Social Network Analysis and Mining (2018) 8:11

Toward Maximizing the Visibility of Content in Social Media Brand Pages: A Temporal Analysis

Nagendra Kumar · Gopi Ande · J. Shirish Kumar · Manish Singh

Abstract A large amount of content is generated everyday in social media. One of the main goals of content creators is to spread their information to a large audience. There are many factors that affect information spread, such as posting time, location, type of information, number of social connections, etc. In this paper, we look at the problem of finding the best posting time(s) to get high content visibility. The posting time is derived taking other factors into account, such as location, type of information, etc. In this paper, we do our analysis over Facebook pages. We propose six posting schedules that can be used for individual pages or group of pages with similar audience reaction profile. We perform our experiment on a Facebook pages dataset containing 0.3 million posts, 10 million audience reactions. Our best posting schedule can lead to seven times more number of audience reactions compared to the average number of audience reactions that users would get without following any optimized posting schedule. We also present some interesting audience reaction patterns that we obtained through daily, weekly and monthly audience reaction analysis.

Keywords Social media analysis, posting time, information spread, characterization

1 Introduction

Social media includes various web-based services that allow users to create and share the content with other users within their social network. A large amount of data is generated daily in social media. One of the main goals of a content creator is to spread the information to a large audience, and thereby receive a large number of audience reactions in the form of likes, comments, shares, etc.

The main obstacle in getting high information spread is that a post has a very short lifetime and within this short lifetime it has to compete with many other posts [1, 14]. In this paper, we use audience reaction as a measure to evaluate the information spread.

Indian Institute of Technology Hyderabad, India – 502285
Email: {cs14resch11005, cs12b1015, cs12b1018, msingh}@iith.ac.in

Many factors affect audience reactions, such as posting time, content type, location, connection in social network, and so on. In this paper, we primarily look at what is the effect of posting time on audience reactions? We propose techniques to compute posting schedules that will lead to increase audience reactions.

In this paper, we use publicly accessible Facebook pages to create our dataset. Facebook pages are maintained by brands, businesses, organizations, etc., to inform customers about their products and services. There are two types of social network relationships: *friend* relationship and *follower-following* relationship. Facebook pages use follower-following kind of relationship. Each page has admin(s) who create contents in the form of posts. Users can follow the page and create reactions in the form of likes, comments, and shares. We call these users as audience. Posted content is broadcasted to the news feed of followers and it has to compete with many other contents to be at the top of the followers' news feed.

In social media, most of the audience reactions are received within first few hours of posting [42]. If a content is posted at a time when audience are not online or not interested in interacting with the content, the content will not receive a large number of audience reactions. Facebook's News Feed algorithm [2] rewards a post if it is getting a large number of audience reactions by increasing rank of the post. If the post appears at the top of the news feed of many users, it would get more audience reactions and thereby becomes more popular.

Apart from looking at the ideal posting time for individual pages, it would be interesting to characterize the pages into groups with similar audience reaction profile. This will enable us to understand what are the factors that determine audience reactions. Given there are millions of Facebook pages, creating page category and then computing the posting schedule for the whole category will give higher statistical confidence while comparing the similarity and differences between various pages. With this characterization, we can also determine what would be the ideal posting schedule for a new page which does not have enough audience interaction history. Let us consider the following example task:

Example 1 Consider a set of traffic related Facebook pages, where each page contains information about traffic updates for a particular city. Following are some of the questions that we address in this paper:

1. What is the best time in a day that one should post about traffic updates to get maximum audience reactions?
2. Is there any difference in the audience reaction pattern over the week?
3. Are there typical periods during the year in which people tend to look more at traffic updates?
4. How audience reaction pattern of traffic pages compare with other types of Facebook pages?

Our key contributions are as follows:

- We analyze post-to-reaction behavior of Facebook pages. We show that 84% of the audience reactions are received within 24 hours after posting.

– We identify top features that affect audience reactions and use these features to categorize pages into groups with similar audience reaction profile.

– We propose six posting schedules for individual pages and groups of similar pages.

– We evaluate our algorithms on a dataset with 0.3 million posts and 10 million audience reactions. Our best posting schedule can lead to seven times more number of audience reactions compared to the average number of audience reactions that one would get without following any optimized posting schedule.

The rest of the paper is organized as follows. We formally define the problem of finding the right time to post to maximize the visibility of content in Section 2. Section 3 presents the audience reaction behaviour on Facebook pages. Section 4 discuss the categorization methods. Section 5 introduces the algorithm for schedule derivation. We proceed by describing schedule evaluations in Section 6. We briefly go through the related work in Section 7 and conclude our work in Section 8.

2 Problem Formulation

In this section, we present the problem definition and details about the used dataset.

2.1 Problem Definition

The problem of finding the right time to post can be defined in terms of the following sequence of sub-problems:

Problem 1 (Schedule for a Facebook page): *Given a Facebook page P, find a set of time-interval(s) T_P such that if a post $p \in P$ is posted during any time-interval $t_k \in T_P$, the post p is likely to get high visibility, which is measured using the number of audience reactions received on p.*

Problem 1 is the right time to create a post for a single Facebook page. If a post is created according to the proposed schedule T_P, it would get more audience reactions. According to Facebook's *News Feed* algorithm [2], if a post is getting a large number of audience reactions, the post will be given a chance to appear on top of the news feed of more number of users, thereby further increasing its likelihood to get high audience reactions. The schedule can be derived by using the posting behaviour of page admins (pages) or the reaction behaviour of audience. We state these two problems below.

Problem 1.1 (Frequent Posting Schedule): *Given a Facebook page P or a page category C, and the post creation profile M, find the frequent posting schedule S^{fp} for the page P or the category C.*

Admins of Facebook pages post a content at the time they receive the content (or just follow a certain personal schedule to post their contents). Although many admins may not be aware of when they should post to get maximum audience reactions, some expert admins with knowledge of social media post ranking might have an intuition of when they should post to get maximum audience reactions. They might realize this by trying out various posting schedules. Thus, our first problem is based on the most frequent posting schedule (*category* is defined later in Problem 2).

Frequent posting schedule can be of three types: *aggregated*, *category specific* and *weighted category specific* denoted as S^{afp}, S^{cfp}, and S^{wcfp} respectively. The aggregated schedule is the common schedule that can be used by all the pages. Categorized schedule is the customized schedule for the categories, and it is the best schedule for all the pages in a given category. Within a given category, all the pages may not have the same importance. Weighted category specific schedule is derived by giving higher weight to more important pages within the category.

Problem 1.2 (Frequent Reaction Schedule): *Given a Facebook page P or a page category C, and the audience reaction profile R, find the frequent reaction schedule S^{fr} for the page P or the category C.*

Since our goal is to maximize the number of audience reactions, the frequent reaction based schedule is derived by analyzing the posting timings that lead to high audience reaction. Frequent reaction schedules are also of three types: *aggregated*, *category specific* and *weighted category specific* denoted as S^{afr}, S^{cfr}, and S^{wcfr} respectively.

