

# Profiling OSN Users Based on Temporal Posting Patterns

Qiu Fang Ying

Dept. of Information Engineering  
The Chinese University of Hong Kong  
qfying@ie.cuhk.edu.hk

Srinivasan Venkatramanan

Virginia Tech  
vsriniv@bi.vt.edu

Dah Ming Chiu

Dept. of Information Engineering  
The Chinese University of Hong Kong  
dmchiu@ie.cuhk.edu.hk

Xiaopeng Zhang

Precise Recommendation Center, Tencent Inc.  
xpzhang@tencent.com

## ABSTRACT

In this paper, we study the posting behavior of OSN users, in particular the posting frequency and temporal patterns, and consider possible interpretations of how users use the platform. At the aggregate (macro) level, we find two distinct peaks, one during morning working hours, and one in the evening. The morning peak is more pronounced for frequent posters, while the evening peak is pronounced for the remaining users. We postulate that this difference results from qualitatively different usage of the OSN platform (e.g. for work, with customers, etc.) than purely social interactions (e.g., friends, family, etc.). We also study user posting behavior at an individual (micro) level and apply LDA to cluster user temporal patterns and interpret our results. Our study provides new insights into user activity in today's OSNs, and suggests a framework for profiling users based on their posting activities. In the process, we provide a novel application of LDA, to temporal user posting behavior by equating the time epochs of posts to words in documents. We believe our approach will complement other methods of user profiling based on static demographic information and friendship network information.

## KEYWORDS

user posting behavior, temporal pattern, time series analysis, LDA

### ACM Reference Format:

Qiu Fang Ying, Dah Ming Chiu, Srinivasan Venkatramanan, and Xiaopeng Zhang. 2018. Profiling OSN Users Based on Temporal Posting Patterns. In *WWW '18 Companion: The 2018 Web Conference Companion, April 23–27, 2018, Lyon, France*. ACM, New York, NY, USA, 6 pages. <https://doi.org/10.1145/3184558.3191592>

## 1 INTRODUCTION

Over the past decade, several Online social network (OSN) platforms have amassed millions, if not billions, of monthly active users (MAUs). As the number of platforms have expanded, the usage patterns and purposes have also evolved significantly. In addition to traditional social communication with friends or family, OSNs are increasingly being used for workplace interactions, news/media

consumption, opinion sharing, brand development and as commercial marketplaces. While some platforms (e.g., Pinterest) are exclusively used for certain functional purposes, other large platforms (e.g., Twitter) contain users with varied motives. These varied motives often reflect in the users' behavioral patterns on the platform. In this paper, we consider the following question: Given the diverse intents and purposes, what can we tell from users' behavior patterns, in particular posting patterns?

In order to address this question, we collected a random sample of 10000 anonymized users from Tencent social platform (China's largest OSN platform), and the ego network of these users comprising of a total of 1,107,782 users for our study. We also collected the time series of their posting activity (but not the content of their posts) within a twelve weeks period. In this paper, we study what we can observe based on the temporal and frequency of posting behaviors.

First, we analyze the aggregate posting behavior, at different time scales, such as weekly and daily trends. The most interesting observation is that there are two peaks for posting during a typical day: one in the morning during work hours and one in the evening. Furthermore, for users with the higher posting frequency, the morning peak is more pronounced. We also apply spectral analysis using the autocorrelation function to identify the periodicity of the posting time series. Then we study the individual posting behaviors, to discover common patterns among users. In particular, propose a novel application of LDA (Latent Dirichlet Allocation), an algorithm popular for text document topic modeling, to cluster temporal posting behavior patterns.

The organization of the paper is as follows. We begin with an overview of related work, and then describe the dataset with some background about the social network platform. We then present the main results pertaining to posting timings. In our conclusions, besides summarizing our findings, we also briefly discuss our ongoing work and future directions related to this paper.

## 2 RELATED WORK

### 2.1 Study of User Activities in OSNs

The topic of social networks has received a lot of attention in the computer science research community for more than two decades [6, 16]. Today's OSNs allow for a diverse set of activities than the mutual social communication and content exchange [1, 7, 18].

