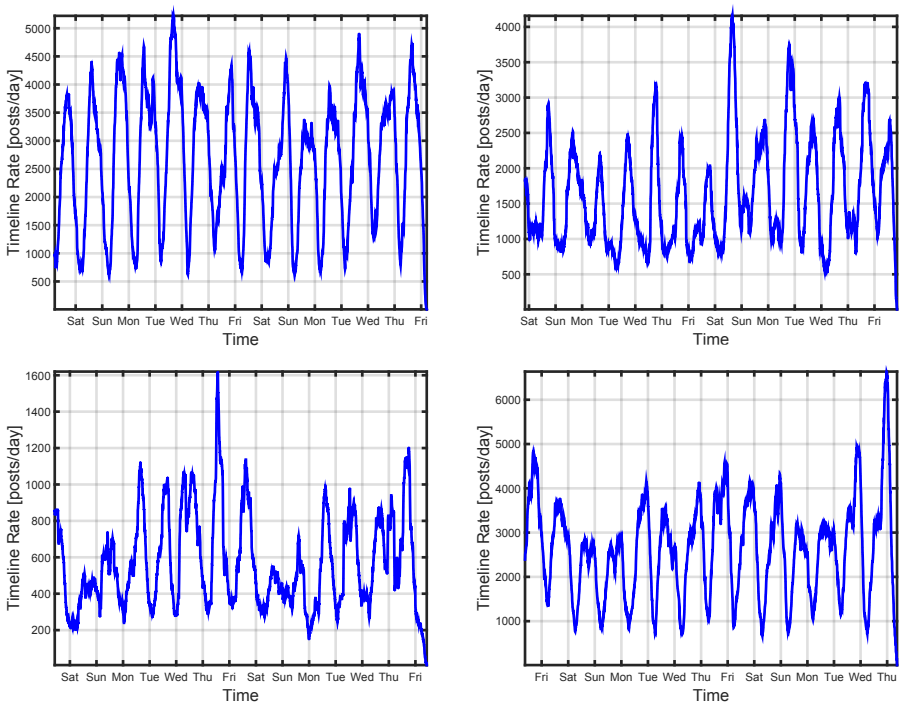

Figure 1 **The smoothed timeline rate for Twitter users with screen names (top left) @444sai, (top right) @AbdeRules, (bottom left) @AsifSKhan, (bottom right) @Birdseye1. Each timeline was constructed from 100 friends of the user. The timeline rate was smoothed using a sliding six hour window.**

We next look more closely at the daily variations in both the timeline and posting proceses. We plot a histogram of the hour of posts in the timeline and posting process for each of the users from Figure 1 in Figure 2. As can be seen, for these users there is considerable variation in the number of posts versus hour for both their timeline and posting processes. Also, the hourly distribution for the two processes can be very different. The one day period of oscillation for the timeline rate is clearly visible in the data. We next examined variations over a one week period. We show a histogram of the day of the posts in the timeline and posting processes for the same users in Figure 3. As with the hour data, we see variation across these users in the day distribution for both processes. Also, the timeline and posting processes can be very different in their day distributions for a given user.

To further show the difference in the peak times for the posting and timeline processes, we plot in Figure 4 the peak hour and peak day for the timeline and posting process for each user in our dataset. The points in the plot are jittered with Gaussian noise with standard deviation 0.25 in order to make them more clear. As can be seen, there is very little correlation between the timeline and posting processes in terms of peak day and peak hour. For the two processes, the correlation of the peak hour is 0.31 (p-value $< 10^{-6}$) and the correlation of the peak day is 0.06 (p-value $= 0.0042$). There is more correlation in the hour than in the day, but both values are much smaller than one.

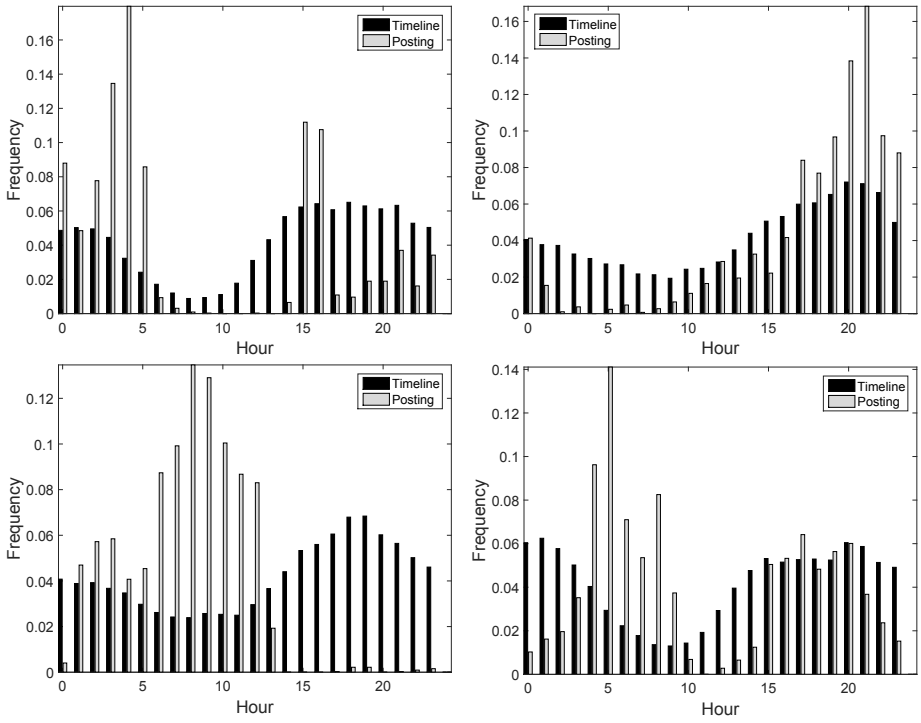

Figure 2 Histograms of the hours of posts from the posting and timeline process for Twitter users with screen names (top left) @444sai, (top right) @AbdeRules, (bottom left) @AsifSKhan, (bottom right) @Birdseye1.

This exploratory analysis provides two important conclusions. The first is that the posting and timeline processes of users show variations with a one day and one week period. This is not a surprising result given the manner in which users typically use social networks. Second, the peak hour and day of the posting and timeline processes are not aligned for many users. We saw in Section 2.3 that when the timeline and arrival rates were equal, then the impression probability was constant in time and there was no gain from strategic timing. However, when the rates are not proportional, strategic timing can produce substantial gains. Because of the observed misalignment of the peak hour and day for the two processes for several users, we expect strategic timing of content to be beneficial in practice.

3.2. Parametric Form of Rate Functions

We now propose a parametric form for the timeline rate and posting rate functions. Because we assume that the posting process is a random sample of the arrival process, this parametric model for the rate of the arrival process will be a scaled version of the rate of the posting process. Timeline posts are an aggregation of user posts, so it is reasonable to assume that the posting and timeline processes have a similar parametric form (but with different parameter values). Based upon the analysis in Section 3.1 we assume that all rates have oscillations with a one day and one week period.

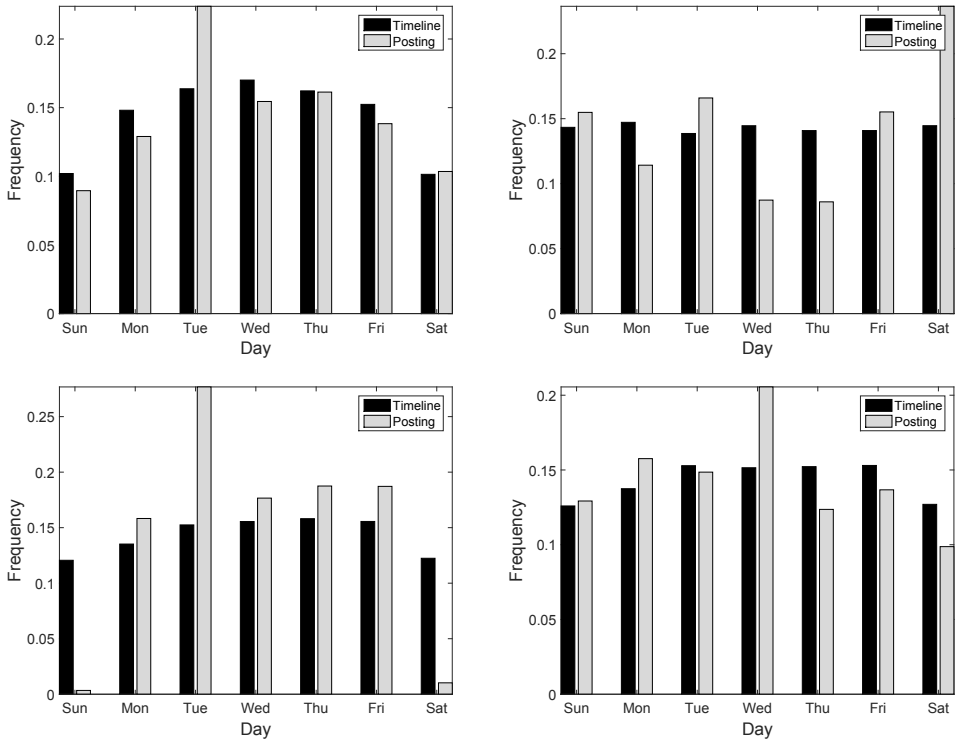

Figure 3 **Histograms of the days of posts from the posting and timeline process for Twitter users with screen names (top left) @444sai, (top right) @AbdeRules, (bottom left) @AsifSKhan, (bottom right) @Birdseye1.**

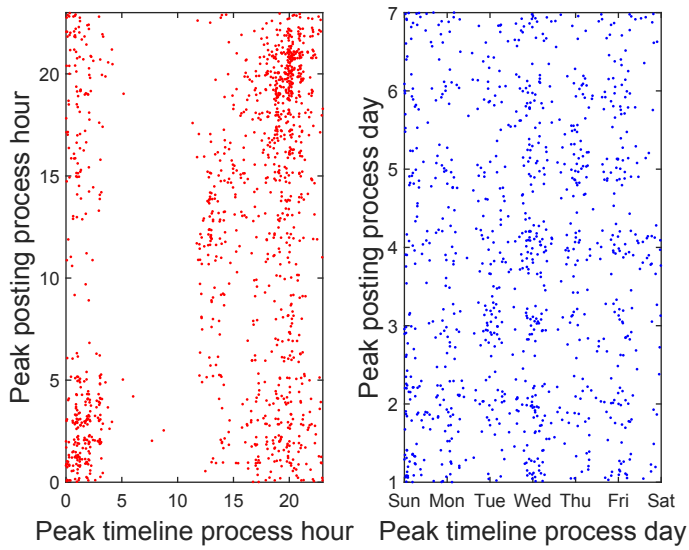

Figure 4 **(left) Plot of the peak hour of the timeline process versus the posting process for several Twitter users. (right) Plot of the peak day of the timeline process versus the posting process for several Twitter users. All data points are jittered with Gaussian noise with a standard deviation of 0.25.**

In our model the timeline process of a user u consists of the following parameters: $\{\lambda_0^u, b_h^u, b_d^u, t_h^u, t_d^u\}$. The timeline rate is then given by

$$\lambda^u(t) = \lambda_0^u \left(1 + b_h^u \cos\left(\omega_h \left(t - t_h^u\right)\right)\right) \left(1 + b_d^u \cos\left(\omega_d \left(t - t_d^u\right)\right)\right). \tag{12}$$

The parameters t_h^u and t_d^u indicate the peak hour and day of the timeline rate. The strength of the hourly and daily variations are modeled by the terms b_h^u and b_d^u. The average timeline rate is captured by λ_0^u. If we measure the time t in days, then the frequencies above become $\omega_h = 2\pi$ (one day period) and $\omega_d = 2\pi/7$ (seven day period).

We model the posting process rate using a similar parametric form as in (12) keeping ω_h and ω_d the same, but using different parameters $\{\psi_0^u, \beta_h^u, \beta_d^u, \tau_h^u, \tau_d^u\}$.

$$\psi^u(t) = \psi_0^u \left(1 + \beta_h^u \cos\left(\omega_h (t - \tau_h^u)\right)\right) \left(1 + \beta_d^u \cos\left(\omega_d (t - \tau_d^u)\right)\right). \tag{13}$$

The parametric form of the rates allow us to express the mean value of the processes in closed form. The mean timeline process is given by

$$\begin{aligned}
\Lambda^u(t, t+s) =& \lambda_0^u s + \lambda_0^u \frac{b_h^u}{\omega_h} \sin\left(\omega_h (r - t_h^u)\right) \Big|_{r=t}^{t+s} + \lambda_0^u \frac{b_d^u}{\omega_d} \sin\left(\omega_d (r - t_d^u)\right) \Big|_{r=t}^{t+s} + \\
& \lambda_0^u \frac{b_h^u b_d^u}{2(\omega_d + \omega_h)} \sin\left((\omega_d + \omega_h)(r - (t_d^u + t_h^u))\right) \Big|_{r=t}^{t+s} + \\
& \lambda_0^u \frac{b_h^u b_d^u}{2(\omega_d - \omega_h)} \sin\left((\omega_d - \omega_h)(r - (t_d^u - t_h^u))\right) \Big|_{r=t}^{t+s}
\end{aligned} \tag{14}$$

and the mean posting process is given by

$$\begin{aligned}
\Psi^u(t, t+s) =& \psi_0^u s + \psi_0^u \frac{\beta_h^u}{\omega_h} \sin\left(\omega_h (r - \tau_h^u)\right) \Big|_{r=t}^{t+s} + \psi_0^u \frac{b_d^u}{\omega_d} \sin\left(\omega_d (r - \tau_d^u)\right) \Big|_{r=t}^{t+s} + \\
& \psi_0^u \frac{\beta_h^u \beta_d^u}{2(\omega_d + \omega_h)} \sin\left((\omega_d + \omega_h)(r - (\tau_d^u + \tau_h^u))\right) \Big|_{r=t}^{t+s} + \\
& \psi_0^u \frac{\beta_h^u \beta_d^u}{2(\omega_d - \omega_h)} \sin\left((\omega_d - \omega_h)(r - (\tau_d^u - \tau_h^u))\right) \Big|_{r=t}^{t+s}.
\end{aligned} \tag{15}$$

3.3. Bayesian Model Specification

To learn the model parameters we will use a Bayesian approach. This will provide us with posterior estimates and credibility intervals for all model parameters in a very natural way. The Bayesian framework also allows for ease of model fitting when there is not sufficient data, as is the case for many users' posting processes. Because of the difference in amount of data for the timeline and posting processes, we use different model structures for each process. For the timeline process we have sufficient data for each user and therefore learn each timeline rate independently. For the posting process, because of sparse data for many

users, we use a hierarchical model which allows for information sharing between users. We now specify these models in detail.

The timeline process of a user u provides us with observations of the n^u times of the timeline posts $\mathbf{t}^u = \{t_1^u, t_2^u, ...t_{n^u}^u\}$. The timeline model parameters for user u are $\Theta^u = \{\lambda_0^u, b_h^u, b_d^u, t_h^u, t_d^u\}$. Because we model the timeline process as an non-homogeneous Poisson process with rate $\lambda^u(t)$, the likelihood of the observations is given by

$$\begin{aligned}\mathbf{P}\left(\mathbf{t}^u|\Theta^u\right) =& \prod_{i=2}^{n^u} \lambda^u\left(t_i^u\right) e^{-\Lambda^u(t_{i-1}^u, t_i^u)} \\ =& e^{-\Lambda^u(t_1^u, t_{n^u}^u)} \prod_{i=2}^{n^u} \lambda^u\left(t_i^u\right), \end{aligned} \tag{16}$$

with $\lambda^u(t)$ and $\Lambda^u(s,t)$ given by equations (12) and (14). We use uninformative hyperpriors for the timeline model parameters. For λ_0^u we use a gamma distribution with shape and scale one and 10,000. For the remaining parameters we use normal priors with zero mean and standard deviation 100.