Problem 2 (Facebook page Categories): *Given a set of Facebook pages $\mathcal{P}$, a set of reaction determining features F_R, categorize the pages in $\mathcal{P}$ into r categories $\{C_1, C_2, .., C_r\}$ such that similarity between reaction profile is high for pages within a category and low across categories.*

Each Facebook page has a unique pattern of audience reaction. The pattern is not same for all the pages. Analyzing these reactions will help the page admins to get a deeper insight into their pages. For example, two e-commerce websites may have the different type of audience reaction patterns, even though they may be from the same location or the similar type of organization. By categorizing pages into categories with similar audience reaction profile, we can understand what are the different types of audience reaction profile? What are the factors that cause one page to get a certain type of audience reaction profile? If an organization wants its page to attain popularity similar to some other organization, what are the factors the organization should focus on to achieve that level of popularity? All these questions can be answered by looking at category-wise reaction behavior.

2.2 Dataset

We do our analysis on publicly accessible Facebook pages having a large number of audience. We obtain the dataset using the Facebook Graph API[1] in a similar way as described by Weaver et al. [40]. Each page has a profile page that contains posts created by page (posts created by the admin of page) and the reactions received on posts from the audience. Each page has a label (organization name) and a set of attributes (features). These attributes can vary across pages. A page can have attributes such as the number of fans (users who liked the page), the number of people talking about the page, type of the page, organization name, post creation time, reaction time, etc.

[1] https://developers.facebook.com/docs/graph-api

Audience can react on the posts created by Facebook pages in the form of like, comment and share. Reactions consist of a textual comment and a unary rating score in form likes and shares. As an audience member reads a post, she can optionally create a reaction to the post created by Facebook Page. Each audience member can contribute one or multiple reactions to a post. Audience are allowed to update previous reactions and add new reactions on the reacted posts. Since we could only access timestamp for comments, we use comments as the reaction and the time of comments creation as the reaction timestamp. Comments can be used to implicitly measure the interest generated by a post [27,37]. We extract the data of 100 Facebook pages from the same location that includes 5 different categories namely, e-commerce, traffic, telecommunication, hospital, and politician. Each of these categories contains the same number of pages to maintain homogeneity in audience reactions across the categories.

Notation	**Number**
R: Reactions	10 million
M: Posts/Messages	0.3 million
Y: Years	5 years
N: Number of pages	100

Table 1: Dataset Statistics

As can be seen in Table 1, our collected dataset contains 0.3 million posts and 10 million reactions that were created in 5 years (2011-2015). As the dataset contains many unimportant and noisy words, we pre-process the data using text-processing techniques [30] such as stop-word removal, stemming, lemmatization, etc. We remove stop words from posts and comments as these words do not contain important significance to be used in the analysis. We also perform stemming and lemmatization to reduce inflected or derived words to their root forms.

3 Audience Reaction Analysis

In this section, we look at the user dynamics in Facebook pages. We analyze the time delay between when a post is created and when the audience react to it. We also show different types of audience reaction that pages receive.

3.1 Post to Reaction Time Analysis

There is some time lag in post creation and audience reaction time [1,42]. It is important to study this time delay as some of the important features used to find the right time to post are derived from this time delay. Typically, a post receives 97% of its total audience reactions within the first week of its posting. So, we consider timespan of one week to analyze post-to-reaction delay. Figure 1 shows the distribution of audience reactions over a period of a week.

We observe in Figure 1 that a post receives around 34% of its total reactions within the 1^{st} hour of its posting, and 84% of reactions within a day. The lifespan of a post is

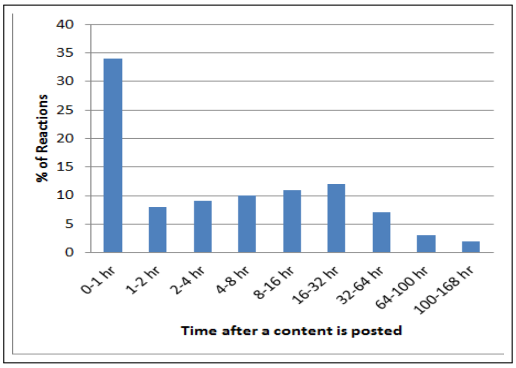

Fig. 1: Distribution of Reactions on a Post

very short, typically few hours and if it is not posted at the right time, it may not get high audience reactions. So it becomes important for Facebook pages to choose a right time of the day to post a content. A Facebook page can post a limited number of posts per day/week. If a page creates fewer posts, it will not engage audience enough for them to maintain a social connection with the page and the page will lose engagement. On the other hand, if a page creates a lot of posts, it will typically lose engagement. So, it is important to know the right time (daily, weekly, monthly) to create a post in Facebook page. This is the motivation for our proposed problem to find the right time to post to get maximum content visibility.

3.2 Audience Reaction Behavior Analysis

We present audience reaction behavior profile of some real world Facebook pages to understand the diversity of audience reaction pattern. We look at individual pages from politics, e-commerce, telecommunication, traffic, and hospital.

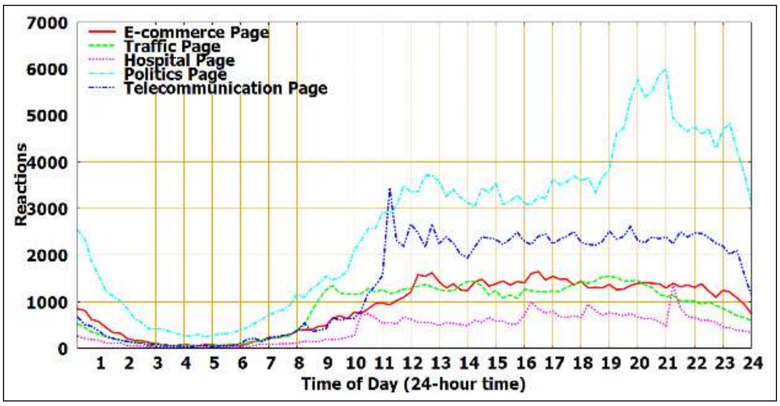

Fig. 2: Audience Reaction Behavior

As can be seen in Figure 2, the audience reaction behavior vary across time and pages. Some pages have one or more peaks per day. Some pages have a uniform peak throughout the day. The page maintained by a politician, receive peak audience reaction between 8:00 pm - 10:00 pm. Audience reaction is much less during rest of the day. For e-commerce and telecommunication related pages the peak is around 11 am, and then it decreases a bit for the rest of the day. It indicates that the audience reactions also depend on the content and characteristics of the page [5,33,10,43]. We give more detailed results on audience reaction analysis in Evaluation Section 6.

4 Categorization of Pages

In this section, we give a solution to Problem 2. We present the reaction determining features and describe the method of feature processing, page categorization.

4.1 Reaction Determining Features

To find features that affect audience reactions, we create 35 features. We use wrapper based feature selection to select the top reaction determining features. The features can be divided into following three types:

4.1.1 Page centric features

These are the features about the pages and signify popularity of the pages. Example features include the number of fans (those who have liked the page), the fan growth rate, the number of people who have created a story about the page on Facebook, and the number of posts per day.