In recent years, there has been renewed interest in understanding the dynamics and temporal patterns of user activity on these

This paper is published under the Creative Commons Attribution 4.0 International (CC BY 4.0) license. Authors reserve their rights to disseminate the work on their personal and corporate Web sites with the appropriate attribution.

WWW '18 Companion, April 23–27, 2018, Lyon, France

© 2018 IW3C2 (International World Wide Web Conference Committee), published under Creative Commons CC BY 4.0 License.

ACM ISBN 978-1-4503-5640-4/18/04.

<https://doi.org/10.1145/3184558.3191592>

platforms. [10] studied the patterns of user content generation, especially posting counts, frequency, and the temporal patterns. They showed that users' posting behavior in the popular knowledge-sharing OSNs exhibits strong daily and weekly patterns. [5] demonstrated that total post count of users follow power law like function, and discuss some outliers. Some outliers are identified based on temporal patterns, but the authors do not attempt to profile users based on their temporal posting patterns.

## 2.2 Time Series Analysis

Models from the area of time series analysis [3] have been applied to study the temporal pattern of online behaviors in many aspects [7]. In particular, [19] applied time series analysis to study the temporal patterns associated with online textual content and derive the shapes characterizing the different types of media. Moreover, [13] presented a model of peoples' Web search behavior, and predicted the changes in query frequency and clicked URLs over time. The temporal modeling framework adapted from physics and signal processing have been used to unearth several properties observed in online user behavior, including trend, periodicity, noise, surprise, and seasonality detection. Furthermore, the more recent work of [4] analyzed temporal patterns of the content changes of three major news websites, using spectral analysis of the autocorrelation function to identify the periodicity. In this paper, we use similar approaches to detect periodicity of user posting patterns.

## 2.3 Topic Modeling

Latent topic analysis has emerged as one of the most effective methods for classifying, clustering and retrieving textual data [2, 9]. Many previous papers have applied topic modeling to study users' temporal online behavior, but most algorithms still relied on the text data [10, 14, 15]. In the work of [15], a temporal topic model was proposed for efficiently mining streams of social text such as a Twitter stream for an author, by modeling the topics and topic temporal transitions. Moreover, the work of [18] hypothesizes posting is generated by three factors: news, social, self-interest. It then builds an LDA model to capture these factors, which is based on the text content of posts. In some circumstance lacking of textual data, authors first built abstraction or categories of users access content before applying the LDA model. [8] develop a hierarchical URL dictionary generated from Yahoo! Directory and a cross-hierarchical matching method of automatic abstraction. This paper also describe the "topics" as intentions of users' next search action. While these topics are transient, in our paper we are interested in extracting the user's purpose for maintaining their OSN account. Other conventional studies apply topic modeling to clustering sequences of temporal activity [17], but the premise is knowing the symbols or features of activity. However, little previous work was found to map time instant as "words" instead of features. From this perspective, our paper is the first work to model user purposes of posting behavior in OSNs by applying the topic modeling to their temporal posting sequences, without disclosing the content of the activity.

## 3 DATASET

### 3.1 Description

In this paper we use a dataset collected from the Tencent social platform. The dataset contains a randomly sampled set of 10K users (with anonymized user IDs). Each record contains the user and the user's neighbors, in other words the user's ego network. Since users are sampled from a much larger pool, these ego networks are relatively disjoint from each other. We refer to each sampled user as *ego*, and the neighbors of the ego as *alters*. With the sampled egos and their neighbors, the dataset covers a total of 1,107,782 users.

### 3.2 Tencent OSN Platform

In order to understand the context the data comes from, we briefly explain the Tencent<sup>1</sup> social network platform. Tencent's original online social network is called QQ. It provides messaging and music streaming service to more than 840 million users. In 2011, a new social network application called WeChat was introduced, and it has become the most popular OSN platform provided by Tencent, with more than 980 monthly million active users. It has become a platform of serving different functions, including messaging, voice calls, group chatting, and commercial transactions. A user can add a friend by scanning the QR-code version of one's ID. Besides the ability to have a private messaging exchange with a friend, a user can also post a variety of content (text, photos, videos, links) to share with their friends. A user can see all the postings from friends in chronological order on a page called *Moments*. When a user comments (or likes) an entry in that user's *Moments*, it can only be seen by the author of the content and their mutual friends.