For the posting process, many times we will have users who do not provide a large amount of data. This is in contrast to the timeline process which typically has a large amount of data for each user. Therefore, it may not be possible to accurately learn the posting rate parameters of each user individually. However, if we assume that similar behavior is shared between different users, then we may be able to better learn the parameter values. For this reason, we use a hierarchical Bayesian model for the posting process. We first define for a user u a set of posting rate parameters $\Phi^u = \{\psi_0^u, \beta_h^u, \beta_d^u, \tau_h^u, \tau_d^u\}$. For each user u we observe the corresponding m^u posting times $\mathbf{s}^u = \left\{s_1^u, s_2^u, ...s_{m^u}^u\right\}$. The likelihood of the observations conditioned on the user parameters is similar to that for the timeline observations and is given by

$$\mathbf{P}\left(\mathbf{s}^u|\Phi^u\right) = e^{-\Psi^u(s_1^u, s_{n^u}^u)} \prod_{i=2}^{n^u} \psi^u\left(s_i^u\right), \tag{17}$$

with $\psi^u(t)$ and $\Psi^u(s,t)$ given by equations (13) and (15). To couple the user parameters, we make them each independent conditional on a set of global paramters. We define this set of global parameters as $\Phi = \left\{\psi_0, \beta_h, \beta_d, \tau_h, \tau_d, \sigma_{\beta_h}, \sigma_{\beta_d}, \sigma_{\tau_h}, \sigma_{\tau_d}\right\}$. The parameters characterize the typical value of the user parameters across all users. The parameter ψ_0 characterizes the typical posting rate. The β terms characterize the typical strength of the hourly and daily variations and the σ_β^2 terms model how variable individual users are. The same goes for the τ and σ_τ^2 terms and the corresponding peak hours and days. The distributions of the user parameters conditioned on the global parameters are

$$\psi_0^u|\psi_0 \sim \text{Exp}\left(\psi_0^{-1}\right) \tag{18}$$

$$\beta_h^u|\beta_h, \sigma_{\beta_h} \sim \mathcal{N}\left(\beta_h, \sigma_{\beta_h}^2\right) \tag{19}$$

$$\beta_d^u|\beta_d, \sigma_{\beta_d} \sim \mathcal{N}\left(\beta_d, \sigma_{\beta_d}^2\right) \tag{20}$$

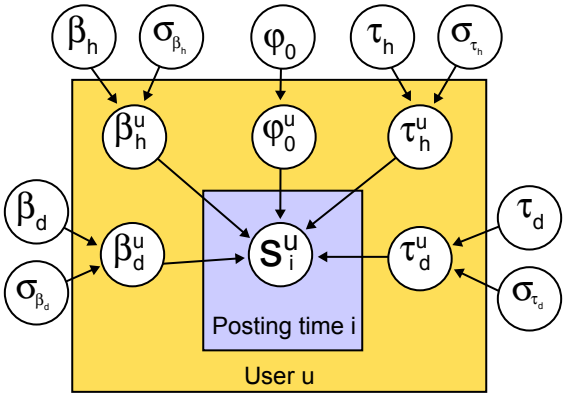

Figure 5 Graphical model of the hierarchcial Bayesian model for the posting process. The plates denote replication over posting times S_i^u and users u. Hyperpriors are omitted for simplicity.

$$\tau_h^u|\tau_h, \sigma_{\tau_h} \sim \mathcal{N}\left(\tau_h, \sigma_{\tau_h}^2\right) \tag{21}$$

$$\tau_d^u|\tau_d, \sigma_{\tau_d} \sim \mathcal{N}\left(\tau_d, \sigma_{\tau_d}^2\right), \tag{22}$$

where $\text{Exp}\,(x)$ denotes an exponential distribution with mean x and $\mathcal{N}\,(\mu, \sigma^2)$ denotes a normal distribution with mean μ and standard deviation σ. Hyperpriors are chosen to be uninformative and conjugate when possible. The hyperprior on ψ_0 is inverse gamma with parameters one and one. For all means of the normal conditional distributions we use normal priors with zero mean and standard deviation 100. All the variances of the normal conditional distributions have inverse gamma hyperpriors with parameters one and one. For convenience we include a visual depiction of the structure of the hierarchical model for the posting process in Figure 5.

For N users, let the observed posting times be $\mathbf{S} = \{\mathbf{s}^1, \mathbf{s}^2, ..., \mathbf{s}^N\}$ and let the observed timeline times be $\mathbf{T} = \{\mathbf{t}^1, \mathbf{t}^2, ..., \mathbf{t}^N\}$. We define the set of all user posting rate parameters as $\mathbf{\Phi}^{user} = \{\Phi^1, \Phi^2, ...\Phi^N\}$ and the set of all user timeline rate parameters as $\mathbf{\Theta}^{user} = \{\Theta^1, \Theta^2, ...\Theta^N\}$. The posterior distribution of the timeline process model parameters given $\mathbf{T}$ is given by

$$\mathbf{P}\left(\mathbf{\Theta}^{user}|\mathbf{T}\right) \propto \prod_{u=1}^{N} \mathbf{P}\left(\mathbf{t}^u|\Theta^u\right)\mathbf{P}\left(\Theta^u\right).$$

The posterior distribution of the posting process model parameters given $\mathbf{S}$ is given by

$$\mathbf{P}\left(\Phi, \mathbf{\Phi}^{user}|\mathbf{S}\right) \propto \mathbf{P}\left(\Phi\right) \prod_{u=1}^{N} \mathbf{P}\left(\mathbf{s}^u|\Phi^u\right)\mathbf{P}\left(\Phi^u|\Phi\right).$$

We sample from these posterior distributions using a Markov Chain Monte Carlo (MCMC) sampler. For the timeline rates the model decouples and we can sample the parameters for each user timeline individually. For the posting process because all users are coupled through the global parameters, we must jointly sample all user parameters. Details of our MCMC sampler are provided in the Appendix.

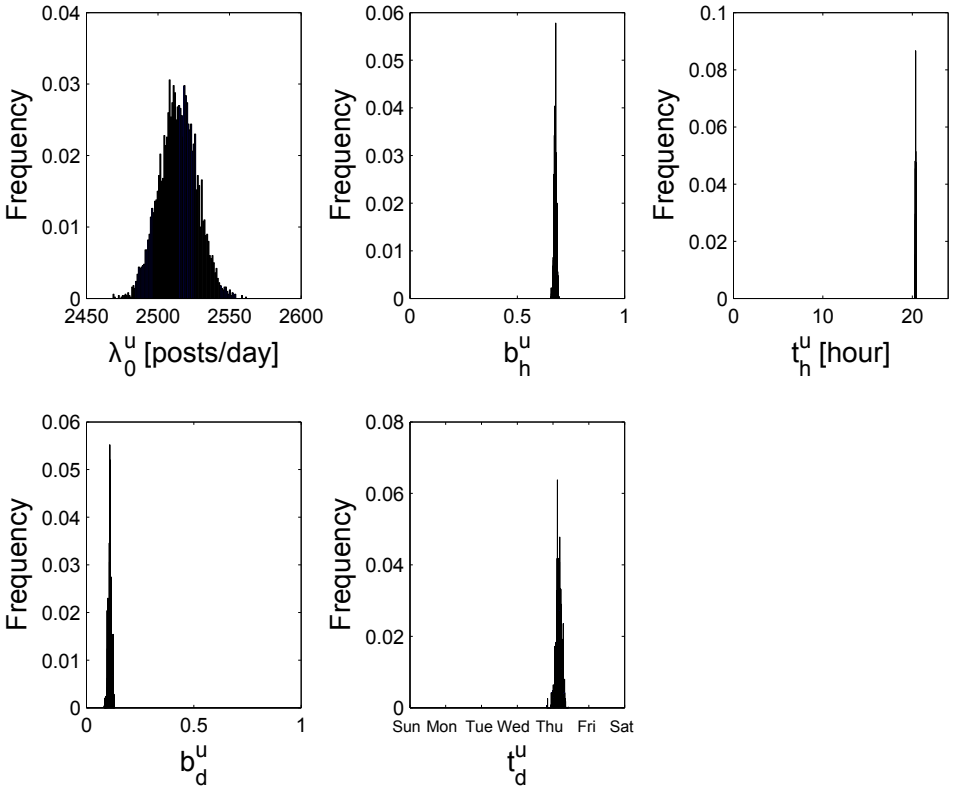

Figure 6 **Posterior histograms of the timeline rate parameters of user @444sai for the timeline process model.**

3.4. Model Estimation Results

We estimate our model on a subset of 94 users from our dataset. These users are selected at random from our dataset with the requirement that they have at least 10 posts in their timeline and posting processes. Each timeline model was learned separately for each user. For the posting processes, we estimated the parameters for all users jointly using the hierarchical model. For each model estimation (timelines and posting process), we generated posterior samples using three independent MCMC chains with dispersed starting points run for 6,000 iterations and discarding a burn-in period of 1,000 iterations. Convergence of the MCMC sampler was assessed using the Gelman-Rubin statistic (Gelman and Rubin 1992).

We plot the posterior histograms of timeline parameters for one user's timeline in Figure 6. To visually demonstrate the quality of the in-sample fit, in Figure 8 we show the posterior in-sample mean timeline process versus the true timeline process for this user, where we define this as $\widehat{\Psi}(0,t) = \mathbf{E}_\Theta\left[\Psi(0,t|\Theta)\right]$. We also plot the empirical and posterior in-sample mean timeline rates in Figure 8. The empirical timeline rate is calculated using a six hour sliding window. As can be seen, the in-sample fit is quite good for the timeline and captures much of the daily and weekly variation, which is most clearly visible in the timeline rate plots. We plot posterior histograms for the posting process for the same user in Figure 7. In Figure 8 we show

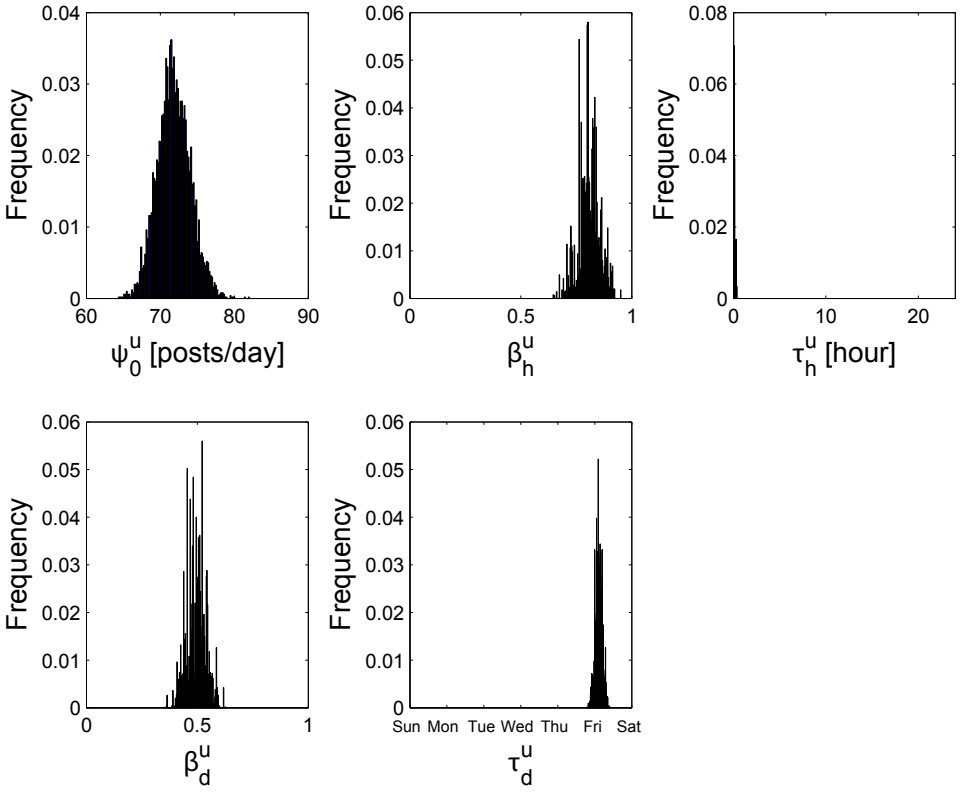

Figure 7 Posterior histograms of the posting rate parameters of user @444sai for the timeline process model.

this user's posterior in-sample mean value posting process $\widehat{\Psi}(0,t) = \mathbf{E}_\Theta\left[\Psi(0,t|\Theta)\right]$ versus the true posting process. As can be seen, the in-sample fits are quite good for the posting process as well and captures much of the daily and weekly variation. We do not plot the posting rates because the sparse number of posts (relative to the timeline process) produces a very noisy empirical rate curve.

The posterior histograms of the global parameters is shown in Figure 9. Here we see that the median posting rate for these users is 107 posts per day. The median peak hour is 13 (in GMT) and the median peak day is between Wednesday and Thursday. The median standard deviation of the user peak hour σ_{τ_h} is 14 hours. This indicates that there is a wide spread in the peak hour of posting across users. However, the median standard deviation for the peak day is 1.3 days, indicating less variation in the day of most user activity.

3.5. Model Comparisons

To assess the quality of fit of our model, which we refer to as Model 3, we compare it to two other benchmark models. The first model, which we refer to as Model 1, assumes no temporal dependence in the timeline and posting rates. The second model, Model 2, only assumes a one day periodicity in the rates, but ignores

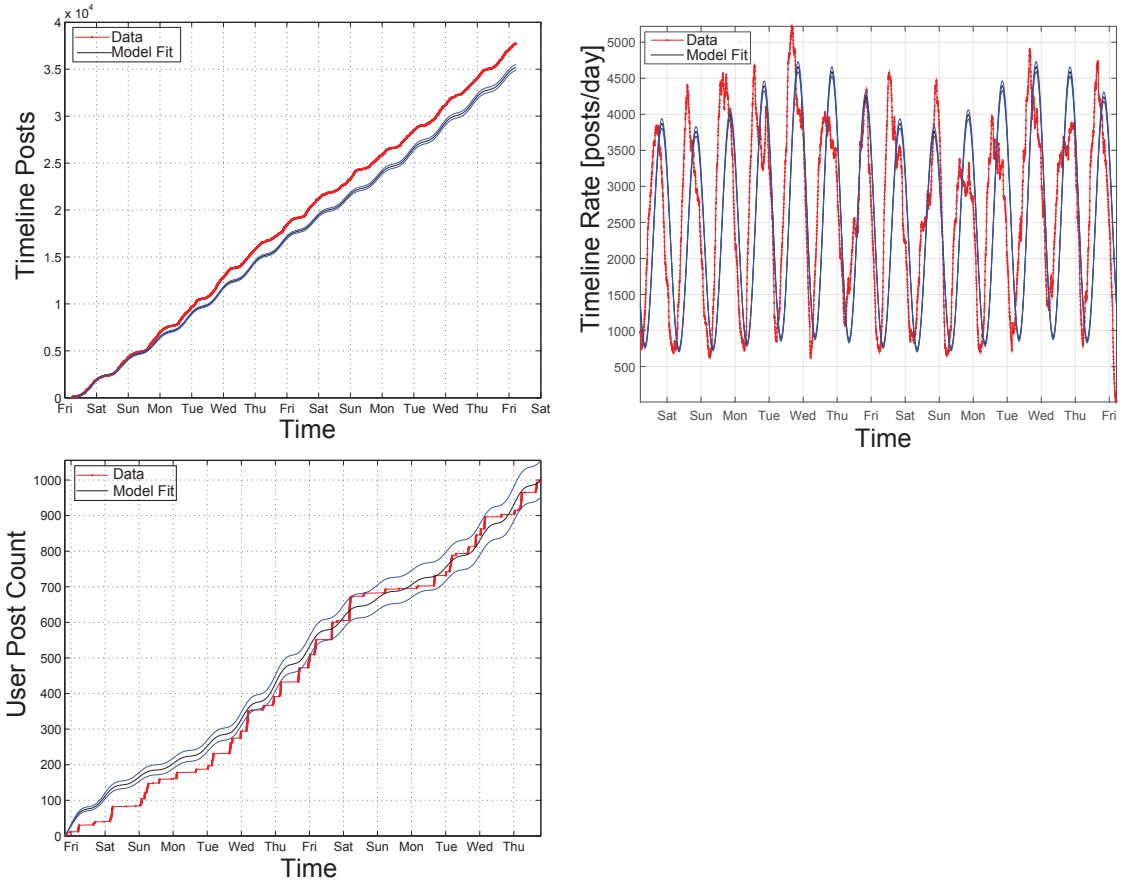

Figure 8 **Posterior fits of the timeline process (top left), timeline rate (top right) and posting process (bottom left) for user @444sai. The blue curves are the 90% posterior credibility intervals for the model fit curves.**

the weekly periodicity. Both models assume a hierarchy on the posting rate parameters, similar to Model 3 proposed in Section 3.3. We now provide detailed Bayesian specifications of these two models.