4.1.2 Content centric features

These are the features about the page content. Example features include type of the page (described in Section 4.2); average number of likes, comments and shares for the whole page; average likes, comments and shares for different types of contents, such as Photos, Links, Videos, and average post length.

4.1.3 Reaction centric features

These are the features about audience reaction. Example features include the average number of audience reactions received within various time intervals after the post is created, such as 0-1 hrs, 1-2 hrs, 2-4 hrs, 4-8 hrs, 8-16 hrs, and 16-32 hrs; the average number of audience reactions received during various day intervals, such as 12:00 am - 4:00 am, 4:00 am - 8:00 am, 8:00 am - 12:00 pm, and so on. These features also include the average number of reactions received on days of a week and months of a year.

4.2 Feature Pre-processing

We perform various pre-processing for the above features, such as correct the time zone, correct the type of page, convert continuous valued attributes to discrete valued attributes. We extract the timestamp associated with each post and reaction. Graph API provides the time in Greenwich Mean Time (GMT) format; we convert it into regional time-zone.

Admins of Facebook pages create the label (or type) for their pages, and they name it based on the domain of the page/organization. There are six primary labels provided by Facebook for pages namely, “Local Business or Place”, “Company Organization or Institution”, “Brand or Product”, “Artist, Band or Public Figure”, “Entertainment”, “Cause or Community”. Each of these labels includes multiple sub-labels such as “Brand or Product” includes “website”, “electronics”, “product/service”, etc. Each page admin has to select one of these labels for their page. There are inconsistency between admins on how they select labels. For example, one e-commerce page is labelled as “Retail Company” and the other is labelled as “Website”. We use Nearest Neighbor algorithm [13] to label pages in a consistent manner, as page label is one of the most important factors in our posting schedule analysis. We use topic modeling to represent the pages in terms of topics, and then use cosine similarity of their topic probability to compute similarity between pages. For each page, we find its k-nearest neighbor pages. We then use majority label from these k neighbors to correct the page label. If the page label is labelled correctly, then majority will also have the same label. If it is not the most appropriate label, then it will differ from the majority and we correct it by assigning the majority label. Since organizations from similar domain post similar type of information, this technique can give all the pages of same domain the most common label used in that domain.

In order to characterize the pages based on these feature attributes, we convert these continuous attributes to discrete attributes. We apply entropy based data discretization [12] method to convert features in discrete attributes because most of the unsupervised data discretization methods require some parameter k such as number of bins. Entropy based method search through all possible values of k and capture inter-dependencies in features.

4.3 Categorization

We use clustering to group the pages with similar audience reaction. We use wrapper method [19] to select the features that are relevant for audience reaction. It considers selection of a subset of features as a search problem, where different combinations of features are used, evaluated and compared to other combinations. In the wrapper method, we use Multinomial Naive Bayes classifier [28] for classification. To create the base classes of Multinomial Naive Bayes, we use k-medoid clustering algorithm over the pages, where k is chosen using elbow method [18]. We define the similarity (refer to Equation 16) between two pages P_i and P_j in k-medoid as the similarity between their reaction profile $R_k(P_i)$ and $R_k(P_j)$ (reaction profile is defined in Section 5.1). We use k-medoid algorithm instead of k-means algorithm because of its robustness to outliers as compared to k-means. Moreover, it uses representative objects as cluster centers instead of taking the mean value of the objects as a cluster center. The top three obtained features are the *reaction within first one hour*, *number of posts posted by the page per day* and *type of the page* in increasing order of usefulness for the categorization. We cluster the pages using the top three reaction determining features as these three features are able to classify the pages into right category with the highest accuracy (90.3%) and increasing the number of features does not make a significant change in accuracy. The similarity in audience reaction within a category is high and across the categories is low when we use theses three features for categorization (as shown in Section 6.3).

5 Schedule Derivation

In this section, we give a solution to Problem 1. First, we describe notations used in schedule derivation and later present six ways to compute posting schedule. The first two schedules are generic schedules that are applicable for all pages, whereas the last four schedules are category specific.

Symbol	**Description**
P	a given set of Facebook pages
C_i	a set of similar Facebook pages, $C_i \subseteq P$
Y_s, d	Y_s is the base year in the dataset, d is the total number of years
t_k	a time bucket of size 15 minute
$r_k(P_x, Y_j)$	reaction profile vector of page P_x in Y_j^{th} year
$R_k(P_x)$	cumulative reaction profile vector of page P_x across d years
$m_k(P_x, Y_j)$	posting profile vector of page P_x in Y_j^{th} year
$M_k(P_x)$	cumulative posting profile vector of a page P_x across d years
$\gamma^r(P_x)$	total number of reactions received in page P_x in d years
$\gamma^m(P_x)$	total number of posts created by page P_x in d years
$\rho^r(C_i)$	total number of reactions received in category C_i in d years
$\rho^m(C_i)$	total number of posts created by category C_i in d years
$W^m(P_x)$	fraction of posts created by page P_x within its own category
$W^r(P_x)$	fraction of reactions received in page P_x within its own category
$\delta(C_i, k)$	reaction per post for category C_i in k^{th} bucket across d years
$\omega(C_i)$	aggregated reaction per post for category C_i in all buckets across d years

Table 2: Notations

Let's assume we have data from d years $\{Y_s, Y_s+1, ..., Y_s+d\}$. We divide a day into 96 discrete buckets $\{t_1, t_2, ..., t_{96}\}$, with each bucket of size 15 minutes as the bucket can capture essential reactions (as shown in Figure 1). By dividing a day time into small size of 96 buckets, we are able to determine right time (or bucket) more precisely. The first bucket t_1 is from 24:00 hrs to 00:15 hrs. We aggregate actions in the same time bucket from multiple years to ensure that our derived results are reliable. We consider two types of actions: creation (posting) and reaction. We denote posting and reaction profile for a given time bucket t_k, page P_z, and year Y_j as $m_k(P_z, Y_j)$ and $r_k(P_z, Y_j)$ respectively. $m_k(P_z, Y_j)$ is the aggregated number of posts created by page P_z at Y_j^{th} year (all days of year Y_j) in the time bucket t_k. For each bucket t_k, $m_k(P_z, Y_j)$ is computed by counting the number of posts created by page P_z in the time bucket t_k over the year Y_j. $r_k(P_z, Y_j)$ is the aggregated number of reactions received in page

P_z at Y_j^{th} year in the time bucket t_k. For each bucket t_k, $r_k(P_z, Y_j)$ is computed by adding all the reactions received by page P_z in the time bucket t_k over the year Y_j. We use these two profiles to compute the schedules.

5.1 Aggregated Schedules

We present two generic schedules, which are common for all the pages. The first schedule ($S_k^{afp}(P)$) is based on the aggregated frequent posting behavior and the second schedule ($S_k^{afr}(P)$) is based on aggregated frequent reaction behavior of all the pages.