### 3.3 User Posting Time Series

When users publish new items in *Moments*, the OSN records the time of such posting events in system operating log. We collect the log of the sampled users from December 2016 to February 2017, spanning 12 weeks. Then we extract the posting behavior as a time series, for temporal analysis. All information that may indicates the identity of the sampled users are removed.

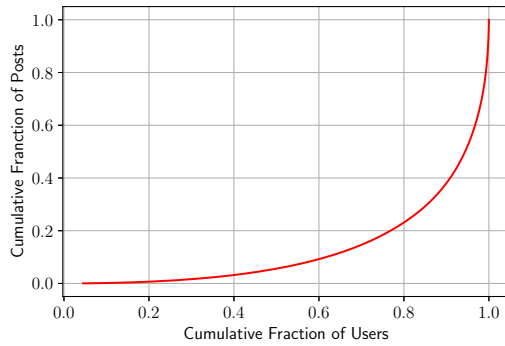
## 4 EMPIRICAL STUDY OF POPULATION BEHAVIOR

A user builds a virtual persona within the online social network through the friends he/she connects to and interacts with, and the content he/she posts on the platform. The activities performed by the user on the platform, do serve a latent purpose, which may include (a) connecting with friends/family, (b) promoting information spread or (c) advertising and selling products. The demographics information used traditionally to profile users in OSNs may not reveal these latent purposes. In this paper, we propose to profile users based on their posting activities. This assumes that there is a relationship between a user's purpose and his/her posting activities, so that we can learn the latent purpose of the user by mining the temporal pattern of user updates. We first study the overall behavior of the aggregate population of users, and then study the diverse behavioral patterns at the individual level.

<sup>1</sup><https://en.wikipedia.org/wiki/Tencent>

#### 4.1 Posting Frequency

We first study the frequency of users' posting activity at the macroscopic scale. Figure. 1 shows the distribution of users' posting frequency: the horizontal axis is the normalized rank of users (equal to the cumulative fraction of users), the vertical axis is the cumulative fraction of posts by the corresponding users. In the 3-month observation period, 20% users made nearly 80% posts, thus demonstrating the 80-20 or the Pareto rule. In order to understand the user behavior better, and since we do not include the content of the posts, we further study this topic by analyzing the temporal pattern of the posting activity.



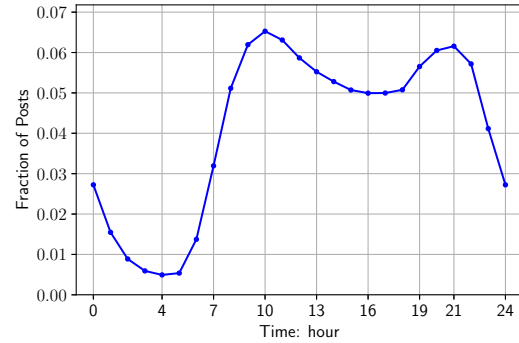
**Figure 1: Distribution of users' posting frequency: the x-axis is the normalized rank of users (equal to the cumulative fraction of users), the y-axis is the cumulative fraction of posts by the corresponding users**

#### 4.2 Aggregate Hourly Posting Pattern

Next, we analyze the temporal pattern of users' posting activity. Each post reports a timestamp indicating the posting time. We sum up all posts on hourly basis over the period of study, i.e., 84 days. Figure. 2 shows this aggregate distribution of posts over 24 hours. We find that the distribution has two peaks: one at 1000 hrs (morning peak) and the other at 2100hrs (evening peak). The morning peak has slightly more traffic than the evening peak, which is different than conventional wisdom [10]. On the other hand, the lowest traffic rate occurs at 0400hrs, which is not surprising. Another trough occurs at 1600hrs. The more pronounced morning peak traffic rate is surprising, compared to the previous study of OSNs. Based on the measurement results of a blog system [10], user generated content (UGC) such as blog articles show an increasing hourly pattern from morning to evening and only has a peak traffic in the evening. This could indicate a trend in the use of current OSN platforms:

- Increased usage of mobile phones, allow users to check their OSN during their morning commutes, and before work.
- Since today's OSNs support diverse functions, besides social communication, the morning work hours are a natural time to post content for people who use OSN for work. Such posts may be for distributing information, or posting daily product promotions.