The timeline and arrival rates of Model 1 are assumed to be constant. The full model is given by

$$\lambda^u(t) = \lambda_0^u$$
$$\psi^u(t) = \psi_0^u$$
$$\psi_0^u | \psi_0 \sim \Gamma(1, \psi_0)$$

The hyperprior for λ_0^u is a gamma distributions with shape and scale parameters one and 10,000 respectively. For ψ_0, the hyperprior is inverse gamma with shape and scale parameters both equal to one.

The timeline and arrival rates of Model 2 are given by

$$\lambda^u(t) = \lambda_0^u \left(1 + a^u \left(1 + b_h^u \cos\left(\omega_h \left(t - t_h\right)\right)\right)\right)$$

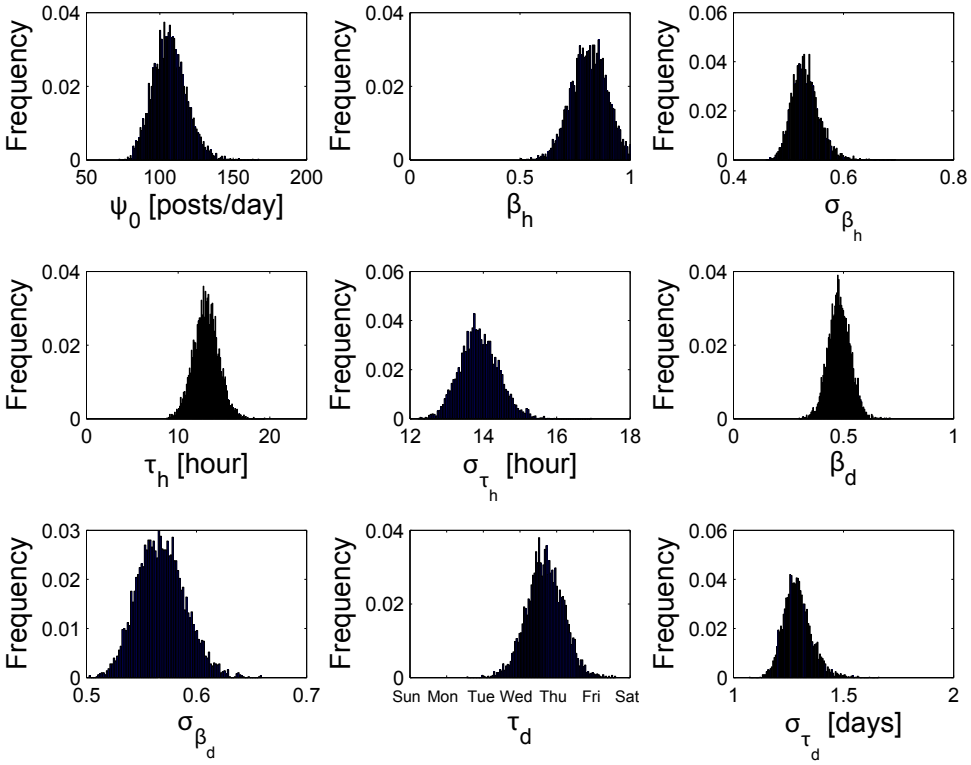

Figure 9 Posterior histograms of the global parameters for the posting process model.

$$\psi^u(t) = \psi_0^u \left(1 + \alpha^u \left(1 + \beta_h^u \cos\left(\omega_h \left(t - \tau_h\right)\right)\right)\right).$$

This is the same form as Equations (12) and (13) but without the temporal dependence on the day of the week. The hyperpriors for all model parameters here are identical to those in Section 3.3.

We use data for the same 94 Twitter users as used to fit Model 3 in Section 3.4 and a similar MCMC procedure to sample from posterior distributions of each model. The quality of the model fit is assessed using the deviance information criteria (DIC) (Spiegelhalter et al. 2002). The DIC is a standard Bayesian model fitting score that rewards models which fit the data better (through a higher likelihood) but penalizes models with a larger number of parameters. Models which fit better have a smaller DIC. The DIC of the three models for the timeline and posting processes are shown in Table 2 and Table 3, respectively. We find that Model 3 has better fit in terms of DIC than Models 1 and 2. Therefore, the inclusion of daily and weekly temporal variations in the arrival and timeline models do lead to a better fit to actual user behavior in Twitter.

4. Impression Probability Calculations

The calculations from Section 3.4 provide us with posterior samples of the timeline and posting rate parameters. We can use these to calculate the impression probability (with credibility intervals) for the users in

Timeline Process Model	Description	DIC
Model 1	No temporal variation	-5.167×10^6
Model 2	Daily variation	-5.217×10^6
Model 3	Daily and weekly variation	-5.225×10^6

Table 2 Deviance information criteria (DIC) for different timeline process models.

Posting Process Model	Description	DIC
Model 1	No temporal variation	-0.912×10^4
Model 2	Daily variation	-0.980×10^4
Model 3	Daily and weekly variation	-1.010×10^4

Table 3 Deviance information criteria (DIC) for different posting process models.

our dataset. We do so as follows. Denote the ith sample of the timeline and posting rate parameters for a user u as Θ_i^u and Φ_i^u. Using these we can calculate the timeline and posting rate for this user using equations (12) and (13). Here we are using the posting rate in place of the arrival rate. For the cutoff window we use a value of 30. This value is a reasonable value given the way in which people use social network applications. From the analysis in Sections 2.5 and 2.6 we expect our calculations not to be affected by using the posting rate or error in the cutoff window. We will also show numerical evidence supporting this claim in Sections 4.2 and 4.1.

With the timeline rate, posting rate, and cutoff window we are able to calculate the impression probability using equation (3) for a time t. Doing this for each posterior sample we obtain a distribution for the impression probability at each time t. We can then obtain an estimate for the impression probability by taking the median of these samples and quantify uncertainty in the value using the 90% credibility interval centered at the median. We show the impression probability plots for several users in Figure 10. The top left plot in the figure is for user @444sai whose posterior timeline and posting rate parameters were shown in Section 3.

We see that there is substantial gain from strategic timing of content. For instance, in the top left plot in Figure 10, random timing would produce an impression probability of approximately 0.50 and if one was unlucky the impression probability could be as low as 0.14. However, with strategic timing the impression probability can go up to 0.90. We will see in Section 5 how much gain in impressions real users can achieve with strategic timing versus their current posting behavior.

4.1. Robustness to Cutoff Error

We next look at the robustness of the impression probability calculations to uncertainty in the cutoff window value. We plot in Figure 11 the impression probability of a user for different values of the cutoff window. We see that with the cutoff window ranging from 20 to 100 there is an appreciable change in the value of the impression probability. However, there is not an appreciable change in the optimal posting time. Therefore, without knowing the actual cutoff window, we can still determine the best time to post with low error.

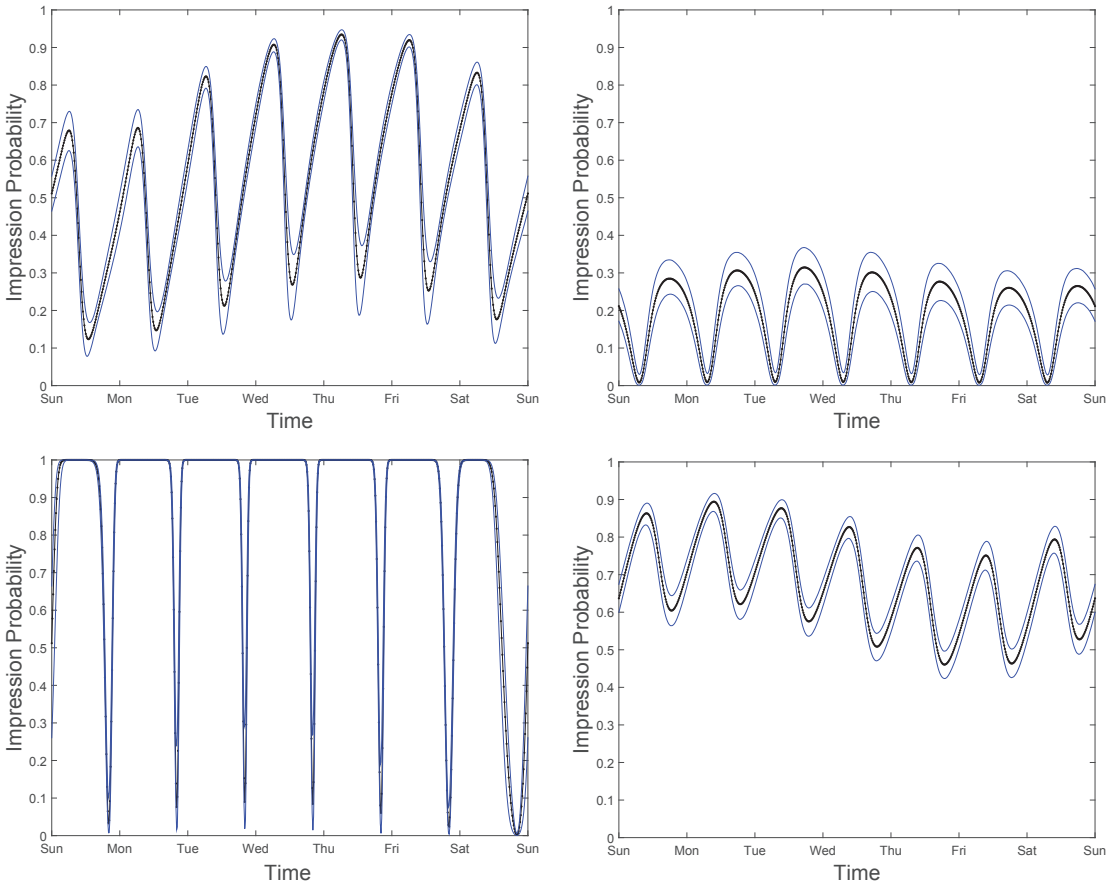

Figure 10 **Impression probability plots for Twitter users with screen names (top left) @444sai, (top right) @AbdeRules, (bottom left) @AsifSKhan, and (bottom right) @Birdseye1**

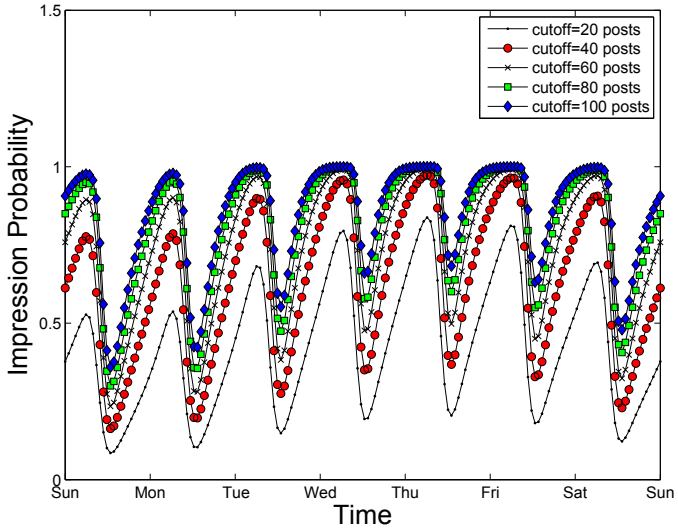

Figure 11 **Impression probability plots for different values of the cutoff window for user @444sai.**

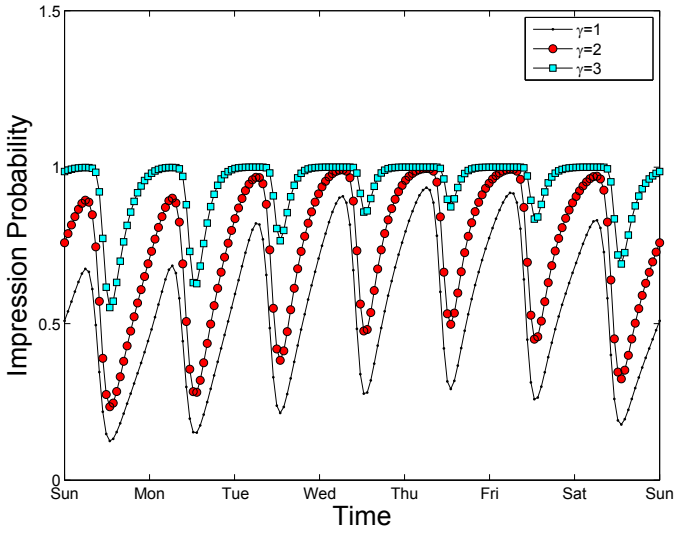

Figure 12 Impression probability plots for different values of γ for user @444sai.

4.2. Robustness to Using the Posting Process Instead of the Arrival Process

Of concern in our analysis is the fact that we do not have the actual arrival times, but instead only the posting times. We saw in Section 2.5 that this would not produce an appreciable change in the achieved impression probability. We now show that this is the case with real data. We scale up the posting rate by a factor $\gamma \geq 1$ and calculate the resulting impression probability. The results for a Twitter user, shown in Figure 12, show that by changing γ we are not affecting the location of the times for maximum and minimum impression probability. Therefore, we can still achieve timing gains using the posting rate as a surrogate for the arrival rate.

4.3. Accuracy of Approximations

We next look at how good the two approximations for the impression probability that were presented in Section 2 are on real data. We plot in Figure 13 the impression probability calculated using the exact equation ($q(t)$ from equation (3)), and the approximations $q_1(t)$ from equation (5), $q_2(t)$ from equation (7),and $q_3(t)$ defined as

$$q_3(t) = \min\left\{q_2(t), 1\right\}. \tag{23}$$

$q_3(t)$ is simply truncating $q_2(t)$ to obtain a proper probability. We assume a cutoff value of 30. As can be seen from the figure, approximation $q_1(t)$ is a very good approximation to the true impression probability. In contrast, $q_2(t)$ and $q_3(t)$ do not perform as well for large values of the impression probability, but are good for smaller values. In fact, $q_2(t)$ many times exceeds one, but does follow the same shape as $q_t(t)$.

We look at the quality of the approximations over a set of 94 Twitter users. We calculate the impression probability at n values of t over a time interval of one week. In our calculations $n = 172$ which corresponds

Approximation	MAPE
$q_1(t)$	13%
$q_2(t)$	3242%
$q_3(t)$	25%

Table 4 Mean absolute percent error (MAPE) of impression probability approximations averaged over 94 Twitter users.

to calculating the impression probability at one hour intervals. We define the mean absolute percent error (MAPE) for approximation k of each user u as

$$\text{MAPE}_k^u = \frac{1}{n}\sum_{i=1}^{n}\frac{|q^u(t_i) - q_k^u(t_i)|}{q^u(t_i)}, \qquad k = 1, 2, 3$$

and the mean absolute percent error of approximation k as MAPE_k^u averaged over all users. We show the values for these errors for the approximations in Table 4.3. As can be seen, the MAPE is below 25% for $q_1(t)$ and $q_3(t)$, and for approximation $q_1(t)$ it is only 13%. The MAPE is very high for $q_2(t)$ because many times the value of the approximation is greater than one, but $q_2(t)$ follows the general shape of the exact impression probability, as can be seen in Figure 13. In practice a user may only want to know the best time to post. Therefore, the actual value of the impression probability is not needed, but just the location of its maximum. Therfore, despite its huge MAPE, $q_2(t)$ is still a useful expression if one does not need the exact value of the impressio probability. This is what is shown in Theorem 1. When more accurate values of the impression probability are required, approximations $q_1(t)$ or $q_3(t)$ can be used in place of the exact expression for the impression probability.

5. Optimal versus Real Posting Times

With our model we are able to now evaluate the impression probability for arbitrary social media users. We saw there can be substantial gain in impressions for strategic timing versus randomly posting. A natural next question to ask is what is the potential gain in impressions from strategic timing compared to what is currently done by real social media users. To investigate this we now consider several high-profile Twitter users that are celebrities, politicians, or media programs and have millions of followers. Due to data limitations we are only able to collect the posting and timeline times for approximately 100 followers per user. We wish to investigate how the real posting times of these users compare with the optimal posting times for our sample of their followers in terms of average impressions. This will provide a sense of how much these users could increase their impressions if strategic timing was used.