Aggregated frequent posting schedule ($S_k^{afp}(P)$) is generated by using cumulative posting profile vector $M_k(P_z)$. For each time bucket t_k, $M_k(P_z)$ is the total number of posts created by page P_z in time bucket t_k across d years. $M_k(P_z)$ is computed by aggregating the posting profile vector $m_k(P_z, Y_j)$ of page P_z across d years as follows:

$$M_k(P_z) = \sum_{j=y_s}^{y_s+d} m_k(P_z, Y_j) \tag{1}$$

$S_k^{afp}(P)$ is a fraction of total number of posts created by all the pages in the t_k^{th} bucket. It is computed as follows:

$$S_k^{afp}(P) = \frac{\sum_{z=1}^{N} M_k(P_z)}{\sum_{z=1}^{N} \sum_{k=1}^{96} M_k(P_z)} \tag{2}$$

where $P_z \in P$ and $S_k^{afp}(P)$ is the fraction of total posts created by pages in k^{th} bucket, which is also defined as the probability of creating a post by pages in k^{th} bucket.

Similarly, aggregated frequent reaction schedule ($S_k^{afr}(P)$) is generated by using cumulative reaction profile vector $R_k(P_z)$. For each time bucket t_k, $R_k(P_z)$ is the total number of reactions received by the page P_z in the time bucket t_k across d years. $R_k(P_z)$ is computed by aggregating the reaction profile vector $r_k(P_z, Y_j)$ of page P_z across d years as follows:

$$R_k(P_z) = \sum_{j=y_s}^{y_s+d} r_k(P_z, Y_j) \tag{3}$$

$S_k^{afr}(P)$ is a fraction of total number of reactions received by all the pages in the t_k^{th} bucket. It is computed as follows:

$$S_k^{afr}(P) = \frac{\sum_{z=1}^{N} R_k(P_z)}{\sum_{z=1}^{N} \sum_{k=1}^{96} R_k(P_z)} \tag{4}$$

where $S_k^{afr}(P)$ is also defined as the probability of receiving audience reaction on pages in the k^{th} bucket. Now, we rank the buckets in decreasing order of $S_k^{afr}(P)$, $S_k^{afp}(P)$ with the first bucket being the best and the last one being the worst time to post according to these schedules respectively.

5.2 Categorized Schedules

As each category has different reaction behavior compared to other categories, we generate customized schedule for each category of Facebook pages. We derive two customized schedules for categories of Facebook pages, namely *categorized frequent posting schedule* and *categorized frequent reaction schedule*.

Categorized frequent posting schedule $S_k^{cfp}(C_i)$ is computed based on number of posts created by category C_i in time bucket t_k, and total number of posts created by category C_i in all the buckets as follows:

$$S_k^{cfp}(C_i) = \frac{\sum_{x=1}^{|C_i|} M_k(P_x)}{\sum_{k=1}^{96} \sum_{x=1}^{|C_i|} M_k(P_x)} \tag{5}$$

where $P_x \in C_i$, $M_k(P_x)$ is the cumulative posting profile vector of page P_x and $|C_i|$ is the total number of pages in category C_i. $S_k^{cfp}(C_i)$ is the fraction of total posts posted by category C_i in k^{th} bucket, which is also defined as the probability of creating a post by category C_i in k^{th} bucket. Similarly, categorized frequent reaction schedule $(S_k^{cfr}(C_i))$ is computed as follows:

$$S_k^{cfr}(C_i) = \frac{\sum_{x=1}^{|C_i|} R_k(P_x)}{\sum_{k=1}^{96} \sum_{x=1}^{|C_i|} R_k(P_x)} \tag{6}$$

where $R_k(P_x)$ is the cumulative reaction profile vector of page P_x. $S_k^{cfr}(C_i)$ is the fraction of total reactions received on category C_i at k^{th} bucket, which is also defined as the probability of receiving audience reaction on category C_i in k^{th} bucket.

We rank the buckets in decreasing order of $S_k^{cfp}(C_i)$ and $S_k^{cfr}(C_i)$. We pick first few buckets from both the schedules which are the right time to post for a category C_i according to these schedules. We compute categorized schedules for all the categories by following the same procedure. First time bucket of ranked schedules is the best time to post for category C_i in order to maximize content visibility.

5.3 Weighted Categorized Schedules

We derive the weighted categorized schedules by assigning weight to the pages of categories based on their importance. Some of the pages receive a large number of audience reactions and some of the pages post a large number of posts compared to other pages. To maintain homogeneity of actions and audience reactions across all pages in a category, we use weight factor $(W^r(P_x), W^m(P_x))$ in computation of the schedules. Weight signifies the importance of each page in its category. It is computed by using two parameters γ and ρ as follows:

$$\gamma^r(P_x) = \sum_{k=1}^{96} R_k(P_x) \tag{7}$$

$$\rho^r(C_i) = \sum_{x=1}^{|C_i|} \gamma^r(P_x) \tag{8}$$

$$W^r(P_x) = \frac{\gamma^r(P_x)}{\rho^r(C_i)} \tag{9}$$

where $\gamma^r(P_x)$ is the total number of reactions received by a page P_x and $\rho^r(C_i)$ is the total number of reactions received by a category C_i (all the pages of the category). Similarly, $\gamma^m(P_x)$, $\rho^m(C_i)$, and $W^m(P_x)$ are computed using cumulative posting profile vector $(M_k(P_x))$. Weighted categorized frequent posting schedule $S_k^{wcfp}(C_i)$ for category (C_i) is computed as follows:

$$S_k^{wcfp}(C_i) = \frac{\sum_{x=1}^{|C_i|} W^m(P_x) \times M_k(P_x)}{\rho^m(C_i)} \tag{10}$$

where $S_k^{wcfp}(C_i)$ computes the probability of creating a post by a category C_i at the k^{th} bucket. Now, we compute weighted categorized frequent reaction schedule $S_k^{wcfr}(C_i)$ for a category (C_i) as follows:

$$S_k^{wcfr}(C_i) = \frac{\sum_{x=1}^{|C_i|} W^r(P_x) \times R_k(P_x)}{\rho^r(C_i)} \tag{11}$$

where $S_k^{wcfr}(C_i)$ computes the probability of receiving audience reaction on category C_i in k^{th} bucket.

Weighted categorized schedule is similar to categorized schedule, the only difference is that weighted categorized schedule is computed by assigning a weight to each page of a category based on its importance in that category. We rank the buckets in decreasing order of $S_k^{wcfp}(C_i)$ and $S_k^{wcfr}(C_i)$ for all the categories. We pick first few buckets from both the schedules which are the right time to post for a category C_i according to these schedules. We compute weighted categorized schedules for all the categories.

6 Evaluations

In this section, we evaluate our proposed schedules, page categorization technique and present the audience reaction behaviour over time. We also discuss how the audience engagement varies with the type of post content.