In other words, our hypothesis is that OSN these days are used for both functional purposes as well as social purposes. Much of the former activities naturally occur in the working hours, especially in the morning.



**Figure 2: Aggregate distribution of posting time in the granularity of hour across over 84 days.**

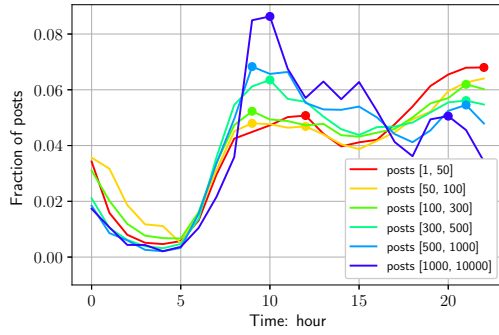
#### 4.3 Influence of Posting Frequency on Hourly Pattern

We look further into the correlation between users' posting frequency and their hourly pattern. We separate users into six groups by their post count in the observation period: [1, 50] [50, 100] [100, 300] [300, 500] [500, 1000] [1000+], and plot the hourly pattern of each group in Figure. 3. For users with fewer posts, the evening peak is more pronounced than the morning peak. As we move to the group of users with more posts, the morning peak is more pronounced. Further, the timing of the evening peak occurs earlier (from 2200hrs to 1900hrs).

Based on our preliminary study, we believe the reason for the movement of the traffic peak time is correlated the purposes why the users operate OSN accounts. Users who post less than 50 times in three months (less than one post everyday) are less likely to be posting for information spreading or product promotion purposes; and these users are more likely to post in the evening (e.g. 9pm) and during lunch break (around noon), which are non-working hours. In the other extreme, users who post more than 1000 times (more than 10 posts everyday) are more likely posting something for functional intention, e.g. posting some product information for commercial reasons. The high-traffic users have a sharp morning peak and relatively low traffic in the evening. Thus we believe the hourly posting pattern can be a good feature for predicting users' purposes.

#### 4.4 Periodicity of Temporal Pattern

In order to study the periodicity of the temporal posting pattern, we treat the aggregate posting count as a time series for 84 days, hence a time series  $\{Y_t\} = \{y_{t_1}, y_{t_2}, \dots, y_{t_N}\}$ , with  $t_1 \leq t_2 \leq \dots \leq t_N$  and  $N = 24 \times 7 \times 12 = 2016$ . The value of  $y_{t_i}$  is the number of posts at  $i^{th}$  hour during the 84 days. We plot the time series of the first



**Figure 3: Comparison of hourly patterns in group of users with different posting frequency.**

two weeks in Figure. 4. We can observe the periodicity exhibited by this time series, at hourly, daily and weekly periods.

To identify the periodicity of the time series, we apply spectral analysis using the autocorrelation function. The autocorrelation function  $R_h$  at lag  $h$  is defined as [4]:

$$R_h = \frac{\sum_{i=1}^{N-h} (y_{t_i} - \bar{y})(y_{t_{i+h}} - \bar{y})}{(y_{t_i} - \bar{y})^2}$$

where  $\bar{y}$  denotes the mean value of the time series. The function computed at varying time lags allows us to check the randomness of the data. If the autocorrelation is close to zero for any time lag, then the time series represents some random behavior, whereas if one or more values of time lag are significantly different from zero, they represent non-random behavior.