5.1. Data analysis

We first analyze the actual posting times of the high-profile users in order to better understand how they are currently posting. In particular, we wish to understand the distribution of the days of the week and hours of the day of their posts. We plot if Figure 14 histograms of the hour and day of the posting times for four

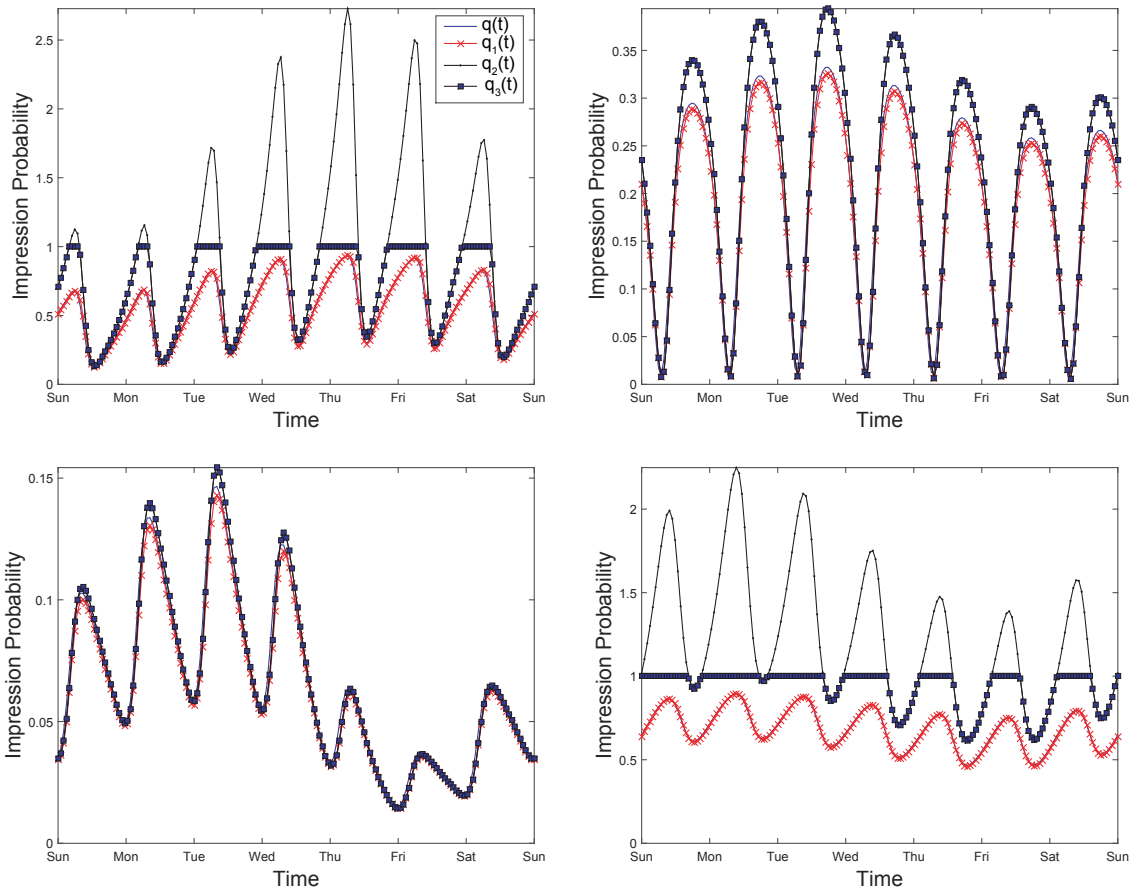

Figure 13 **Plots of the impression probability ($q(t)$) and approximations ($q_1(t)$, $q_2(t)$, and $q_3(t)$) for Twitter users with screen names (top left) @444sai, (top right) @AbdeRules, (bottom left) @AZ5566, and (bottom right) @Birdseye1.**

different users. These histograms represent the main patterns seen in the dataset and most likely correspond to different time zones. The Pacific, Eastern, and Central time zones in the United States are represented in our sample. We also have a user (@Cristiano) located in Madrid. We indicate the time zones listed by Twitter of these users in the figure. As can be seen, the posting patterns do not seem totally random when it comes to the hour of the post. Rather, it seems that the posting times correspond to the daytime hours in the different time zones. However, there is a more uniform distribution of the posts over days of the week, with an exception being @Oprah who posts heavily on Sunday. To quantify this we calculate the entropy of the empirical distribution of the hours and days of the posts for each user. For a discrete probability mass function $p(k), k \in \{1, 2, ..., n\}, \sum_{k=1}^{n}, p(k) = 1$, the entropy is defined as $H(p) = -\sum_{k=1}^{n} p(k) \log(p(k))$. The entropy is zero when there is no randomness, which occurs if all the probability mass concentrates on a single element. The entropy achieves its maximum value of $\log(n)$ if there is maximum randomness, which occurs if the probability mass is uniformly distributed across all elements. We plot in Figure 15 the entropy

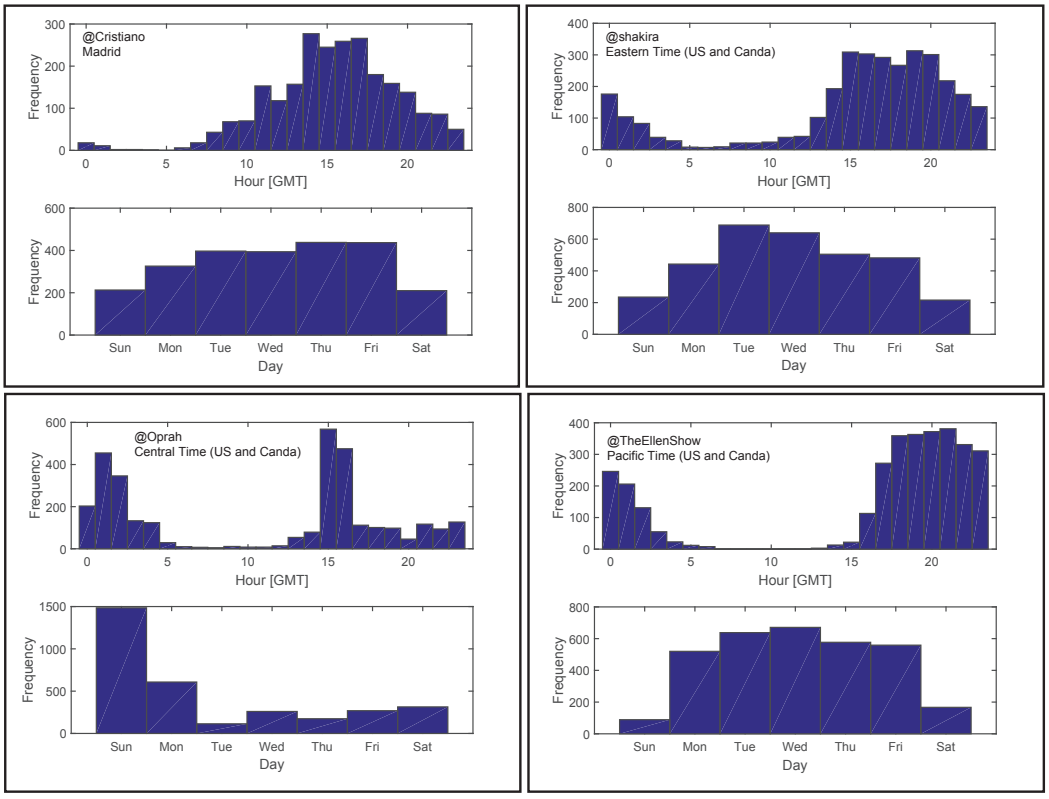

Figure 14 **Histogram of hour and day of posting times for four different Twitter users. The screen names and time zones of the users are (top left) @Cristiano, Madrid, (top right) @shakira, Eastern Time (US and Canada), (bottom left) @Oprah Central Time (US and Canada), and (bottom right) @TheEllenShow, Pacific Time (US and Canada).**

corresponding to the hour and day distributions for the users. The maximum entropy is also indicated in the figure. As can be seen, the day distributions are very close to maximum randomness, while the hour distributions appear to be much less random. The conclusion from this is that users are not strongly biased towards any day of the week when they post, but do have significant biases as to the hour when they post, most likely because they are awake at certain hours.

5.2. Impression Gain

We next wish to see how many impressions these users would gain with respect to our set of their sampled followers if they had posted optimally. We conduct the following simulation. We assume that each post of the user must be constrained to the day it was posted, but there is total flexibility as to what time in that day to post. This is generally true for most posts not involving time sensitive information. The only exception would be posts about a breaking news story. Other than those posts, most posts generally have information pertinent to the current day, but the exact timing within the day is not critical.

For each user in our dataset we estimate the timeline and arrival process parameters for all of their sampled followers using our Bayesian model which incorporates hourly and daily periodicity. We then use these

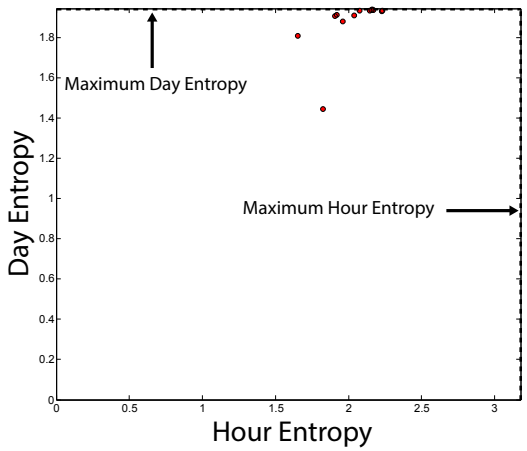

Figure 15 **Plots of the day and hour entropy of several Twitter users. The maximum day and hour entropies are indicated in the figure.**

estimated parameters to evaluate the impression probability for each follower. We assume a cutoff window of 30 posts in this analysis. We then take each post by the user and evaluate its impression probability for each of the user's sampled followers. We use this to find the optimal posting time for each post within its posting day for each follower. This allows us to see how much gain in impressions could have been achieved within the day for each post.

We plot in Figure 16 the impression probability at the actual and optimal posting times for different followers of user @Cristiano. The times have all been mapped into a one week window because our model has a one week periodicity in the timeline and arrival rates. Each day has its own optimal posting time and from the figure it can be seen how close each actual post is to the optimal posting time for each day. We also show the aggregate average impressions and optimal times for all sampled followers of the user. The aggregate average impressions are obtained by adding up each follower's impression probability curve. As can be seen, the gain obtained in aggregate is typically lower than the individual gain for a follower. This is mainly due to the heterogeneity in the follower's impression probabilities reducing the temporal variations when the impression probabilities are aggregated. However, these users have millions of followers, so even if the gain is only a few percent, this can still result in a significant number of extra impressions. In addition, aggregate average impressions may not always be the metric of interest. For instance, there may be specific high impact followers with large influence that a user is targeting with their posts. For these followers, the gain of strategic timing can be very large.

To quantify the gain at the individual follower level, we define the *post gain* as the percent increase in the average impressions of the follower for the optimal posting time in the day versus the actual posting time for a post. We define the *impression gain per follower* as the average of the post gains over all posts of the user for a specific follower. We also define the *aggregate impression gain* as the average of the post gain for the aggregate average impressions curve. We plot the two types of gains for each user in Figure 17. As

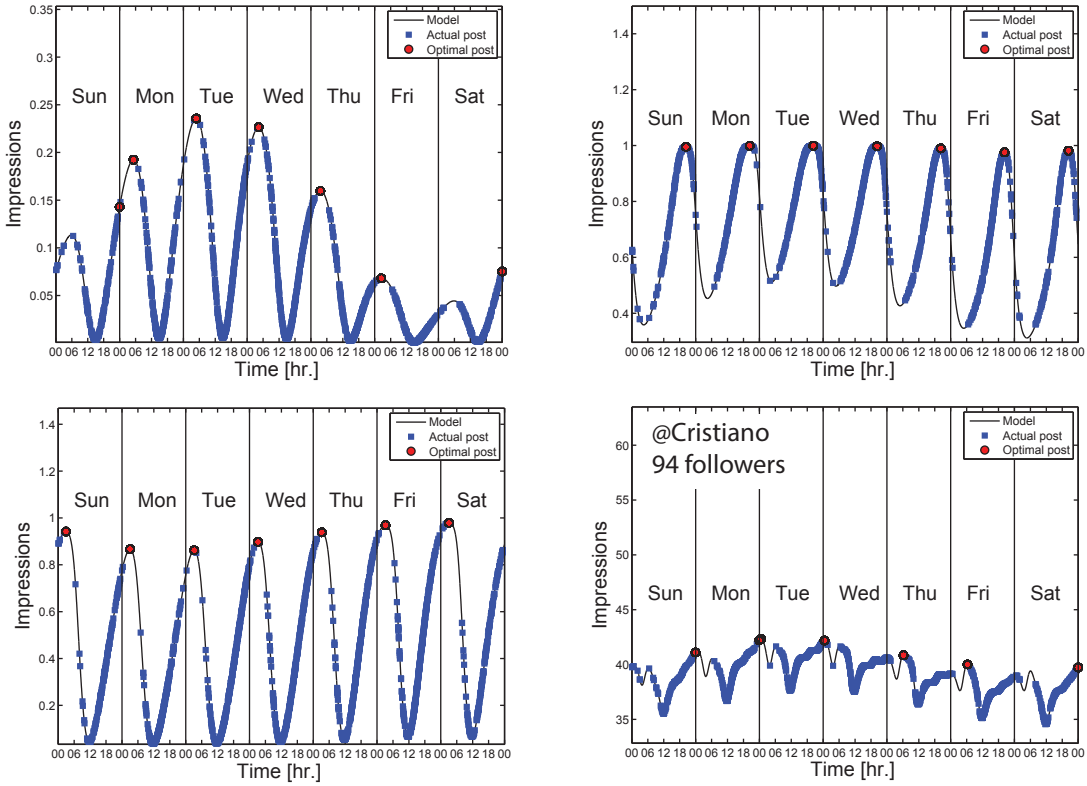

Figure 16 **(top left, top right, bottom left) The impression probability for three different followers of @Cristiano and (bottom right) the aggregate average impressions of 94 followers of @Cristiano. Each figure shows the impression probability/aggregate impressions curve given by the model, along with the values at the times of the actual posts of @Cristiano and at the optimal posting time for each day of the week.**

can be seen, the aggregate impression gain is typically lower than the impression gain per follower, but still not insignificant. The median impression gain per follower for a user ranges from 32 to 223%, while the aggregate impression gain ranges from 2 to 43%. This suggests that these users could significantly increase their impressions for select target followers or in aggregate by using strategic timing of their posts.

6. Optimizing the Timing of Content for Multiple Users

We have seen that it is possible for an individual user to optimize their posting time to target any individual or set of followers. This raises the question: what will happen if many users begin optimizing their posting times. Given the prevalence of scheduling applications for Twitter (Social Times 2010) it is easy to imagine an application that also calculates the optimal posting times for all of its users. For such an application to be successful it must try to provide favorable posting times for all its users. However, if there are common followers of these users there will likely be a trade-off between the aggregate performance of the system and its fairness, that is between the average (or sum) and the minimum of the impression probabilities over all users. Maximizing either of these metrics already leads to a potentially intractable global optimization

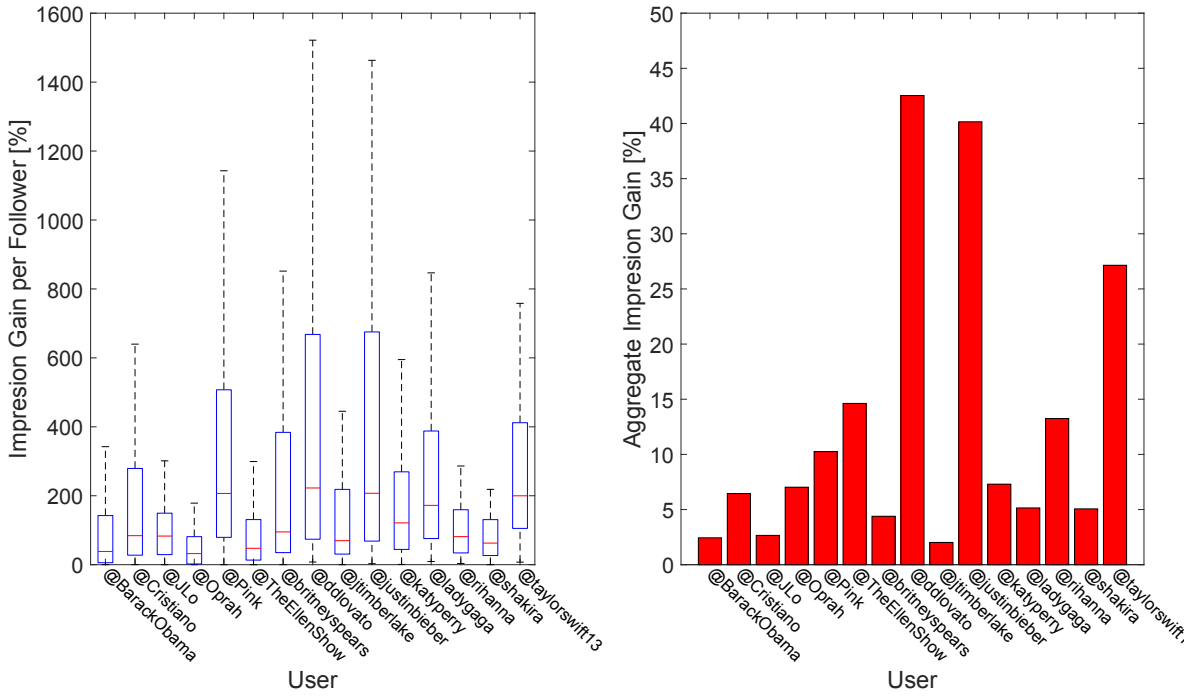

Figure 17 **(left) Boxplot of the impression gain per follower and (right) plot of the aggregate impression gain for different users.**

problem. However, in this section we show how to construct a solution that is nearly optimal for the aggregate performance and is naturally fair. In fact, we will provide some evidence that suggest that improving fairness actually improves aggregate performance.