6.1 Evaluation Metrics

We use *reaction gain* to evaluate the schedules and *correlation* to evaluate the quality of our categorization function.

6.1.1 Reaction Gain

Reaction gain metric is used to compute the performance of proposed schedules. It measures the change in reactions received in a particular time bucket, compared to the average reactions per post. Before computing the reaction gain for a schedule (S), we first rank the time buckets of schedule (S) over a period of 24 hours and compute two parameters: *reaction per post* (δ) and *aggregated reaction per post* (ω). Reaction per post (δ) is the total number of reactions received on pages within category C_i at time bucket t_k in d years divided by the total number of posts created at time bucket t_k by category C_i in d years. For the k^{th} rank bucket as per schedule (S) of category C_i, *reaction per post* (δ) is computed as follow:

$$\delta(C_i, k) = \frac{R_k(C_i)}{M_k(C_i)} \tag{12}$$

where $R_k(C_i)$ and $M_k(C_i)$ are the cumulative reaction profile vector and cumulative posting profile vector for the category C_i respectively. $R_k(C_i)$ and $M_k(C_i)$ are computed by aggregating the cumulative reaction profile vectors, cumulative posting profile vectors of all the pages in its own category respectively.

Aggregated reaction per post (ω) is the total number of reactions received on pages of category C_i divided by the total number of posts created by pages of category C_i.

$$\omega(C_i) = \frac{\sum_{k=1}^{96} R_k(C_i)}{\sum_{k=1}^{96} M_k(C_i)} \tag{13}$$

Now, reaction gain (RG) for time bucket t_k and category C_i is defined as:

$$RG(C_i, k) = \frac{\delta(C_i, k)}{\omega(C_i)} \tag{14}$$

where $RG(C_i, k)$ signifies the increase or decrease in reactions received by the category C_i when it posts in time bucket t_k, compared to the average reactions per post it receives.

Similarly, we compute the reaction gain ($RG(P, k)$) for the aggregated schedules by using $\delta(P, k)$, $\omega(P)$, $R_k(P)$, and $M_k(P)$. $R_k(P)$ and $M_k(P)$ are determined by aggregating the cumulative reaction profile vector and cumulative posting profile vector of all the pages respectively. Next, we compute average reaction gain for the categorized and weighted categorized schedules as these schedules contain multiple categories.

$$RG_{avg}(k) = \frac{\sum_{i=1}^{r} RG(C_i, k)}{r} \tag{15}$$

where average reaction gain ($RG_{avg}(k)$) for k^{th} time bucket is the average of $RG(C_i, k)$ across all the r categories. We use $RG_{avg}(k)$, $RG(P, k)$ to evaluate the performance of categorized schedules and aggregated schedules respectively.

6.1.2 Correlation

We use correlation metric to evaluate the effectiveness of the categorization method. We compute correlation across the categories by using the cumulative reaction profile vector of categories as follows:

$$Co(C_i, C_s) = \frac{\sum_{k=1}^{96}(R_k(C_i) - \bar{R}(C_i)) * (R_k(C_s) - \bar{R}(C_s))}{\sqrt{\sum_{k=1}^{96}(R_k(C_i) - \bar{R}(C_i))^2} * \sqrt{\sum_{k=1}^{96}(R_k(C_s) - \bar{R}(C_s))^2}} \tag{16}$$

where C_i and C_s are two different categories. $R_k(C_i)$ is the cumulative reaction profile vector (audience reaction) of category C_i in k^{th} bucket and $\overline{R}(C_i)$ is the average audience reaction of category C_i.

Similarly, we use the cumulative reaction profile vectors of categories of pages $(R_k(P_x))$ to compute the correlation within the category. We determine the correlation within the category by taking the average of correlation computed between each pair of the pages which belong to the same category.

6.2 Effect of Schedule

We evaluate our proposed six schedules using reaction gain metric defined in Section 6.1.1. As there are no previous baselines on best time to post for Facebook pages, we consider the first two generic schedules, namely aggregated frequent posting schedule and aggregated frequent reaction schedule as baseline schedules. We compute the average reaction gain for all the categorized schedules, aggregated schedules and pick the top-30 time buckets.

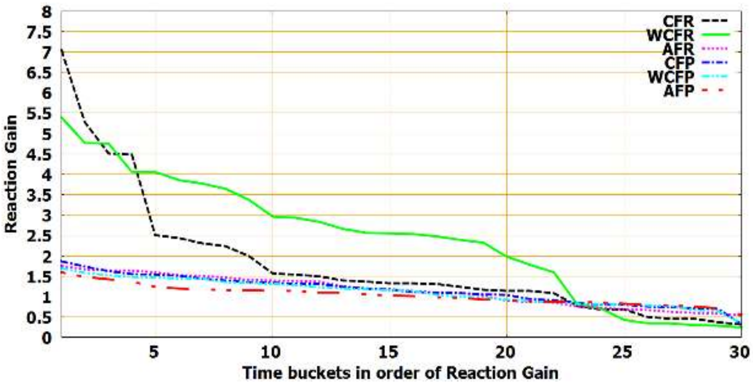

Fig. 3: Reaction Gain

We observe in Figure 3 that all the posting based schedules, such as S^{AFP}, S^{CFP}, and S^{WCFP} have reaction gain less than 2.0 even in their top bucket and their overall performance is also not as good as reaction based schedules. The reason is that most page admins do not know what is the right time to post a content. They may not be even aware of the fact that they can get better audience reaction by just choosing a better time for posting.

On the other hand, reaction based schedules perform far better compared to posting based schedules. It is also observed that category-wise schedules perform better than aggregated schedules (baseline schedules). Reaction gain of categorized frequent reaction schedule (S^{CFR}) is highest (i.e., seven times better) in its top bucket. Weighted categorized frequent reaction (S^{WCFR}) schedule shows a reaction gain of 5.4 in the top bucket. S^{WCFR} performs better than the S^{CFR} for all the buckets except the first two buckets. The reason could be that S^{CFR} is biased towards those buckets which receive a large number of audience reactions. If a page or category receives a large number of audience reactions in few buckets, it reflects high reaction gain in these buckets. However, S^{WCFR} is a normalized schedule and it does not show high reaction gain if few buckets receive high audience reaction.

6.3 Effectiveness of Categorization

We compute the correlation within and across categories to show the effectiveness of our categorization method. Let's consider five categories: C1, C2, C3, C4 and C5. We label these categories using the type of most frequent pages in that category. With this, the categories C1, C2, C3, C4, C5 represent e-commerce, telecommunication, hospital, politics, traffic respectively. We consider two ways of doing categorization: using single feature and using multiple features. From the top reaction determining features, we select the best feature for single feature case. In multiple feature case, we consider all the top reaction determining features.