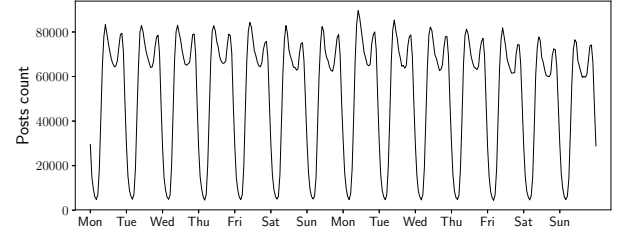
We compute the autocorrelation functions with time lags varying from one hour up to one week (168 hours) and plot  $R$  in Figure. 5. The plot shows repeated patterns, with almost the same magnitude. In particular, the test of independence performed at 5% significance level fails, as most of the values fall outside the 95% confidence bands delimited by the horizontal dashed lines, thus confirming the non-randomness of these patterns.

The nature of the periodicity can be derived from the power spectrum of the autocorrelation functions. The discrete Fourier coefficients  $f_k$  associated with the  $\frac{k}{T}$  frequencies is calculated as:

$$f_k = \sum_{i=0}^{N-1} y_{t_i} e^{-i2\pi \frac{j}{N} k}, \quad k = 0, 1, 2, \dots, N-1.$$

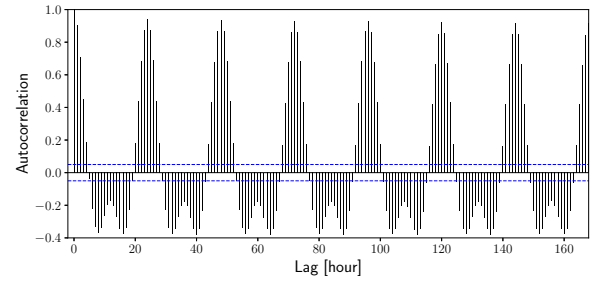
$T$  refers to the observation interval of the time series. We plot the periodogram of  $f_k^2$  as a function of frequency in Figure. 6. Here we choose  $T = 7$  days. The highest peak lies on the frequency of  $\frac{7}{T}$ , which corresponds to daily period of one day. There is also a smaller peak lies on the frequency of  $\frac{14}{T}$ , which corresponds to 12 hours. This period of 12 hours corresponds to the interval of two peaks in the daily posting pattern, that the morning peak and evening peak are separated by an interval of roughly 12 hours.

From the analysis of periodicity, we find that the users as an aggregate population show regular hourly and daily periodic pattern, but weak weekly or monthly patterns (the long-term trend). These

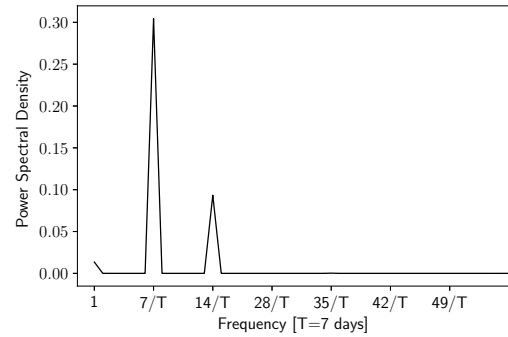


**Figure 4: Dynamics of the post count by the population of user over a two weeks interval.**

observation may allow us in the future to build models of population posting behavior based on this time series analysis, similar to [4, 13].



**Figure 5: Autocorrelation function computed at varying time lags for the users post count.**



**Figure 6: Power spectrum of the autocorrelation function of the users post count.**

## 5 LDA BASED MODELING OF INDIVIDUAL TIME SERIES

In section 4, we looked at macroscopic user posting behavior. In section 4.3, we found that users who have high posting frequency show different hourly pattern than those who have low posting

frequency, which actually indicates significant difference in behavior at the individual level, or microscopic level. In this section, we study individual user's posting behaviors and identify their hidden patterns.

As hinted earlier, the posting frequency to some extent reflects the latent purpose, i.e., why an individual operates an OSN account. A user who posts more than 10 items on their timeline each day is quite likely to have some functional intention (or purpose), compared to a typical social user who may only post less than 1 item daily. Also, these latent purposes tend to lead to specific temporal posting patterns, i.e., posting timings. Based on this intuition, we hypothesize we can learn users' latent purpose by profiling and clustering users posting activities.