Assume there is a set $\mathcal{U}$ of users who want to optimize their posting times and let $|\mathcal{U}| = U$. Each user $u \in \mathcal{U}$ has a set of followers $\mathcal{F}_u$ and let the set of all followers of users in $\mathcal{U}$ be $\mathcal{F} = \bigcup_{u\in\mathcal{U}} \mathcal{F}_u$, with $|\mathcal{F}| = F$. Each follower $f \in \mathcal{F}$ has a set of users it follows (its friends) denoted by $\mathcal{U}_f$. We assume that all friends of the followers are using the application to schedule their posts. That is, $\mathcal{U} = \bigcup_{f\in\mathcal{F}} \mathcal{U}_f$. This way, the application completely determines the timelines of the followers.

Assume at first that the main goal of the application is to maximize the average total impressions of all followers. To simplify the analysis and more closely model how the application would operate in practice, we assume that each user will post once during a fixed period. For simplicity we will assume that the period T is one day. We define the posting time of a user u as $t_u \in [0, T]$ and assume that this will be the time that the user will post his content, i.e., the posting times of user u are given by $t_u + kT$ for $k = 1, 2, ...$ Given a set of posting times $\mathbf{t} = \{t_1, t_2, ..., t_U\}$, we define the impression probability of content posted at time t for follower f as $q_f(t|\mathbf{t})$. With this notation, the optimization problem can be written as

$$\begin{aligned} \underset{\mathbf{t}\in[0,T]^U}{\text{maximize}} \quad & \sum_{u\in\mathcal{U}} \sum_{f\in\mathcal{F}_u} q_f(t_u|\mathbf{t}) \\ \text{subject to} \quad & 0 \le t_u \le T,\ u = 1, \ldots, U. \end{aligned} \tag{24}$$

The objective function of this problem may be non-concave so it is unclear if an optimal solution can easily be found. To gain insight into the solution, we consider a simpler version where there is one follower who follows all users in $\mathcal{U}$. We will find that by studying this problem we will gain insights into the solution of the full optimization problem. We assume that the cutoff window of the follower is C and that $C < U$. If the cutoff window is larger than U, then all posts will be seen by the follower and the problem is trivial. We still assume that the follower's timeline consists of only tweets from the set of U users. The follower's arrival rate $\psi(t)$ is known and is periodic with period T. We define the amount of time until there are C posts on the follower's timeline after time t as $\Delta(t|C, \mathbf{t})$. To simplify our analysis, we assume the posting times of the users are ordered such that $t_u \leq t_{u+1}, 1 \leq u \leq U-1$. Because we know the times of each post exactly, each post lifetime can easily be calculated. For time t_u, this time is given by

$$\Delta(t_u|C, \mathbf{t}) = t_{u+C} - t_u \tag{25}$$

Above we have used the periodicity of the posting times so that $t_{u+U} = T + t_u$. The follower arrival process is modeled as a non-homogeneous Poisson process with rate $\psi(t)$. The post of user u is seen if the follower arrives (checks his timeline) at least once during the interval $[t_u, t_{u+C})$. It can be shown that the exact impression probability of user u's post is then given by

$$\begin{aligned} q(t_u|\mathbf{t}) &= 1 - e^{-\Psi(t_u, t_u + \Delta(t_u|C,\mathbf{t}))} \\ &= 1 - e^{-\Psi\left(t_u, t_{u+C}\right)}, \end{aligned} \tag{26}$$

where we define the mean of the arrival process as $\Psi(t, t+s) = \int_t^{t+s} \psi(x)dx$.

For this simplified setting of one follower, we consider two different scenarios. In one scenario, the application will use historic data to calculate the posting time of each user independently. In the other scenario, the application will try to jointly optimize the posting times by solving the optimization problem in equation (24). We will next examine what happens in these two scenarios and how to schedule the posts for the joint optimization.

6.1. Multiple Users Optimizing Individually

Consider a scenario where during each period the application selects the posting time by individually optimizing each user's impression probability. There is a single follower so all users will see the same impression probability function. Therefore, if they each choose their posting times to maximize this function and its global maximum is unique, they will end up posting at the same time t. Then the set of posting times is $\mathbf{t}^{ind} = \{t, t, ..., t\}$. When all users post at the same time we can calculate the average total impressions achieved. There are two types of users in this scenario. Since all users post at the same time, only the final C posts are seen and the first $U - C$ posts will not be seen. We assume that the ordering of the U posts

that occur at the same time is given by a permutation chosen uniformly at random. Then the impression probability for any user posting at t is

$$\begin{aligned} q(t|\mathbf{t}^{ind}) &= \frac{C}{U}(1-e^{-\Psi(t,t+T)}) \\ &= \frac{C}{U}(1-e^{-\Psi(0,T)}). \end{aligned} \tag{27}$$

Above we used the fact that $\psi(t)$ is periodic over an interval T. We define the total average impressions of a set of posting times $\mathbf{t} = \{t_1, t_2, ..., t_U\}$ as $I(\mathbf{t}) = \sum_{i=1}^{U} q(t_i|\mathbf{t})$. The average total impressions for individual optimization is given by

$$\begin{aligned} I(\mathbf{t}^{ind}) &= Fq(t|\mathbf{t}^{ind}) \\ &= C\left(1-e^{-\Psi(0,T)}\right). \end{aligned} \tag{28}$$

We see here that each user has the same impression probability if the ordering is chosen randomly. However, for a given slot, the impression probabilities are not equal. The final C posts will have $q(t|\mathbf{t}^{ind}) = 1 - e^{-\Psi(t,t+T)}$, while for the first $F - C$ posts $q(t|\mathbf{t}^{ind}) = 0$. Therefore, each day some users are guaranteed to not have their post seen, resulting in an unfair allocation of the possible impressions. If the impression probability has multiple global maxima (or multiple local maxima that are nearly global maximum) that are sufficiently spread out the unfairness may reduced by spreading the posts among these maxima. However, this approach will fail for large F or small C. Fortunately, as we show next, the unfairness can always be eliminated if we spread the posts in a way that adapts to the shape of the follower's arrival rate. In addition, this approach actually leads to a near optimal solution with respect to total average impressions.

6.2. Multiple Users Optimizing Jointly

We now consider the situation where all posting times are jointly optimized. It is difficult to solve the optimization problem exactly. Instead, here we will propose a solution and show its good properties in terms of the total impressions it achieves and fairness across users. We have the following result.

THEOREM 6.1. Consider the optimization problem (24) with $F = 1$ and $C < U$. Let the optimal solution be $\mathbf{t}^*$. Define the distribution function corresponding to follower arrival rate $\psi(t)$ as

$$G(t) = \frac{\Psi(0,t)}{\Psi(0,T)}, \tag{29}$$

and define the set of posting times $\mathbf{t}^{\psi} = \left\{G^{-1}\left(\frac{u}{U}\right)\right\}_{u=1}^{U}$. Then the impression probability of user u is

$$q(t_u^{\psi}) = 1 - e^{-\frac{C}{U}\Psi(0,T)}, \quad u = 1, 2, ..., U \tag{30}$$

the average total impressions is

$$I(\mathbf{t}^{\psi}) = U\left(1-e^{-\frac{C}{U}\Psi(0,T)}\right), \tag{31}$$

and we have that

$$|I(\mathbf{t}^*) - I(\mathbf{t}^\psi)| \leq \begin{cases} \frac{C^2\Psi^2(0,T)}{U}\left(\frac{1}{2} - \frac{1}{e}\right), & \Psi(0,T) \leq 1 \\ \frac{C^2\Psi(0,T)}{U}\left(\frac{\Psi(0,T)}{2} - \frac{1}{e}\right), & \Psi(0,T) > 1 \end{cases}. \tag{32}$$

This result shows that if we choose the posting times as the quantiles of the distribution derived from the arrival rate $\psi(t)$, then we will come very close to maximizing the average total impressions and also, we will give each slot an equal impression probability. Therefore, we achieve fairness in this allocation of posting times, unlike when users optimize independently. This fairness is not surprising given the result from Section 2.3 which showed that when the timeline and arrival rates are proportional, the impression probability is time independent. In this case the solution creates a timeline rate that tries to match the arrival rate.

If we compare the average total impressions from the approximate solution to the joint optimization ($I(\mathbf{t}^\psi)$ from equation (31)) to that from independently optimizing ($I(\mathbf{t}^{ind})$ from equation (28)), we find that their difference can be significant depending on the values of U, C, and $\Psi(0,T)$. We plot the ratio of the average total impressions from joint optimization to independent optimization versus different values of the parameters in Figure 18. We find that when $\Psi(0,T)$ is large and U is large, there is a large gain in average total views from joint optimization. This gain is much less for small $\Psi(0,T)$ and U. If we look at the asymptotic gain in average total views as a function of the number of users posting we find

$$\begin{aligned} \lim_{U\to\infty} \frac{I(\mathbf{t}^\psi)}{I(\mathbf{t}^{ind})} &= \lim_{U\to\infty} \frac{U(1 - e^{-C/U\Psi(0,T)})}{C(1 - e^{-\Psi(0,T)})} \\ &\lim_{U\to\infty} \frac{(1 - e^{-C/U\Psi(0,T)})}{\frac{C}{U}(1 - e^{-\Psi(0,T)})} \\ &\lim_{U\to\infty} \frac{-\frac{C\Psi(0,T)}{U^2} e^{-C/U\Psi(0,T)}}{-\frac{C}{U^2}(1 - e^{-\Psi(0,T)})} \\ &= \frac{\Psi(0,T)}{1 - e^{-\Psi(0,T)}}. \end{aligned} \tag{33}$$

Thus, for a large number of users U, the gain from joint optimization is monotonically increasing in the average user arrivals in a period $\Psi(0,T)$ and approaches $\Psi(0,T)$ for larger values of $\Psi(0,T)$. On the other hand, for slowly arriving users (small $\Psi(0,T)$), there is very little gain from joint optimization.

6.3. Optimization Example: Sinusoidal Arrival Rate

We will now show a numerical example of our optimization problem. Based upon our modeling in Section 3 we will assume the user arrival rate is $\psi(t) = \psi_0\left(1 + \beta_h \cos\left(\omega_h\left(t - \tau_h\right)\right)\right)$, where we have defined $\omega_h = 2\pi$ when t is measured in days. We use the parameter values shown in Table 5 for the follower arrival rate which come from the posterior median values for a Twitter user. This optimization problem is already very hard to solve, but after one hour of computing time the global optimization solver Couenne (Belotti et al. 2009) was able to produce a solution that is guaranteed to be within 14% of the global optimum. We define

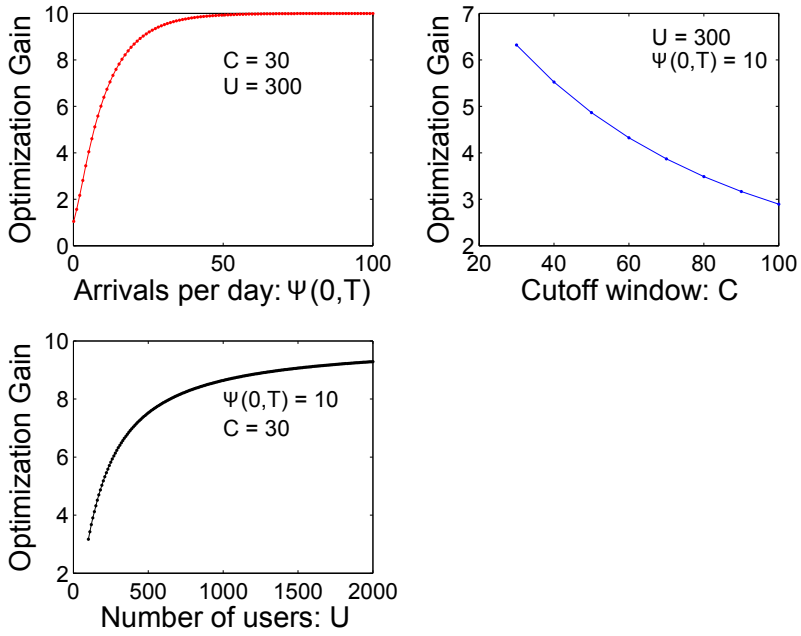

Figure 18 **Plot of the gain in average total impressions from joint optimization versus individual optimization of posting times $I(\mathbf{t}^{\psi})/I(\mathbf{t}^{ind})$.**

Parameter	Value
ψ_0	1 day^{-1}
T	1 Day
U	300
C	30
β_h	0.9 days
τ_h	0.5 days

Table 5 **Parameter values for optimization problem.**

the posting times for this solution as $\mathbf{t}^c = \{t_1^c, t_2^c, ..., t_U^c\}$. To visualize the solution, we first look at the impression probability for each user i as a function of his posting time in Figure 19. It can be seen that the impression probability of each user is almost constant. This aligns with the idea that maximizing the average total impressions tends to induce fairness in the impression probabilities, as suggested by Theorem 6.1.

Also from that theorem, we expect the posting times to be distributed according to the distribution function defined by $\psi(t)$ which is given by $G(t) = \Psi(0,t)/\Psi(0,T)$. We define the empirical distribution of the optimized posting times as $\widehat{G}(t) = \sum_{i=1}^{U} \mathbf{1}\{t_i^c \leq t\}$. We show a plot of the two functions in Figure 19 and see that they are nearly equal, as expected. Hence, selecting the posting times to match the user arrival rate $\psi(t)$ does seem like a good approach to produce high quality solutions and to induce fairness.