Categories	**Single Feature**	**Multiple Feature**
C1 & C2	0.547	0.503
C1 & C3	0.392	0.341
C1 & C4	0.418	0.367
C1 & C5	0.519	0.470
C2 & C3	0.403	0.378
C2 & C4	0.353	0.302
C2 & C5	0.510	0.473
C3 & C4	0.351	0.305
C3 & C5	0.448	0.416
C4 & C5	0.440	0.419

Table 3: Correlation across the categories

Category	**Single Feature**	**Multiple Features**
C1	0.634	0.768
C2	0.703	0.848
C3	0.621	0.702
C4	0.650	0.771
C5	0.672	0.778

Table 4: Correlation within the category

We show across and within category correlation in Table 3 and 4 respectively for both types of categorization. Ideally, we would want within category correlation high and across category correlation low. In case of single feature case, we find that within and across category correlation is almost same. However, in case of multi-feature categorization, there is a large difference between within and across category correlation. These results indicate that our categorization function is able to categorize the pages effectively using multiple features. A new page that doesn't have enough reactions, can use this analysis to determine its right category and can post the accordingly (as described in Section 6.4) to get a large number of audience reactions. For ease of presentation, in rest of the paper, we refer the categories as e-commerce, politicians, etc. Each of these categories contains the same number of pages to maintain homogeneity in audience reaction across the categories.

6.4 Trend Analysis

We present some examples of audience reaction patterns which is observed in daily, weekly and monthly analysis.

6.4.1 Daily Analysis

For daily analysis, we analyze the reaction behavior for all the above mentioned five categories, for 24 hours period over a duration of 5 years. Unlike Figure 2 which shows audience reaction behavior of individual pages, Figure 4 shows the aggregated audience reaction behavior of the categories.

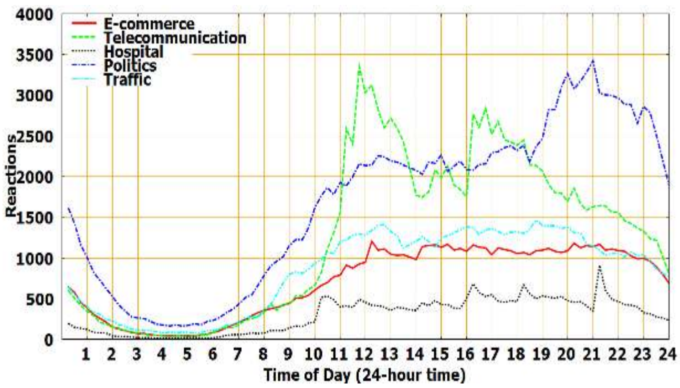

Fig. 4: Audience Reaction Pattern on Daily basis

We observe in Figure 4 that categories can have audience reaction in different ways, such as multiple peaks, single peak and uniform peak during a day.

First, we analyze the categories which have multiple reaction peaks in a day (i.e., traffic, telecommunication). Reactions on traffic category are high during the start of office hours (11 AM) and end of office hours (6 PM to 8 PM). One of the reasons is that there is high traffic in these time-periods and people react in Facebook pages about the traffic problems which they have faced while going or coming back from offices.

Similarly, telecommunications category has two peaks in the day: first is around 10 AM to 12 AM and second is around 4 PM to 6 PM. One of the reasons for this is that most of the people interact to social media pages in the morning to complain about an issue or to get the information related to tariffs, vouchers, special offers so that they can fill their balance and can use it throughout the day without out of balance problem. Some people prefer to do the same activity in the evening so that they can talk to family, friends, and relatives in the night when they become free from regular activities.

E-commerce category has uniform reactions from 12 PM to 10 PM (mostly during office hours) and drops after these hours. One of the possible reason is that people usually take the opinion of their colleagues and friends working in the same office about the product. If they found any issue, they often bring it to the notice of that e-commerce business immediately using Facebook page due to its quick response.

Pages related to politics and hospitals have single reaction peak per day. There is a high peak of audience reactions on politics category between 8 PM to 9 PM. One of the possible reasons is that people become free from their daily work by this time and spend some time in knowing the political updates which are posted during the daytime. Similarly, people complain more about hospital related issues in the evening which they faced during the daytime.

6.4.2 Weekly Analysis

In weekly analysis, we analyze audience reaction behavior on two categories namely telecommunication and traffic over the period of a week.

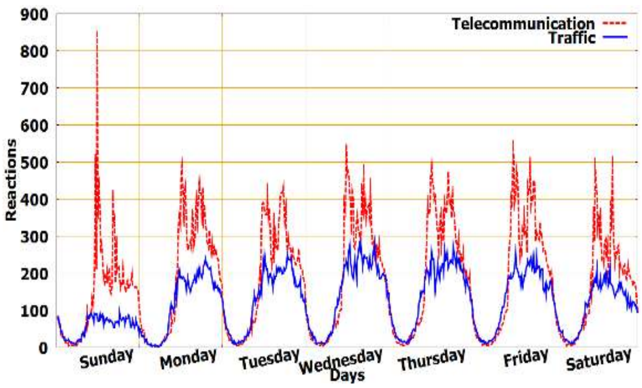

Fig. 5: Audience Reaction Pattern on Weekly basis

Telecommunication category has the highest peak during Sundays compared to other days of the week. One of the reasons is that most of the people are free on Sundays and they prefer to fill their mobile and data balances. People react more to posts related to telecommunication such as special offers, vouchers during these days. Therefore, it is better to post important updates and offers on Sundays instead of other weekdays to get a large number of audience reactions.

Reactions on traffic category are high during working days and drop slightly during weekends. One of the reasons is that people do not go to offices on weekends as they have holidays. Audience reactions drop to half during Sundays compared to other days of the week because even on Saturday some people still go to offices, but most of the people don't go to offices on Sunday. Most of the people stay at home and react less in traffic pages during weekends.

6.4.3 Monthly Analysis

In monthly analysis, we present audience reaction pattern on two categories namely e-commerce and politics over the period of a year.

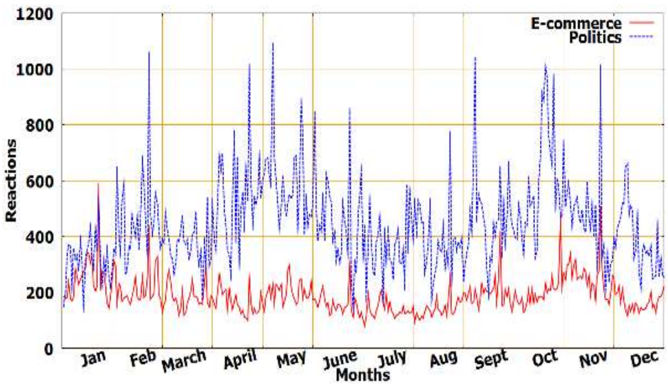

Fig. 6: Audience Reaction Pattern on Monthly basis

As can be seen in Figure 6, politics pages received more number of reactions in the months of April and May. One of the possible reasons is that the politics pages included in dataset had their elections in these months. People are more active on social media pages during election period. The peak in month of October and November is due to the introduction of new fiscal policies. People react more about the advantages and disadvantages of new policies through the social media pages during these periods.