We are inspired by the concept of topic modeling which is a powerful tool used in the text mining domain. In a typical topic model, documents are represented as a mixture of topics, where a topic is a probability distribution over words [8]. In our case, our model is that an individual user ("document") has a mixture of purposes ("topics"), and each purpose leads to a probability distribution over posting actions ("words"). The aim is to characterize each user by a probabilistic mixture of purposes. To realize it, we employ LDA (Latent Dirichlet Allocation), a common topic modelling algorithm, to model the time series of use posting behavior. To distinguish from the traditional LDA in texting mining, we call our model (applying to time series) T-LDA.

### 5.1 T-LDA Formulation

In most previous works that applied LDA to model user behavior, "words" are texts or abstract "categories" derived from texts. In our problem, we do not have the content or even categories of users' posts. Instead, we divide the time over 24 hours into a sequence of discrete and relative time epochs (e.g. each epoch is an hour); so a topic is a bag of "timestamps" (in terms of epochs). The basic ingredients of LDA and their corresponding meaning in T-LDA are therefore:

- Documents: users, each has a time series of posts.
- Topics: purposes of users posting behavior.
- Words: the timestamps of posts, in terms of discrete epochs, 00-23 hrs.

In our posting data, timestamps are in terms of hours, since the beginning of the measurement period. Formally,  $U$  indicates the set of users,  $T$  is the set of latent topics,  $N_u$  means the length of time series of user  $u$ , and  $D$  is the set of discretized times. It assumes a Dirichlet prior  $\beta$  for the topics and a Dirichlet prior  $\alpha$  for the users. The generative process of an individual's hourly posting pattern is as follows:

- (1) For each topic  $k$ , draw  $\varphi_k$  from  $Dir(\beta)$
- (2) For each user  $u$ , draw  $\theta_u$  from  $Dir(\alpha)$
- (3) For the  $i$ -th post published by  $u$  at time  $w_i$ :
  - (a) Choose a topic  $z_i$  from  $Multinomial(\theta_u)$
  - (b) Choose a time  $w_i$  from  $Multinomial(\varphi_{z_i})$

The input is a matrix  $M$  of user-timestamps (document-words), where each element of the matrix represents the count the corresponding user posted at the corresponding timestamp. The outputs are a matrix  $\Theta$  of user-topics (document-topics) denoting the topic distribution of each user, and a matrix  $\Phi$  of topic-times (topic-words)

denoting time distributions of each topic. The goal of topic modeling is to derive the optimal outputs  $\Theta$  and  $\Phi$ .

### 5.2 Experiments and Results

We first eliminated inactive users who posted fewer than 5 times in the 3 months period. Then we ran LDA following [11, 12], which used online variational Bayes algorithm as the inference algorithm.

We evaluate the optimality of the T-LDA model by perplexity, a common evaluation metric of clustering quality. Perplexity is a measure of the ability of a model to generalize documents. The definition of perplexity is as follows:

$$perplexity = \exp\left\{-\frac{\sum_{m=1}^U \log p(w_m^{test})}{\sum_{m=1}^U N_m}\right\}$$

$N_m$  is the length of time series for user  $m$ , and  $p(w_m^{test})$  denotes the predictive probability of a test word. A lower perplexity indicates better performance. Figure. 7 shows the results for the proposed model with different number of topics. The perplexity changes little after  $K > 6$ .

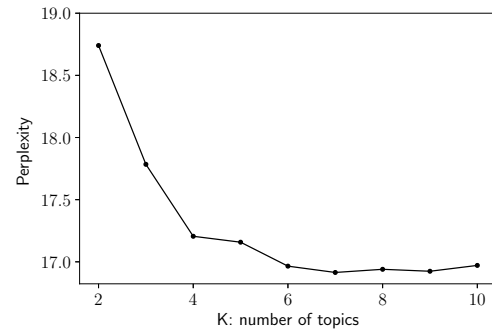


Figure 7: Perplexity of the T-LDA model.