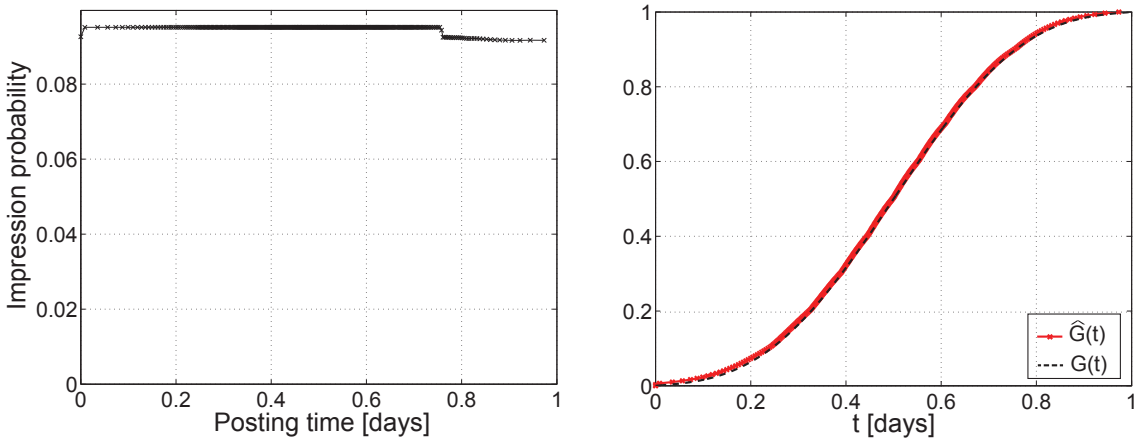

Figure 19 **(left) Plot of the impression probability of each user for the posting times found using Couenne. (right) Plot of the empirical CDF of the posting times obtained by Couenne $\widehat{G}(t)$ and the CDF $G(t) = \Psi(0,t)/\Psi(0,T)$ corresponding to the follower arrival rate $\psi(t)$.**

7. Discussion

We have shown here that strategic timing of content is an effective and simple method which can substantially increase the reach of posts in social networks. Our work presented a model for the use of social networks which allowed us to calculate impression probabilities for posts. This model is very general and applies to any social network where posts are shown to users in chronological order, such as Twitter and Instagram. We also presented a model for the dynamics of user behavior on social networks, which was used as an input to our social network use model. We were able to obtain impression probabilities as functions of the posting times for several real users in Twitter and found that strategic timing can greatly increase the impressions received by posts. In addition, we found the the optimal posting times were robust to uncertainty in several model parameters which are difficult to measure in practice, such as the arrival rate and cutoff window. We then extended our model to multiple users to show what happens when everyone in a social network posts content strategically. We showed that jointly optimizing posting times to maximize the average total impressions tends to induce a fair allocation of the impression probability for these users.

The most important aspect of this work is that it can be easily applied by companies trying to reach a specific audience in social networks such as Twitter which provide publicly accessible data. For Twitter users, determining the optimal time to post in order to reach a group of target followers can be done as follows.

1. Collect the posting times and timeline posting times for the target followers using the Twitter API.

2. Estimate the posting and timeline rate parameters for each follower using the Bayesian model in Section 3.

3. Use the estimated rate parameters and a reasonable value for the cutoff window to calculate the impression probability of each follower using equation (3), or one of the approximations given by equations (5) and (7).

4. Post at the time which maximizes the impression probabilities of the target followers.

All of these calculations are straightforward and can be done with publicly available data in Twitter and the results are robust to errors in several model parameters. Also, given that selecting the posting time incurs no extra cost, strategic timing can very easily be implemented using our approach.

One key assumption in our model is that posts are shown in the timeline in chronological order, with the most recent shown first. Our model does not apply to social networks where the order of posts is determined by an arbitrary ranking algorithm, such as in Facebook. In this case our model of the timeline process is no longer accurate. However, an interesting direction for future work is to incorporate different ranking algorithms into our timeline model. This way strategic timing could be extended to other social networks such as Facebook which do not rank posts chronologically.

Appendix A: Proof of Theorem 2.1

Define S_t as the random variable corresponding to the residual time until the next follower arrival in the arrival process conditioned on no arrival since time t. Because we model the arrival process as an non-homogeneous Poisson process, standard results (Cox and Isham 1980) show that we can write the cumulative distribution function (CDF) of S_t as

$$\begin{aligned} F_{S_t}(s) &= \mathbf{P}\left(S_t \leq s\right) \\ &= \mathbf{P}\left(M(t, t+s) > 0\right) \\ &= 1 - e^{-\Psi(t,t+s)}, \end{aligned} \tag{34}$$

and the corresponding density as

$$f_{S_t}(s) = \psi(t+s) e^{-\Psi(t,t+s)}. \tag{35}$$

Because the timeline process is a non-homogeneous Poisson process, we have that $N(t, t+s)$ is a Poisson random variable with mean value $\Lambda(t, t+s)$. For any fixed t and s, the post is seen if $N(t, t+s) \leq C$. Let us define the probability of this event as $q(t, t+s)$, which is given by the CDF of a Poisson random variable with mean value $\Lambda(t, t+s)$:

$$\begin{aligned} q(t, t+s) &= \mathbf{P}(N(t, t+s) \leq C) \\ &= e^{-\Lambda(t,t+s)} \sum_{k=0}^{C} \frac{\Lambda^k(t, t+s)}{k!}. \end{aligned}$$

The value s is determined by the time until the next arrival of the follower after time t, which is the random variable S_t. Therefore, to obtain the impression probability $q(t)$ we simply need to average $q(t, t+S_t)$ over S_t. This results in the following expression for the impression probability:

$$\begin{aligned} q(t) &= \mathbf{E}\left[q(t, t+S_t)\right] \\ &= \int_0^\infty q(t, t+s) f_{S_t}(s) ds \\ &= \sum_{k=0}^{C} \frac{1}{k!} \int_0^\infty \psi(t+s) \Lambda^k(t, t+s) e^{-(\Psi(t,t+s)+\Lambda(t,t+s))} ds. \end{aligned} \tag{36}$$

Appendix B: Proof of Lemma 1

The result is established by a direct application of equation (3).

$$\begin{aligned}
q(t) &= \sum_{k=0}^{C} \frac{1}{k!} \int_0^\infty \psi(t+s)\Lambda^k(t,t+s)e^{-(\Psi(t,t+s)+\Lambda(t,t+s))}ds \\
&= a\sum_{k=0}^{C} \frac{1}{k!} \int_0^\infty \lambda(t+s)e^{-(a+1)\Lambda(t,t+s)}\Lambda^k(t,t+s)ds \\
&= a\sum_{k=0}^{C} \frac{1}{k!(a+1)^{k+1}} \int_0^\infty e^{-(a+1)\Lambda(t,t+s)}(a+1)^k\Lambda^k(t,t+s)(a+1)\lambda(t+s)ds \\
&= \frac{a}{a+1}\sum_{k=0}^{C} \frac{1}{(a+1)^k} \\
&= \frac{a}{a+1}\frac{1-(a+1)^{-(C+1)}}{1-(a+1)^{-1}} \\
&= 1-(a+1)^{-(C+1)}.
\end{aligned}$$

Appendix C: Proof of Theorem 2.2

The number of arrivals in the timeline in the interval $(t, t+s]$ is a Poisson random variable $N(t,t+s)$ with mean value $\Lambda(t,t+s)$. Define the cumulative distribution function (CDF) of $N(t,t+s)$ as $F_{N(t,t+s)}(\cdot)$. Recall the definition of Δ as $\Delta = \sup\{s : \Lambda(t,t+s) \le C\}$. With this notation, the error of $q_1(t)$ is given by

$$\begin{aligned}
|q(t)-q_1(t)| &= \left| \int_0^\infty ds e^{-\Psi(t,t+s)}\psi(t+s)F_{N(t,t+s)}(C) - \int_0^\Delta ds e^{-\Psi(t,t+s)}\psi(t+s) \right| \\
&= \left| \int_0^\Delta ds e^{-\Psi(t,t+s)}\psi(t+s)\left(F_{N(t,t+s)}(C)-1\right) + \int_\Delta^\infty ds e^{-\Psi(t,t+s)}\psi(t+s)F_{N(t,t+s)}(C) \right| \\
&\le \int_0^\Delta ds e^{-\Psi(t,t+s)}\psi(t+s)\left(1-F_{N(t,t+s)}(C)\right) + \int_\Delta^\infty ds e^{-\Psi(t,t+s)}\psi(t+s)F_{N(t,t+s)}(C)
\end{aligned}$$

We will now upper bound $1-F_{N(t,t+s)}(C)$ for $s \le \Delta$ and $F_{N(t,t+s)}(C)$ for $s > \Delta$. The CDF of $N(t,t+s)$ is given by

$$F_{N(t,t+s}(C) = e^{-\Lambda(t,t+s)}\sum_{k=0}^{C}\frac{\Lambda^k(t,t+s)}{k!}. \tag{37}$$

For $s < \Delta$ we can upper bound $1-F_{N(t,t+s)}(C)$ with a straightforward application of the Chernoff bound. For any $\theta > 0$, we have

$$\begin{aligned}
1-F_{N(t,t+s)}(C) &= \mathbf{P}\left(N(t,t+s) \ge C\right) \\
&\le \mathbf{E}\left[e^{\theta N(t,t+s)}\right]e^{-\theta C} \\
&\le e^{\Lambda(t,t+s)\left(e^\theta-1\right)-\theta C}
\end{aligned}$$

The upper bound above is minimized for $\theta = \log(C/\Lambda(t,t+s))$ and the corresponding optimized upper bound is

$$1-F_{N(t,t+s)}(C) \le e^{C-\Lambda(t,t+s)}\left(\frac{\Lambda(t,t+s)}{C}\right)^C. \tag{38}$$

For $s > \Delta$ we can also upper bound $F_{N(t,t+s)}(C)$ with a straightforward application of the Chernoff bound. Using Markov's inequality we have that for any $\theta > 0$

$$\begin{aligned} F_{N(t,t+s)}(C) &= \mathbf{P}\left(N(t,t+s) \leq C\right) \\ &\leq \mathbf{E}\left[e^{-\theta N(t,t+s)}\right] e^{\theta C} \\ &\leq e^{\Lambda(t,t+s)\left(e^{-\theta}-1\right)+\theta C} \end{aligned}$$

The upper bound above is minimized for $\theta = \log(\Lambda(t,t+s)/C)$ and the corresponding optimized upper bound is

$$F_{N(t,t+s)}(C) \leq e^{C-\Lambda}\left(\frac{\Lambda}{C}\right)^{C}. \tag{39}$$

To simplify the next step in the analysis, we will use Stirling's approximation for the factorial ($n! \geq \sqrt{2\pi} n^{1/2+n} e^{-n}$) to obtain

$$e^{C-\Lambda}\left(\frac{\Lambda}{C}\right)^{C} \leq \frac{e^{-\Lambda}\Lambda^{C}\sqrt{2\pi C}}{C!}. \tag{40}$$

Next, we note that the conditions of the theorem give us that $F_1 < \lambda(t)/\psi(t)$ for some $1 < F_1$. Using this fact, along with the bounds on the tail probabilities of $N(t,t+s)$, we bound the error of $q_1(t)$.

$$\begin{aligned} |q(t) - q_1(t)| \leq & \int_0^{\Delta} ds e^{-\Psi(t,t+s)}\psi(t+s)\left(1 - F_{N(t,t+s)}(C)\right) + \int_{\Delta}^{\infty} ds e^{-\Psi(t,t+s)}\psi(t+s)F_{N(t,t+s)}(C) \\ \leq & \frac{\sqrt{2\pi C}}{C!}\int_0^{\Delta} ds e^{-\Psi(t,t+s)-\Lambda(t,t+s)}\psi(t+s)\Lambda^{C}(t,t+s) + \\ & \frac{\sqrt{2\pi C}}{C!}\int_{\Delta}^{\infty} ds e^{-\Psi(t,t+s)-\Lambda(t,t+s)}\psi(t+s)\Lambda^{C}(t,t+s) \\ \leq & \frac{\sqrt{2\pi C}}{C!}\int_0^{\infty} ds e^{-\Psi(t,t+s)-\Lambda(t,t+s)}\psi(t+s)\Lambda^{C}(t,t+s) \\ \leq & \frac{\sqrt{2\pi C}}{C!}\int_0^{\infty} ds e^{-\Lambda(t,t+s)}\frac{\lambda(t+s)}{F_1}\Lambda^{C}(t,t+s) \\ \leq & \frac{\sqrt{2\pi C}}{F_1}. \end{aligned} \tag{41}$$

Appendix D: Proof of Theorem 2.3

We begin with the following simple lemma regarding the exponential function.

LEMMA 2. *Let $f(x) = 1 - e^{-x}$. We can upper bound $f(x)$ by*

$$1 - e^{-x} \leq \begin{cases} x\left(1 - e^{-1}x\right), & x \leq 1 \\ x\left(1 - e^{-1}\right) & x > 1 \end{cases} \tag{42}$$

and lower bound $f(x)$ by

$$1 - e^{-x} \geq x - \frac{1}{2}x^2, \quad x \geq 0. \tag{43}$$

Recall that by the conditions of the Theorem, $\lambda(t)/F \leq \psi(t) \leq \lambda(t)/F_1$. Using this and the definition of Δ given by equation (6) we have that $C/F \leq \Psi(t,t+\Delta) \leq C/F_1$. Using this, we can upper bound the difference between $q_1(t)$ and $q_2(t)$ as

$$|q_1(t) - q_2(t)| \leq \left|1 - e^{-\Psi(t,t+\Delta)} - \Psi(t,t+\Delta) + \Psi(t,t+\Delta) - C\frac{\psi(t)}{\lambda(t)}\right|$$

$$
\begin{aligned}
&\leq \left|1 - e^{-\Psi(t,t+\Delta)} - \Psi(t,t+\Delta)\right| + \left|\Psi(t,t+\Delta) - C\frac{\psi(t)}{\lambda(t)}\right| \\
&\leq \frac{\Psi^2(t,t+\Delta)}{2} + \frac{C}{F_1} - \frac{C}{F} \\
&\leq \frac{1}{2}\left(\frac{C}{F_1}\right)^2 + C\frac{F-F_1}{FF_1} \\
&\leq \frac{C^2 + C(F-F_1)}{F_1^2}
\end{aligned}
$$

Above we have used Lemma 2 to bound $\left|1 - e^{\Psi(t,t+\Delta)} - \Psi(t,t+\Delta)\right|$. Combining this bound with equation (41) we obtain our final result,

$$
\begin{aligned}
|q(t) - q_2(t)| &\leq |q(t) - q_1(t)| + |q_1(t) - q_2(t)| \\
&\leq \frac{\sqrt{2\pi C}}{F_1} + \frac{C^2 + C(F-F_1)}{F_1^2}.
\end{aligned}
$$

Appendix E: Proof of Corollary 1

First, we note that by definition $q(t^*) \geq q(\widehat{t}_2)$ because t^* maximizes the true impression probability. Define $t_2^* = \arg\max_t q_2(t)$. Because $\widehat{q}_2(t) = \gamma q_2(t)$, we have that $\widehat{t}_2 = t_2^*$. This means that $q_2(\widehat{t}_2) \geq q_2(t^*)$. Consider the values $q(t^*)$, $q(\widehat{t}_2)$, $q_2(t^*)$, and $q_2(\widehat{t}_2)$. Given the previous inequalities, there are six possible ways these values can be ranked in decreasing order. Each possible permutation will give us a bound on $|q(t^*) - q(\widehat{t}_2)|$ which we will write using terms of the form $|q(t) - q_2(t)|$ in order to apply Theorem 2.3. We will obtain our final result by taking the maximum value of these bounds. We now enumerate each permutation and the bound it provides.

1. If $q(t^*) \geq q(\widehat{t}_2) \geq q_2(\widehat{t}_2) \geq q_2(t^*)$, then $|q(t^*) - q(\widehat{t}_2)| \leq |q(t^*) - q_2(t^*)|$.
2. If $q_2(\widehat{t}_2) \geq q(t^*) \geq q(\widehat{t}_2) \geq q_2(t^*)$, then $|q(t^*) - q(\widehat{t}_2)| \leq |q(t^*) - q_2(t^*)| + |q(\widehat{t}_2) - q_2(\widehat{t}_2)|$.
3. If $q_2(\widehat{t}_2) \geq q_2(t^*) \geq q(t^*) \geq q(\widehat{t}_2)$, then $|q(t^*) - q(\widehat{t}_2)| \leq |q(\widehat{t}_2) - q_2(\widehat{t}_2)|$.
4. If $q(t^*) \geq q_2(\widehat{t}_2) \geq q(\widehat{t}_2) \geq q_2(t^*)$, then $|q(t^*) - q(\widehat{t}_2)| \leq |q(t^*) - q_2(t^*)|$.
5. If $q_2(\widehat{t}_2) \geq q(t^*) \geq q_2(t^*) \geq q(\widehat{t}_2)$, then $|q(t^*) - q(\widehat{t}_2)| \leq |q(\widehat{t}_2) - q_2(\widehat{t}_2)|$.
6. If $q(t^*) \geq q_2(\widehat{t}_2) \geq q_2(t^*) \geq q(\widehat{t}_2)$, then $|q(t^*) - q(\widehat{t}_2)| \leq |q(t^*) - q_2(t^*)| + |q(\widehat{t}_2) - q_2(\widehat{t}_2)|$.