E-commerce category has more number of reactions during the month of October and November because these are the festive months in India and people buy new goods on the occasion of festivals. Increase in reactions during mid of December, January is due to Christmas and end of the year sale. These are the festive occasions and people like to purchase new items during these occasions. They would be interested to know about offers and sales during these periods. If an e-commerce page post a news related to these sales and offers, people tend to react on it. Moreover, during these sales, lots of people purchase new items and a large fraction of these people face the problems such as delivery issue, product issue, etc. People share their experiences[2] and complaints[3] about the issue through the social media pages of e-commerce.

[2] Today discounts looks impressive hope big billion days rock coming days?

[3] I ordered product exchange offer honor b 15oct 2015 but yesterday cancelled order without information.

6.5 Audience Engagement with Contents

In this section, we show through empirical results that audience engagement depends on the type of content. Facebook page admins create different types of content, such as photo, video, link, and status. Some of these types of content receive more number of audience reactions compared to others. Pages can achieve higher audience engagement by creating contents of the type that receives more audience reaction. The results in Table 5 are based on the dataset mentioned in Section 2.2.

Content type	**Posts %**	**Reactions %**
Link	78.60%	54.16%
Photo	8.55%	18.46%
Status	1.59%	1.21%
Video	11.26%	26.17%

Table 5: Posts and reactions of different types of contents

In Table 5, for each content type, the second column shows the percentage of posts created by all the pages of that type, and the third column shows the percentage of reactions received by all the posts of that type. From the first row, we observe that although pages post 78.60% of the content as links, they get only 54.16% reactions from such content. In other words, links give less reaction (or audience engagement) per post. On the other hand, pages post only 8.55% and 11.26% content as images and videos, which brings 18.46% and 26.17% audience reaction respectively. Videos can bring highest audience engagement.

7 Related Work

One can use influential users to increase content visibility. There are existing works [3, 4, 7, 15, 17, 31] that find influential users in the social network, and use these influential users to spread information. Recently, researchers [32] at Klout developed a influence scoring system that measures influence of users for targeted search and marketing. In this paper, we look at the problem of spreading information by finding what is the right time to create a post so that the post can get high audience reactions. If a post is getting high audience reactions, it will automatically become popular because it would be shown at the top of audience news feed. Our approach is complementary to the existing approach of using influential users to increase content visibility.

To spread information in a social network, we need to understand the flow and diffusion of information in the social network [6, 8, 11, 14, 23, 25]. It requires an understanding of the topological structure and temporal characteristics of the social network [21, 24, 29, 36, 39]. For example, if a user connected to many users, posts some information it will automatically reach to many users. In this paper, we do not use topological structure. However, we use the right time to post to increase information diffusion. In future work, we would look at combining topological analysis with our approach to get even higher information diffusion.

To understand audience reaction behavior, we need to understand the user dynamics in social networks [9, 20, 22, 41, 44]. User dynamics changes across different social networks [26]. For example, in Twitter, the lifetime of content is quite short compared to other social networks. Some of the topics end in just 20-40 minutes [1]. Wu et al. [42] show that regardless of the type of content, all contents have a very short life span that usually drops exponentially after a day. One of the most recent work in user dynamics carried out by Rizoiu et al. [34] models the popularity dynamics of online items. They investigate the factors that influence the forecast of future popularity under promotion and use it to quatify expected attention generated by external promotion. In this paper, we study the user dynamics in Facebook pages. This social network is somewhat different compared to Twitter. Here user dynamics is somewhat slower compared to Twitter. Moreover, there have been few studies on finding the right posting schedule for social network users which stated that posting time also depends on the user dynamics [16, 35, 45]. However, these works mainly focused on finding the right posting schedule for individual users in social network. Their posting schedules are derived based and the users' social connections and locations. They do not look at many other features that can affect audience reactions, such as features about the content [5, 10, 33] or features about the content creator [43]. Our work is complementary to existing approaches that attempts to find the right time to spread the information of social media brand pages towards a large audience.

In this paper, we look at Facebook pages, which has follower-following type of relationship. A page can have unlimited number of followers, whereas a user can have at most few thousand friends [38]. We look at large number of features to find the best posting schedule. In addition to compute schedule for individual pages, we also look at the problem of finding schedule for a group of pages with similar audience reaction.

8 Conclusion

In this paper, we looked at the problem of how to increase the visibility of a content in social media brand pages by posting messages at a time that increases the likelihood of getting audience reactions. We analyzed user dynamics for individual Facebook pages as well as for a group of Facebook pages, with similar reaction profile, which we call page category. We proposed six schedules for getting high audience reaction, amongst which the best schedule leads to seven times higher reaction gain. We presented interesting audience reaction patterns in the form of daily, weekly and monthly temporal patterns.

References

1. Asur, S., Huberman, B.A., Szabo, G., Wang, C.: Trends in social media: Persistence and decay. In: Fifth International AAAI Conference on Weblogs and Social Media. pp. 434–437 (2011)
2. Backstrom, L.: News feed fyi: A window into news feed. https://www.facebook.com/business/news/News-Feed-FYI-A-Window-Into-News-Feed (2013)