In the output topic-time matrix  $\Phi$ , each row represents the topic's distribution over 24 hours. A straightforward way to visualize the topics is to plot the distribution of the row vectors. Considering the distribution of perplexity, we first plot the different topics when  $K=5$  in Figure. 8. Each curve plots the probability distribution over 24 hours, which reflects the hourly pattern of a topic, corresponding to a cluster of common user posting behavior. We make the following observations:

- Both Topic 1 and 2 have high morning peak traffic, but Topic 1 has peak at 7am to 8 am, and Topic 2 has peak at 10 am to 11 am. These two topics capture user behaviors with high activity rate in the early morning and at the beginning of work hours.
- Topic 3 and 4 represents user posting patterns with high activity rates in the evening. The hourly pattern of Topic 3 is similar to the typical pattern in the previous study of OSNs [10], which is consistent of posting more during leisure time of a day. And Topic 4 shows an increasing activity rate from daytime to nighttime and reaches the peak at 8pm.
- Topic 0 shows rather abnormal hourly pattern, contradicting most users' natural daily activity pattern. From checking

manually, we find this pattern belongs to a very small number of users who repeat exactly the same temporal pattern every day; so we believe that they may not be generated by human activity.

As a comparison for the case of five topics ( $K=5$ ), we also plot the results of  $K=3$  in Figure. 9. This time, Topic 0 of Figure. 8 is no longer detected, and Topic3 is merged into Topic 1. Obviously, some details are omitted in the  $K=3$  case. The parameter  $K$ , as we can see, plays a key role in controlling the right number of topics to be discovered.

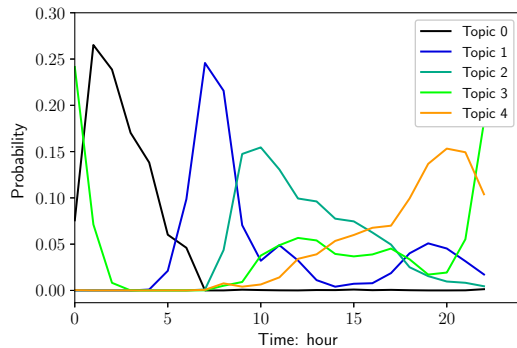


Figure 8: Distributions of topic-time for  $K=5$ .

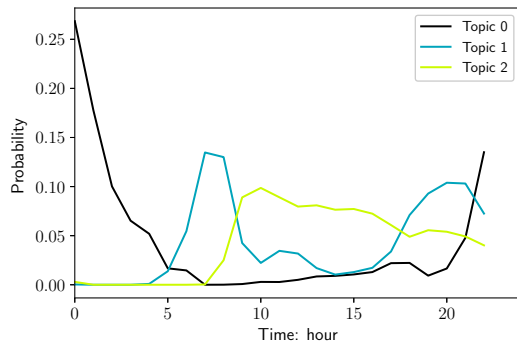


Figure 9: Distributions of topic-time for  $K=3$ .

## 6 DISCUSSION AND CONCLUSION

Our overall objective is to understand user behavior in online social networks, especially their posting patterns and how it relates to their latent purpose of using the platform. A great deal of past literature studied community detection, leveraging the socializing behavior of OSNs to discover communities and hence users' involvement in them. We believe today's OSN have become increasingly multi-functional, where users use OSN for work and business related purposes. It is an interesting research question to ask what we can learn about these users and their purposes based on less sensitive data, such as friend connectivity graph and activity event

logs. In this paper, we focus on user posting data, and try to identify clusters of common temporal patterns of posting behavior. Our main contributions in this paper are the new insights based on the analysis of aggregate frequency and temporal posting behaviors that lead to plausible indications of functional versus social users. Our application of LDA to cluster temporal user behavioral patterns is novel and shed light on different common behavior patterns. Future extensions include validating the observations against some subset of users in our dataset manually tagged to be functional or social users. For future study, we are working towards combining multiple features including network connectivity, demographics and posting behavior to more accurately profile users.