Applying Theorem 2.3 we have that

$$
\begin{aligned}
|q(t^*) - q(\widehat{t}_2)| &\leq |q(t^*) - q_2(t^*)| + |q(\widehat{t}_2) - q_2(\widehat{t}_2)| \\
&\leq 2\frac{\sqrt{2\pi C}}{F_1} + 2\frac{C^2 + C(F-F_1)}{F_1^2}.
\end{aligned}
$$

E.1. Proof of Corollary 2

The proof of this corollary is accomplished by applying the proof of Theorem 1 with C' replacing γ. All other steps of the proof remain the same once this substitution is made because the approximation $q_2(t)$ is proportional to both the cutoff window and the follower arrival rate.

Appendix F: Proof of Theorem 6.1

We start with the following Lemma.

LEMMA 1. Consider the optimization problem (24) with $C < U$ and define $\Psi(0,T) = \int_0^T \psi(t)dt$ as the average number of follower arrivals in $[0,T]$. Then we have the following upper and lower bounds for the average total impressions in $[0,T]$.

$$I(\mathbf{t}) \leq \begin{cases} C\Psi(0,T) - (eU)^{-1}(C\Psi(0,T))^2, & \Psi(0,T) \leq 1 \\ C\Psi(0,T) - (eU)^{-1}C^2\Psi(0,T) & \Psi(0,T) > 1 \end{cases}. \tag{44}$$

$$I(\mathbf{t}) \geq C\Psi(0,T) - \frac{C\Psi^2(0,T)}{2} \tag{45}$$

Define the distribution function derived from $\psi(t)$ as

$$G(t) = \frac{\Psi(0,t)}{\Psi(0,T)}, \quad t \in [0,T] \tag{46}$$

Then for $i = 1, 2, ..., U$ set $t_i = G^{-1}(i/U)$ and $\mathbf{t}^\psi = \{t_i\}_{i=1}^U$. We then find that

$$\begin{aligned} \Psi(t_i, t_{i+C}) &= \Psi(0, t_{i+C}) - \Psi(0, t_i) \\ &= \Psi(0,T)\left(G(t_{i+C}) - G(t_i)\right) \\ &= \Psi(0,T)\left(\frac{i+C}{U} - \frac{i}{U}\right) \\ &= \Psi(0,T)\frac{C}{U}. \end{aligned}$$

The impression probability of user i is given by

$$\begin{aligned} q(t_i|\mathbf{t}^\psi) &= 1 - e^{-\Psi(t_i, t_{i+C})} \\ &= 1 - e^{-\frac{C}{U}\Psi(0,T)} \end{aligned} \tag{47}$$

and the average total impressions will be

$$I(\mathbf{t}^\psi) = F\left(1 - e^{-\frac{C}{U}\Psi(0,T)}\right). \tag{48}$$

Using Lemma 2 we can lower bound this by

$$I(\mathbf{t}^\psi) \geq C\Psi(0,T) - \frac{(C\Psi(0,T))^2}{2U} \tag{49}$$

From Lemma 1 we know that the optimal average total views is upper bounded by

$$I(\mathbf{t}) \leq \begin{cases} C\Psi(0,T) - (eU)^{-1}(C\Psi(0,T))^2, & \Psi(0,T) \leq 1 \\ C\Psi(0,T) - (eU)^{-1}C^2\Psi(0,T) & \Psi(0,T) > 1 \end{cases}. \tag{50}$$

We also know that because $\mathbf{t}^*$ is optimal, $I(\mathbf{t}^\Psi) \leq I(\mathbf{t}^*)$. Therefore, we can bound the difference in these values by

$$|I(\mathbf{t}^*) - I(\mathbf{t}^\psi)| \leq \begin{cases} \frac{(C\Psi(0,T))^2}{U}\left(\frac{1}{2} - \frac{1}{e}\right), & \Psi(0,T) \leq 1 \\ \frac{C^2\Psi(0,T)}{U}\left(\frac{\Psi(0,T)}{2} - \frac{1}{e}\right), & \Psi(0,T) > 1. \end{cases} \tag{51}$$

Appendix G: Proof of Lemma 1

We start by using Lemma 2 from Section D to upper bound the impression probability by

$$q(t_i|\mathbf{t}) \leq \Psi(t_i, t_{i+C}) - a\Psi^2(t_i, t_{i+C}) \tag{52}$$

where we have define a as

$$a = \begin{cases} e^{-1}, & \Psi(0,T) \leq 1 \\ e^{-1}\Psi^{-1}(0,T) & \Psi(0,T) > 1 \end{cases}. \tag{53}$$

Above we have used the fact that $\Psi(t_i, t_{i+C}) \leq \Psi(0,T)$ because $C < U$ and $\psi(t)$ has period T.

Again using Lemma 2, the impression probability can be lower bounded by

$$q(t_i|\mathbf{t}) \geq \Psi\left(t_i, t_{i+C}\right) - \frac{1}{2}\Psi^2\left(t_i, t_{i+C}\right) \tag{54}$$

The objective function is then bounded by

$$\sum_{i=1}^{U} \Psi(t_i, t_{i+C}) - \frac{1}{2}\sum_{i=1}^{U} \Psi^2(t_i, t_{i+C}) \leq I(\mathbf{t}) \leq \sum_{i=1}^{U} \Psi(t_i, t_{i+C}) - a\sum_{i=1}^{U} \Psi^2(t_i, t_{i+C}). \tag{55}$$

The periodicity of the arrival rate and the posting times t_i means that $\Psi(t_i, t_{i+F}) = \Psi(0,T)$. Using this we can simplify the first summation that appears in equation (55) as

$$\begin{aligned} \sum_{i=1}^{U} \Psi(t_i, t_{i+C}) &= \sum_{i=1}^{U}\sum_{j=1}^{C} \Psi(t_{i+j-1}, t_{i+j}) \\ &= \sum_{j=1}^{C}\sum_{i=1}^{U} \Psi(t_{i+j-1}, t_{i+j}) \\ &= \sum_{j=1}^{C} \Psi(t_j, t_{j+U}) \\ &= C\Psi(0,T). \end{aligned} \tag{56}$$

We next lower bound the second sum that appears in equation (55) as

$$\begin{aligned} \sum_{i=1}^{U} \Psi^2\left(t_i, t_{i+C}\right) &\geq \frac{C^2}{U^2}\sum_{i=1}^{U} \Psi^2\left(0,T\right) \\ &\geq \frac{C^2}{U}\Psi^2(0,T). \end{aligned} \tag{57}$$

Above we have used the fact that we want to minimize the square of the l^2-norm of the vector $\left[\Psi\left(t_1, t_{1+C}\right), \Psi\left(t_2, t_{2+C}\right), ..., \Psi\left(t_U, t_{U+C}\right)\right]$ when we have fixed its l^1-norm in equation (56). In this situation it is well known that the minimum is achieved by the vector with all elements equal. In our case, this would mean that $\Psi\left(t_i, t_{i+C}\right) = C\Psi(0,T)/U$ for $i = 1, 2, ..., U$.

We next need to upper bound this sum. It is known that the upper bound of the square of the l^2-norm of a vector when we have fixed its l^1-norm is achieved by the vector with all the mass in one entry. However, such a vector is not achievable in our situation because $\Psi\left(t_i, t_{i+C}\right) = \sum_{j=1}^{C} \Psi\left(t_{i+j-1}, t_{i+j}\right)$ and $\sum_{i=1}^{U} \Psi\left(t_i, t_{i+1}\right) = \Psi(0,T) > 0$. Therefore, there must be at least C adjacent elements of the vector $\left[\Psi\left(t_1, t_{1+C}\right), \Psi\left(t_2, t_{2+C}\right), ..., \Psi\left(t_U, t_{U+C}\right)\right]$ which are non-zero. With this constraint, the vector with $\Psi(t_i, t_{i+1}) = 0$ for $1 \leq i \leq U-1$ and $\Psi(t_{U-1}, t_U) = \Psi(0,T)$

achieves the upper bound. For this vector, we will have that $\Psi(t_i, t_{i+C}) = 0$ for $1 \leq i \leq U - C$ and $\Psi(t_i, t_{i+C}) = \Psi(0, T)$ for $U - C \leq i \leq U$. With this vector, the upper bound is

$$\begin{aligned}\sum_{i=1}^{U} \Psi^2 (t_i, t_{i+C}) &\leq \sum_{i=1}^{C} \Psi^2 (0, T) \\ &\leq C\Psi^2 (0, T) .\end{aligned}$$

Combining the expressions for the two summations we obtain

$$C\Psi(0,T) - \frac{C}{2}\Psi^2(0,T) \leq I(\mathbf{t}) \leq C\Psi(0,T) - a\frac{C^2}{U}\Psi^2(0,T), \tag{58}$$

with a given by equation (53). This holds for any feasible $\mathbf{t}$ and therefore holds for the optimal value $\mathbf{t}^*$.

Appendix H: Proof of Lemma 2

The function $f(x) = 1 - e^{-x}$ is monotonically increasing in x. For $x > 1$ we have

$$\begin{aligned}x &> 1 \\ &> \frac{1 - e^{-x}}{1 - e^{-1}}\end{aligned}$$

Rearranging this expression provides the upper bound for $x > 1$. For $x \leq 1$ we have that

$$\begin{aligned}1 - e^{-x} &\leq \frac{x}{2} \\ &\leq x - x^2 e^{-1}.\end{aligned} \tag{59}$$

Inequality (59) is established by noting that for $0 \leq x \leq 1$, $x/2 - x^2 e^{-1}$ is 0 for $x = 0$, positive for $x = 1$, and achieves its unique maximum value of $e/16$ at $x = e/4 < 1$. The lower bound is obtained using Taylor's theorem applied to $1 - e^{-x}$, which states that for $x \geq 0$, there exists a $\xi \in [0, x]$ such that

$$\begin{aligned}1 - e^{-x} &= x - \frac{x^2}{2} + \frac{e^{-\xi}}{3!}x^3 \\ &\geq x - \frac{x^2}{2}\end{aligned}$$

Appendix I: Details of MCMC Sampler

We use a Metropolis-within-Gibbs scheme to sample from the posterior distribution of the model parameters. We assume we have timeline and posting data for N users. We begin by establishing our notation. We define the set of timeline model parameters for a user u as $\Theta^u = \{\lambda_0^u, b_h^u, b_d^u, t_h^u, t_d^u\}$. The set of posting model parameters for user u is given by $\Phi^u = \{\psi_0^u, \beta_h^u, \beta_d^u, \tau_h^u, \tau_d^u\}$. We define the set of posting model parameters for all users as $\mathbf{\Phi}^{user} = \{\Phi^1, \Phi^2, ..., \Phi^N\}$. The set of global parameters is given by $\Phi = \{\psi_0, \beta_h, \beta_d, \tau_h, \tau_d, \sigma_{\beta_h}, \sigma_{\beta_d}, \sigma_{\tau_h}, \sigma_{\tau_d}\}$. For each user u we observe the corresponding m^u posting times $\mathbf{s}^u = \{s_1^u, s_2^u, ...s_{m^u}^u\}$ and n^u timeline times $\mathbf{t}^u = \{t_1^u, t_2^u, ...t_{n^u}^u\}$. We define the set of posting times of all users as $\mathbf{S} = \{\mathbf{s}^1, \mathbf{s}^2, ..., \mathbf{s}^N\}$. Given a set of parameters A, for any parameter $\gamma \in A$ we define $A_{-\gamma} = A/\gamma$ (the set A excluding γ).

I.1. MCMC Sampler for Timeline Model

For our MCMC sampler for the timeline model, we can sample the parameters of each user independently. To implement our Metropolis-within-Gibbs scheme we must sample from the conditional distribution $\mathbf{P}(\gamma|\mathbf{t}^u, \Theta^u_{-\gamma})$ for every timeline model parameter for every user. We will now derive these conditional distributions and show how to sample from them.

Parameter λ_0^u. The prior distribution for λ_0^u is a gamma distribution with shape and scale one and 10,000. The conditional distribution of λ_0^u is given by

$$\begin{aligned}\mathbf{P}(\lambda_0^u|\mathbf{t}^u, \Theta^u_{-\lambda_0^u}) \propto& \mathbf{P}(\mathbf{t}^u|\Theta^u)\mathbf{P}(\lambda_0^u) \\ \propto& \exp\left(-\lambda_0^u(10^{-4}+\xi)\right)\left(\lambda_0^u\right)^{n^u},\end{aligned}$$

where we define ξ as

$$\begin{aligned}\xi =& \frac{\Lambda^u(t_1^u, t_{n^u}^u)}{\lambda_0^u} \\ =& t_{n^u}^u - t_1^u + \frac{b_h^u}{\omega_h}\sin\left(\omega_h(r-t_h^u)\right)\Big|_{r=t_1^u}^{t_{n^u}^u} + \frac{b_d^u}{\omega_d}\sin\left(\omega_d(r-t_d^u)\right)\Big|_{r=t_1^u}^{t_{n^u}^u} + \\ & \frac{b_h^u b_d^u}{2(\omega_d+\omega_h)}\sin\left((\omega_d+\omega_h)(r-(t_h^u+t_d^u))\right)\Big|_{r=t_1^u}^{t_{n^u}^u} + \frac{b_h^u b_d^u}{2(\omega_d-\omega_h)}\sin\left(\omega_d-\omega_h(r-(t_h^u-t_d^u))\right)\Big|_{r=t_1^u}^{t_{n^u}^u}.\end{aligned}$$

We see from this that the conditional distribution of λ_0^u is again a gamma distribution with shape and scale parameters $1+n^u$ and $1/(10^{-4}+\xi)$. Therefore we can directly sample λ_0^u.

Parameters b_h^u, b_d^u,t_h^u, *and* t_d^u. The prior distribution for the parameters b_h^u, b_d^u,t_h^u, and t_d^u are normal with zero mean and standard deviation of 100. For b_d^u the conditional distribution is given by

$$\begin{aligned}\mathbf{P}(b_h^u|\mathbf{t}^u, \Theta^u_{-b_h^u}) \propto& \mathbf{P}(\mathbf{t}^u|\Theta^u)\mathbf{P}(b_h^u) \\ \propto& \left(\Lambda^u(t_1^u, t_{n^u}^u)\right)^{n^u}\exp\left(-\Lambda^u(t_1^u, t_{n^u}^u) - \frac{(b_h^u)^2}{2(100^2)}\right).\end{aligned}$$

For b_d^u the conditional distribution is given by

$$\begin{aligned}\mathbf{P}(b_d^u|\mathbf{t}^u, \Theta^u_{-b_d^u}) \propto& \mathbf{P}(\mathbf{t}^u|\Theta^u)\mathbf{P}(b_d^u) \\ \propto& \left(\Lambda^u(t_1^u, t_{n^u}^u)\right)^{n^u}\exp\left(-\Lambda^u(t_1^u, t_{n^u}^u) - \frac{(b_d^u)^2}{2(100^2)}\right).\end{aligned}$$

For t_d^u the conditional distribution is given by

$$\begin{aligned}\mathbf{P}(t_h^u|\mathbf{t}^u, \Theta^u_{-t_h^u}) \propto& \mathbf{P}(\mathbf{t}^u|\Theta^u)\mathbf{P}(t_h^u) \\ \propto& \left(\Lambda^u(t_1^u, t_{n^u}^u)\right)^{n^u}\exp\left(-\Lambda^u(t_1^u, t_{n^u}^u) - \frac{(t_h^u)^2}{2(100^2)}\right).\end{aligned}$$

For t_d^u the conditional distribution is given by

$$\begin{aligned}\mathbf{P}(t_d^u|\mathbf{t}^u, \Theta^u_{-t_d^u}) \propto& \mathbf{P}(\mathbf{t}^u|\Theta^u)\mathbf{P}(t_d^u) \\ \propto& \left(\Lambda^u(t_1^u, t_{n^u}^u)\right)^{n^u}\exp\left(-\Lambda^u(t_1^u, t_{n^u}^u) - \frac{(t_d^u)^2}{2(100^2)}\right).\end{aligned}$$

To sample from the conditional distributions, we use a random walk Metropolis-Hastings step, which we now describe. We consider the parameter b_h^u. For b_h^u we define the ith sample as b_{hi}^u, and the proposal for the $(i+1)$ sample is drawn from a normal distribution with mean b_{hi}^u and standard deviation 0.1, where this value is chosen to balance the acceptance rate with step size. We use this exact same procedure to sample from the conditional distributions of b_d^u, t_h^u,and t_d^u, except that we use different values for the standard deviation. For b_d^u and t_h^u we choose a standard deviation of 0.1 and for t_d^u we choose a standard deviation of one.