3. Bapna, R., Umyarov, A.: Do your online friends make you pay? a randomized field experiment on peer influence in online social networks. Management Science 61(8), 1902–1920 (2015)
4. Bharathi, S., Kempe, D., Salek, M.: Competitive influence maximization in social networks. Internet and Network Economics pp. 306–311 (2007)
5. Biswas, R., Riffe, D., Zillmann, D.: Mood influence on the appeal of bad news. Journalism Quarterly 71(3), 689–696 (1994)
6. Centola, D.: The spread of behavior in an online social network experiment. science 329(5996), 1194–1197 (2010)
7. Chen, W., Wang, Y., Yang, S.: Efficient influence maximization in social networks. In: Proceedings of the 15th ACM SIGKDD international conference on Knowledge discovery and data mining. pp. 199–208. ACM (2009)
8. Cheng, J., Adamic, L., Dow, P.A., Kleinberg, J.M., Leskovec, J.: Can cascades be predicted? In: Proceedings of the 23rd international conference on World wide web. pp. 925–936. ACM (2014)
9. Das, A., Gollapudi, S., Munagala, K.: Modeling opinion dynamics in social networks. In: Proceedings of the 7th ACM international conference on Web search and data mining. pp. 403–412. ACM (2014)
10. Esiyok, C., Kille, B., Jain, B.J., Hopfgartner, F., Albayrak, S.: Users' reading habits in online news portals. In: Proceedings of the 5th Information Interaction in Context Symposium. pp. 263–266. ACM (2014)
11. Farajtabar, M., Ye, X., Harati, S., Song, L., Zha, H.: Multistage campaigning in social networks. In: Advances in Neural Information Processing Systems. pp. 4718–4726 (2016)
12. Fayyad, U.M., Irani, K.B.: On the handling of continuous-valued attributes in decision tree generation. Machine learning 8(1), 87–102 (1992)
13. García-Laencina, P.J., Sancho-Gómez, J.L., Figueiras-Vidal, A.R., Verleysen, M.: K nearest neighbours with mutual information for simultaneous classification and missing data imputation. Neurocomputing 72(7), 1483–1493 (2009)
14. Guille, A.: Information diffusion in online social networks. In: Proceedings of the 2013 SIGMOD/PODS Ph. D. symposium. pp. 31–36. ACM (2013)
15. Kao, L.J., Huang, Y.P., Sandnes, F.E.: Mining time-dependent influential users in facebook fans group. In: IEEE International Conference on Systems, Man, and Cybernetics, SMC 2016. pp. 718–723. IEEE (2016)
16. Karimi, M.R., Tavakoli, E., Farajtabar, M., Song, L., Gomez Rodriguez, M.: Smart broadcasting: Do you want to be seen? In: Proceedings of the 22nd ACM SIGKDD. pp. 1635–1644. ACM (2016)
17. Kempe, D., Kleinberg, J., Tardos, É.: Maximizing the spread of influence through a social network. In: Proceedings of the ninth ACM SIGKDD international conference on Knowledge discovery and data mining. pp. 137–146. ACM (2003)
18. Kodinariya, T.M., Makwana, P.R.: Review on determining number of cluster in k-means clustering. International Journal 1(6), 90–95 (2013)
19. Kohavi, R., John, G.H.: Wrappers for feature subset selection. Artificial intelligence 97(1-2), 273–324 (1997)
20. Kwak, H., Lee, C., Park, H., Moon, S.: What is twitter, a social network or a news media? In: Proceedings of the 19th international conference on World wide web. pp. 591–600. ACM (2010)
21. Kwon, S., Cha, M., Jung, K., Chen, W., Wang, Y.: Prominent features of rumor propagation in online social media. In: Data Mining (ICDM), 2013 IEEE 13th International Conference on. pp. 1103–1108. IEEE (2013)
22. Lehmann, J., Gonçalves, B., Ramasco, J.J., Cattuto, C.: Dynamical classes of collective attention in twitter. In: Proceedings of the 21st international conference on World Wide Web. pp. 251–260. ACM (2012)
23. Lerman, K., Ghosh, R.: Information contagion: An empirical study of the spread of news on digg and twitter social networks. ICWSM 10, 90–97 (2010)
24. Leskovec, J., Backstrom, L., Kleinberg, J.: Meme-tracking and the dynamics of the news cycle. In: Proceedings of the 15th ACM SIGKDD international conference on Knowledge discovery and data mining. pp. 497–506. ACM (2009)
25. Leskovec, J., McGlohon, M., Faloutsos, C., Glance, N., Hurst, M.: Patterns of cascading behavior in large blog graphs. In: Proceedings of the 2007 SIAM international conference on data mining. pp. 551–556. SIAM (2007)

26. Li, J., Peng, W., Li, T., Sun, T., Li, Q., Xu, J.: Social network user influence sense-making and dynamics prediction. Expert Systems with Applications 41(11), 5115–5124 (2014)
27. Mazloom, M., Rietveld, R., Rudinac, S., Worring, M., van Dolen, W.: Multimodal popularity prediction of brand-related social media posts. In: Proceedings of the 2016 ACM on Multimedia Conference. pp. 197–201. ACM (2016)
28. McCallum, A., Nigam, K., et al.: A comparison of event models for naive bayes text classification. In: AAAI-98 workshop on learning for text categorization. vol. 752, pp. 41–48. Madison, WI (1998)
29. Pei, S., Makse, H.A.: Spreading dynamics in complex networks. Journal of Statistical Mechanics: Theory and Experiment 2013(12), P12002 (2013)
30. Perkins, J.: Python 3 Text Processing with NLTK 3 Cookbook. Packt Publishing Ltd (2014)
31. Probst, F., Grosswiele, D.K.L., Pfleger, D.K.R.: Who will lead and who will follow: Identifying influential users in online social networks. Business & Information Systems Engineering 5(3), 179–193 (2013)
32. Rao, A., Spasojevic, N., Li, Z., DSouza, T.: Klout score: Measuring influence across multiple social networks. In: Big Data (Big Data), 2015 IEEE International Conference on. pp. 2282–2289. IEEE (2015)
33. dos Rieis, J.C.S., de Souza, F.B., de Melo, P.O.S.V., Prates, R.O., Kwak, H., An, J.: Breaking the news: First impressions matter on online news. In: Ninth International AAAI Conference on Web and Social Media. pp. 357–366 (2015)
34. Rizoiu, M.A., Xie, L.: Online popularity under promotion: Viral potential, forecasting, and the economics of time. In: ICWSM. pp. 182–191 (2017)
35. Spasojevic, N., Li, Z., Rao, A., Bhattacharyya, P.: When-to-post on social networks. In: Proceedings of the 21th ACM SIGKDD International Conference on Knowledge Discovery and Data Mining. pp. 2127–2136. ACM (2015)
36. Szabo, G., Huberman, B.A.: Predicting the popularity of online content. Communications of the ACM 53(8), 80–88 (2010)
37. Tatar, A., Antoniadis, P., De Amorim, M.D., Fdida, S.: From popularity prediction to ranking online news. Social Network Analysis and Mining 4(1), 174 (2014)
38. Taylor, D.G., Lewin, J.E., Strutton, D.: Friends, fans, and followers: do ads work on social networks? Journal of advertising research 51(1), 258–275 (2011)
39. Tsytsarau, M., Palpanas, T., Castellanos, M.: Dynamics of news events and social media reaction. In: Proceedings of the 20th ACM SIGKDD international conference on Knowledge discovery and data mining. pp. 901–910. ACM (2014)
40. Weaver, J., Tarjan, P.: Facebook linked data via the graph api. Semantic Web 4(3), 245–250 (2013)
41. Wu, S., Das Sarma, A., Fabrikant, A., Lattanzi, S., Tomkins, A.: Arrival and departure dynamics in social networks. In: Proceedings of the sixth ACM international conference on Web search and data mining. pp. 233–242. ACM (2013)
42. Wu, S., Hofman, J.M., Mason, W.A., Watts, D.J.: Who says what to whom on twitter. In: Proceedings of the 20th international conference on World wide web. pp. 705–714. ACM (2011)
43. Yan, R., Huang, C., Tang, J., Zhang, Y., Li, X.: To better stand on the shoulder of giants. In: Proceedings of the 12th ACM/IEEE-CS joint conference on Digital Libraries. pp. 51–60. ACM (2012)
44. Yu, L., Cui, P., Song, C., Zhang, T., Yang, S.: A temporally heterogeneous survival framework with application to social behavior dynamics. In: Proceedings of the 23rd ACM SIGKDD International Conference on Knowledge Discovery and Data Mining. pp. 1295–1304. ACM (2017)
45. Zarezade, A., Upadhyay, U., Rabiee, H.R., Gomez-Rodriguez, M.: Redqueen: An online algorithm for smart broadcasting in social networks. In: Proceedings of the Tenth ACM International Conference on Web Search and Data Mining. pp. 51–60. ACM (2017)