## REFERENCES

- [1] Fabricio Benevenuto, Tiago Rodrigues, Meeyoung Cha, and Virgilio Almeida. 2009. Characterizing user behavior in online social networks. In *Proceedings of the 9th ACM SIGCOMM conference on Internet measurement conference*. ACM, 49–62.
- [2] David M Blei, Andrew Y Ng, and Michael I Jordan. 2003. Latent dirichlet allocation. *Journal of machine Learning research* 3, Jan (2003), 993–1022.
- [3] George EP Box, Gwilym M Jenkins, Gregory C Reinsel, and Greta M Ljung. 2015. *Time series analysis: forecasting and control*. John Wiley & Sons.
- [4] Maria Carla Calzarossa and Daniele Tessa. 2015. Modeling and predicting temporal patterns of web content changes. *Journal of Network and Computer Applications* 56 (2015), 115–123.
- [5] Koutra D. Faloutsos M. et al. Devineni, P. 2017. Facebook Wall Posts: A Model of User Behavior. *Social Network Analysis and Mining* (2017). <https://doi.org/10.1007/s13278-017-0422-9>
- [6] David Easley and Jon Kleinberg. 2010. *Networks, crowds, and markets: Reasoning about a highly connected world*. Cambridge University Press.
- [7] Alceu Ferraz Costa, Yuto Yamaguchi, Agma Juci Machado Traina, Caetano Traina Jr, and Christos Faloutsos. 2015. Rsc: Mining and modeling temporal activity in social media. In *Proceedings of the 21th ACM SIGKDD International Conference on Knowledge Discovery and Data Mining*. ACM, 269–278.
- [8] Hiroshi Fujimoto, Minoru Etoh, Akira Kinno, and Yoshikazu Akinaga. 2011. Topic analysis of web user behavior using LDA model on proxy logs. In *Pacific-Asia Conference on Knowledge Discovery and Data Mining*. Springer, 525–536.
- [9] Thomas L Griffiths and Mark Steyvers. 2004. Finding scientific topics. *Proceedings of the National academy of Sciences* 101, suppl 1 (2004), 5228–5235.
- [10] Lei Guo, Enhua Tan, Songqing Chen, Xiaodong Zhang, and Yihong Eric Zhao. 2009. Analyzing patterns of user content generation in online social networks. In *Proceedings of the 15th ACM SIGKDD international conference on Knowledge discovery and data mining*. ACM, 369–378.
- [11] Matthew Hoffman, Francis R Bach, and David M Blei. 2010. Online learning for latent dirichlet allocation. In *advances in neural information processing systems*. 856–864.
- [12] Matthew D Hoffman, David M Blei, Chong Wang, and John Paisley. 2013. Stochastic variational inference. *The Journal of Machine Learning Research* 14, 1 (2013), 1303–1347.
- [13] Kira Radinsky, Krysta Svore, Susan Dumais, Jaime Teevan, Alex Bocharov, and Eric Horvitz. 2012. Modeling and predicting behavioral dynamics on the web. In *Proceedings of the 21st international conference on World Wide Web*. ACM, 599–608.
- [14] Zhiyong Shen, Ping Luo, Yuhong Xiong, Jun Sun, and Yidong Shen. 2009. Topic modeling for sequences of temporal activities. In *Data Mining, 2009. ICDM'09. Ninth IEEE International Conference on*. IEEE, 980–985.
- [15] Yu Wang, Eugene Agichtein, and Michele Benzi. 2012. TM-LDA: efficient online modeling of latent topic transitions in social media. In *Proceedings of the 18th ACM SIGKDD international conference on Knowledge discovery and data mining*. ACM, 123–131.
- [16] S Wasserman and Katherine Faust. 1994. *Social Network Analysis: Methods and Analysis*. Cambridge. (1994).
- [17] Jobin Wilson, Santanu Chaudhury, and Brejesh Lall. 2017. Clustering short temporal behaviour sequences for customer segmentation using LDA. *Expert Systems* (2017).
- [18] Zhiheng Xu, Yang Zhang, Yao Wu, and Qing Yang. 2012. Modeling user posting behavior on social media. In *Proceedings of the 35th international ACM SIGIR conference on Research and development in information retrieval*. ACM, 545–554.
- [19] Jaewon Yang and Jure Leskovec. 2011. Patterns of temporal variation in online media. In *Proceedings of the fourth ACM international conference on Web search and data mining*. ACM, 177–186.