I.2. MCMC Sampler for Posting Model

For our MCMC sampler for the posting model, we must sample the parameters of each user jointly because of the hierarchical structure of our model. For each user specific parameter γ^u we must sample from the conditional distribution $\mathbf{P}(\gamma^u|\mathbf{S},\Phi,\mathbf{\Phi}^{user}_{-\gamma^u})$ and for each global parameter γ we must sample from the conditional distribution $\mathbf{P}(\gamma|\mathbf{S},\Phi_{-\gamma},\mathbf{\Phi}^{user})$. We will now derive these conditional distributions and show how to sample from them. We note that there are two types of parameters in this model: user specific parameters and global parameters.

Parameter ψ_0^u. The conditional distribution of ψ_0^u is given by

$$\begin{aligned}\mathbf{P}(\psi_0^u|\mathbf{S},\Phi,\mathbf{\Phi}^{user}_{-\psi_0^u}) \propto& \mathbf{P}(\mathbf{s}^u|\Phi^u)\mathbf{P}(\psi_0^u|\Phi)\\ \propto& \exp\left(-\psi_0^u\left(\frac{1}{\psi_0}+\zeta\right)\right)(\lambda_0^u)^{m^u},\end{aligned}$$

where we define ζ as

$$\begin{aligned}\zeta =& \frac{\Psi^u(s_1^u,s_{n^u}^u)}{\psi_0^u}\\ =& s_{n^u}^u - s_1^u + \frac{\beta_h^u}{\omega_h}\sin\left(\omega_h(r-\tau_h^u)\right)\Big|_{r=s_1^u}^{s_{m^u}^u} + \frac{\beta_d^u}{\omega_d}\sin\left(\omega_d(r-\tau_d^u)\right)\Big|_{r=s_1^u}^{s_{m^u}^u} +\\ & \frac{\beta_h^u\beta_d^u}{2(\omega_d+\omega_h)}\sin\left((\omega_d+\omega_h)(r-(\tau_h^u+\tau_d^u))\right)\Big|_{r=s_1^u}^{s_{m^u}^u} + \frac{\beta_h^u\beta_d^u}{2(\omega_d-\omega_h)}\sin\left(\omega_d-\omega_h(r-(\tau_h^u-\tau_d^u))\right)\Big|_{r=s_1^u}^{s_{m^u}^u}.\end{aligned}$$

We see from this that the conditional distribution of ψ_0^u is a gamma distribution with shape and scale parameters $1+m^u$ and $1/(\psi_0^{-1}+\zeta)$, so we can sample it directly.

Parameters β_h^u, β_d^u,τ_h^u, *and* τ_d^u. For β_h^u the conditional distribution is given by

$$\begin{aligned}\mathbf{P}(\beta_h^u|\mathbf{S},\Phi,\Phi^{user}_{-\beta_h^u}) \propto& \mathbf{P}(\mathbf{s}^u|\Phi^u)\mathbf{P}(\beta_h^u|\Phi)\\ \propto& \left(\Psi^u(s_1^u,s_{m^u}^u)\right)^{m^u}\exp\left(-\Psi^u(s_1^u,s_{m^u}^u) - \frac{(\beta_h^u-\beta_h)^2}{2(\sigma^2_{\beta_h})}\right).\end{aligned}$$

For β_d^u the conditional distribution is given by

$$\begin{aligned}\mathbf{P}(\beta_d^u|\mathbf{S},\Phi,\Phi^{user}_{-\beta_d^u}) \propto& \mathbf{P}(\mathbf{s}^u|\Phi^u)\mathbf{P}(\beta_d^u|\Phi)\\ \propto& \left(\Psi^u(s_1^u,s_{m^u}^u)\right)^{m^u}\exp\left(-\Psi^u(s_1^u,s_{m^u}^u) - \frac{(\beta_d^u)^2}{2(\sigma^2_{\beta_d})}\right).\end{aligned}$$

For τ_h^u the conditional distribution is given by

$$\begin{aligned}\mathbf{P}(\tau_h^u|\mathbf{S},\Phi,\Phi^{user}_{-\tau_h^u}) \propto& \mathbf{P}(\mathbf{s}^u|\Phi^u)\mathbf{P}(\tau_h^u|\Phi)\\ \propto& \left(\Psi^u(s_1^u,s_{m^u}^u)\right)^{m^u}\exp\left(-\Psi^u(s_1^u,s_{m^u}^u) - \frac{(\tau_h^u)^2}{2(\sigma^2_{\tau_h})}\right).\end{aligned}$$

Parameter	Posterior Mean	Posterior Standard Deviation
β_h	$(N+\sigma_{\beta_h}^2/100^2)^{-1}\sum_{u=1}^N \beta_h^u$	$\sigma_{\beta_h}(N+\sigma_{\beta_h}^2/100^2)^{-1/2}$
β_d	$(N+\sigma_{\beta_d}^2/100^2)^{-1}\sum_{u=1}^N \beta_d^u$	$\sigma_{\beta_d}(N+\sigma_{\beta_d}^2/100^2)^{-1/2}$
τ_h	$(N+\sigma_{\tau_h}^2/100^2)^{-1}\sum_{u=1}^N \tau_h^u$	$\sigma_{\tau_h}(N+\sigma_{\tau_h}^2/100^2)^{-1/2}$
τ_d	$(N+\sigma_{\tau_d}^2/100^2)^{-1}\sum_{u=1}^N \tau_d^u$	$\sigma_{\tau_d}(N+\sigma_{\tau_d}^2/100^2)^{-1/2}$

Table 6 Posterior mean and standard deviation of β_h, β_d, τ_h, and τ_d for the posting model.

For τ_d^u the conditional distribution is given by

$$\begin{aligned}\mathbf{P}(\tau_d^u|\mathbf{S},\Phi,\Phi_{-\tau_d^u}^{user}) \propto& \mathbf{P}(\mathbf{s}^u|\Phi^u)\mathbf{P}(\tau_d^u|\Phi)\\ \propto& \left(\Psi^u(s_1^u,s_{m^u}^u)\right)^{m^u}\exp\left(-\Psi^u(s_1^u,s_{m^u}^u)-\frac{(\tau_d^u)^2}{2(\sigma_{\tau_d}^2)}\right).\end{aligned}$$

To sample from the conditional distributions, we use a random walk Metropolis-Hastings step as was done for the timeline model. We consider the parameter β_h^u. For β_h^u we define the ith sample as β_{hi}^u, and the proposal for the $(i+1)$ sample is drawn from a normal distribution with mean β_{hi}^u and standard deviation 0.1, where this value is chosen to balance the acceptance rate with step size. We use this exact same procedure to sample from the conditional distributions of β_d^u, τ_h^u,and τ_d^u, except that we use different values for the standard deviation. For β_d^u and τ_h^u we choose a standard deviation of 0.1 and for τ_d^u we choose a standard deviation of one.

Parameter ψ_0. The prior distribution for the global parameter ψ_0 is inverse gamma with parameters one and one. The posterior distribution is given by

$$\begin{aligned}\mathbf{P}(\psi_0|\mathbf{S},\Phi_{-\psi_0},\Phi^u) \propto& \prod_{u=1}^N \mathbf{P}(\psi_0^u|\Phi)\mathbf{P}(\psi_0)\\ \propto& \psi_0^{-(N+1)-1}\exp\left(-\frac{1+\sum_{u=1}^N\psi_0^u}{\psi_0}\right)\end{aligned}$$

We see that the posterior distribution is again inverse gamma with shape parameter $N+1$ and scale parameter $1+\sum_{u=1}^N\psi_0^u$ and we can sample ψ_0 directly.

Parameters β_h, β_d, τ_h, *and* τ_d. The prior distribution for the global parameters β_h, β_d,τ_h, and τ_d are normal with zero mean and standard deviation of 100. Their posteriors also have a similar structure. We consider β_d. The posterior distribution is given by

$$\begin{aligned}\mathbf{P}(\beta_d|\mathbf{S},\Phi_{-\beta_d},\Phi^{user}) \propto& \prod_{u=1}^N \mathbf{P}(\psi_0^u|\Phi)\mathbf{P}(\psi_0)\\ \propto& \exp\left(-\frac{1}{2\sigma_{\beta_h}^2}\sum_{u=1}^N(\beta_h^u-\beta_h)^2-\frac{1}{2(100)^2}\sum_{u=1}^N\beta_h^2\right).\end{aligned}$$

We see that the posterior is normal with mean $(N+\sigma_{\beta_h}^2/100^2)^{-1}\sum_{i=1}^N\beta_h^u$ and standard deviation $\sigma_{\beta_h}(N+\sigma_{\beta_h}^2/100^2)^{-1/2}$, so we can sample this directly. A similar calculation shows that β_d,τ_h, and τ_d also have normal posteriors. We show the posterior means and standard deviations for these parameters in Table 6.

Parameters $\sigma_{\beta_h}^2$, $\sigma_{\beta_d}^2$, $\sigma_{\tau_h}^2$, *and* $\sigma_{\tau_d}^2$. The prior distribution for the global parameters $\sigma_{\beta_h}^2$, $\sigma_{\beta_d}^2$, $\sigma_{\tau_h}^2$, and $\sigma_{\tau_d}^2$ are inverse gamma with shape and scale parameters one and one. Their posteriors also have a similar structure. Consider $\sigma_{\beta_h}^2$. Its posterior distribution is given by

$$\begin{aligned}\mathbf{P}(\sigma_{\beta_h}^2|\mathbf{S},\Phi_{-\sigma_{\beta_h}^2},\Phi^{user}) \propto& \prod_{u=1}^N \mathbf{P}(\psi_0^u|\Phi)\mathbf{P}(\sigma_{\beta_h}^2)\\ \propto& \exp\left(-\frac{1}{2\sigma_{\beta_h}^2}\sum_{u=1}^N(\beta_h^u-\beta_h)^2-\frac{1}{\sigma_{\beta_h}^2}\right)\left(\sigma_{\beta_h}^2\right)^{-2-N/2}.\end{aligned}$$

Parameter	Posterior Shape Parameter	Posterior Scale Parameter
$\sigma^2_{\beta_h}$	$1+N/2$	$1+1/2\sum_{u=1}^{N}(\beta_h^u-\beta_h)^2$
$\sigma^2_{\beta_d}$	$1+N/2$	$1+1/2\sum_{u=1}^{N}(\beta_d^u-\beta_d)^2$
$\sigma^2_{\tau_h}$	$1+N/2$	$1+1/2\sum_{u=1}^{N}(\tau_h^u-\tau_h)^2$
$\sigma^2_{\tau_d}$	$1+N/2$	$1+1/2\sum_{u=1}^{N}(\tau_d^u-\tau_d)^2$

Table 7 Posterior shape and scale parameters of $\sigma^2_{\beta_h}$, $\sigma^2_{\beta_d}$, $\sigma^2_{\tau_h}$, and $\sigma^2_{\tau_d}$ for the posting model.

We see that the posterior distribution is again inverse gamma with shape parameter $1+N/2$ and scale parameter $1+1/2\sum_{u=1}^{N}(\beta_h^u-\beta_h)^2$ and we can sample $\sigma^2_{\beta_h}$ directly. A similar calculation shows that $\sigma^2_{\beta_d}$, $\sigma^2_{\tau_h}$, and $\sigma^2_{\tau_d}$ also have inverse gamma posteriors. We show their posterior shape and scale parameters in Table 7.

Acknowledgments

Sina Modaresi was partially supported by the National Science Foundation under grant CMMI-1233441.

References

Anderson, Simon P, André De Palma. 2012. Competition for attention in the information (overload) age. *The RAND Journal of Economics* **43**(1) 1–25.

Anderson, Simon P, André De Palma. 2013. Shouting to be heard in advertising. *Management Science* **59**(7) 1545–1556.

Aral, Sinan, Dylan Walker. 2012. Identifying influential and susceptible members of social networks. *Science* **337**(6092) 337–341.

Bandari, Roja, Sitaram Asur, Bernardo A Huberman. 2012. The pulse of news in social media: Forecasting popularity. *ICWSM*. 26–33.

Belotti, Pietro, Jon Lee, Leo Liberti, Francois Margot, Andreas Wächter. 2009. Branching and bounds tightening techniques for non-convex MINLP. *Optimization Methods & Software* **24**(4-5) 597–634.

Chen, Wei, Chi Wang, Yajun Wang. 2010. Scalable influence maximization for prevalent viral marketing in large-scale social networks. *Proceedings of the 16th ACM SIGKDD international conference on Knowledge discovery and data mining*. ACM, 1029–1038.

Chen, Wei, Yajun Wang, Siyu Yang. 2009. Efficient influence maximization in social networks. *Proceedings of the 15th ACM SIGKDD international conference on Knowledge discovery and data mining*. ACM, 199–208.

Cox, David Roxbee, Valerie Isham. 1980. *Point processes*, vol. 12. CRC Press.

Gelman, A., D. B. Rubin. 1992. Inference from iterative simulation using multiple sequences. *Statis. Sci.* **7** 457–472.

Hong, Liangjie, Ovidiu Dan, Brian D Davison. 2011. Predicting popular messages in twitter. *Proceedings of the 20th international conference companion on the world wide web*. ACM, 57–58.

Instagram. 2015. 300 million: Sharing real moments. `http://blog.instagram.com/post/104847837897/141210-300million`. Accessed: 2015-06-25.

Iyer, Ganesh, Zsolt Katona. 2015. Competing for attention in social communication markets. *Management Science* .

Katona, Zsolt, Peter Pal Zubcsek, Miklos Sarvary. 2011. Network effects and personal influences: The diffusion of an online social network. *Journal of Marketing Research* **48**(3) 425–443.

Kempe, David, Jon Kleinberg, Éva Tardos. 2003. Maximizing the spread of influence through a social network. *Proceedings of the ninth ACM SIGKDD international conference on Knowledge discovery and data mining*. ACM, 137–146.

Kempe, David, Jon Kleinberg, Éva Tardos. 2005. Influential nodes in a diffusion model for social networks. *Automata, languages and programming*. Springer, 1127–1138.

Naveed, Nasir, Thomas Gottron, Jérôme Kunegis, Arifah Che Alhadi. 2011. Bad news travel fast: A content-based analysis of interestingness on twitter. *Proceedings of the 3rd International Web Science Conference*. ACM, 8.

Peng, Huan-Kai, Jiang Zhu, Dongzhen Piao, Rong Yan, Ying Zhang. 2011. Retweet modeling using conditional random fields. *Data Mining Workshops (ICDMW), 2011 IEEE 11th International Conference on*. IEEE, 336–343.

Petrovic, Sasa, Miles Osborne, Victor Lavrenko. 2011. Rt to win! predicting message propagation in twitter. *ICWSM*.

Pozin, Ilya. 2010. 20 companies you should be following on social media. `http://www.forbes.com/sites/ilyapozin/2014/03/06/20-companies-you-should-be-following-on-social-media/`. Accessed: 2015-12-06.

Social Times. 2010. 10 apps to schedule future tweets on twitter. `http://www.adweek.com/socialtimes/10-apps-to-schedule-future-tweets-on-twitter/9868`. Accessed: 2015-12-06.

Spiegelhalter, D. J., N. G. Best, B. P. Carlin, A. van der Linde. 2002. Bayesian measures of model complexity and fit. *Journal of the Royal Statistical Society* **64** 583–639.

Suh, Bongwon, Lichan Hong, Peter Pirolli, Ed H Chi. 2010. Want to be retweeted? large scale analytics on factors impacting retweet in twitter network. *Second international conference on social computing*. IEEE, 177–184.

Twitter. 2015. About twitter, inc. `http://about.twitter.com/company`. Accessed: 2015-06-25.

Ugander, Johan, Lars Backstrom, Cameron Marlow, Jon Kleinberg. 2012. Structural diversity in social contagion. *Proceedings of the National Academy of Sciences* **109**(16) 5962–5966.

Van Zandt, Timothy. 2004. Information overload in a network of targeted communication. *The RAND Journal of Economics* **35**(3) 542–560.

Zaman, Tauhid, Emily B Fox, Eric T Bradlow, et al. 2014. A bayesian approach for predicting the popularity of tweets. *The Annals of Applied Statistics* **8**(3) 1583–1611.