CHAPTER 3

Predicting Content Popularity in Social Networks

Yan Yan, Ruibo Zhou, Xiaofeng Gao and Guihai Chen

CONTENTS

CENTRALIZED with users being the creators and propagators, social networks tend to be an indispensable part of modern people's lives, in the era of Web 2.0. Massive amount of users' thoughts and friendships are implied in social networks, which becomes a promising source of big data. One of the most significant meanings for data mining is to analyze the underlined relations among data, and use it for the future. In a social network, the limitation of users' time and attention determines that users will only focus on what they are interested in and what is popular for the time being. Predicting what is popular in time will not only improve the utilization of users' time and attention, but also benefit social websites to offer better services to their users. In this chapter, we intend to research the popularity prediction of textual content, using big data in social networks. We focus on methods and models of prediction, which are well classified by elements the models consider, such as user behaviors, the life cycles of information, and the social network topology. We also reveal researchers' work on classifying social networks, evaluating metrics, as well as feature selection, and what remains to be done.

3.1 INTRODUCTION

Social networks such as Twitter, Facebook, Flickr and Instagram are well-known to people all over the world. People, especially the young, usually use more than one kind of social network service simultaneously in order to keep in touch with friends in different circles. Besides that, social networks also provide ordinary people with a fancy chance to keep a close eye on the life of celebrities, such as President Barack Obama, Taylor Swift, and Kobe Bryant. According to the Twitter Company Statistics, the total number of registered Twitter users is 645,750,000, and active users are 289,000,000. The total number of monthly active users in Facebook reaches 1,310,000,000. Via analyzing the large volume of data, we can know better what users need and what they like, so as to make more contributions to attract and keep more users in social network. Take Twitter as an example. Each Twitter user has a Homepage, from which we can view what is new of the person whom we are following. Moreover, on the left part of Homepage, there is a column called "trend". If the team of Twitter can predict accurately and instantaneously what is popular recently, and how long the trend will last, then they can put these popular contents on the trend, drawing more users' attention, which also benefits Twitter itself.

Social network has so many topics that we can dig deeper into, such as social group recognition and influential users discovery. Social group recognition is to divide social network users into groups by their hobbies, interests or occupations, so that organizers can recommend goods to each group easily. Discovering influential users is also important, in that these vital users may contribute a lot in topic spreading. Furthermore, big data in social network has the potential to solve cross-domain challenges, such as surveillance of public health. To keep pace with new developing technologies, social network is now combining with crowdsourcing, which is an up-to-date method to gather the wisdom of all people who take part in it. In addition to structural-hole theory, information-diffusion theory, the challenges in social network seem to have new ways to solve.

As our predicting basement is social network, we first give a brief introduction to the social network.

3.1.1 What is a Social Network?

A social network is a social structure made up of a set of social actors (such as individuals or organizations) and a set of the dyadic ties between these actors. The social network perspective provides a set of methods for analyzing the structure of the whole social entity as well as a variety of theories explaining the patterns observed in these structures [41]. Studies of these structures use social network analysis to identify local and global patterns, locate influential entities, and examine network dynamics.

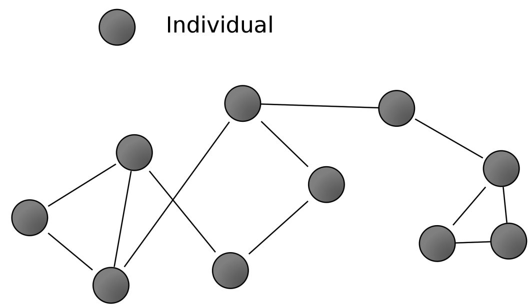
3.1.2 Levels of Social Network

Social network is self-organized, emergent, and complex. Since there are so many aspects to analyze, we should firstly provide a partition for it. In general, social networks can be divided into three levels:[1] micro level (Figure 3.1(a)), meso level (Figure 3.1(b)), and macro level (Figure 3.1(c)). In the micro level, researchers focus on the individuals in the social network. Thus we dig deeper into users' behaviors in this level. In the meso level, the formation of groups draws our attention. Correspondingly, we find groups in social network, and try to pick out the leader of the group. In the macro level, we take the whole large-scale social network graph into account. At this moment, we are more interested in hierarchical structures.

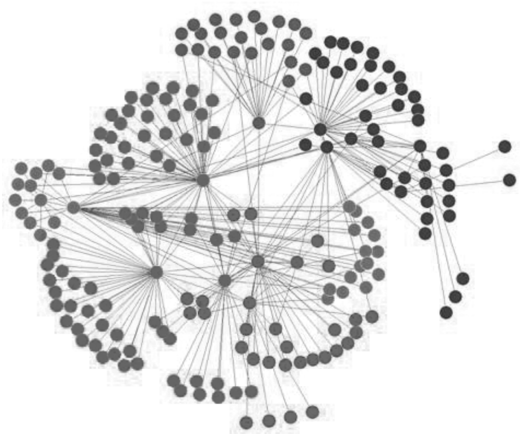
3.1.3 The Long Tail

The long tail shape (shown in Figure 3.2) is a common shape to illustrate the 2-dimensional relationship in social network, such as the relationship between the number of users and their followers, or the number of users and things they post [25]. An example can be shown in Figure 3.2. From this figure we can observe users' relation shapes in the social network like Twitter. Users

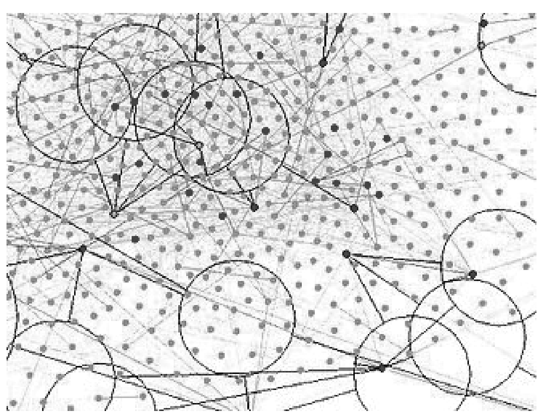
[1] From Wikipedia: *https : //en.wikipedia.org/wiki/Socialnetwork*



(a) Micro level



(b) Meso level



(c) Macro level

FIGURE 3.1 Different levels of social network (From: Wikipedia).

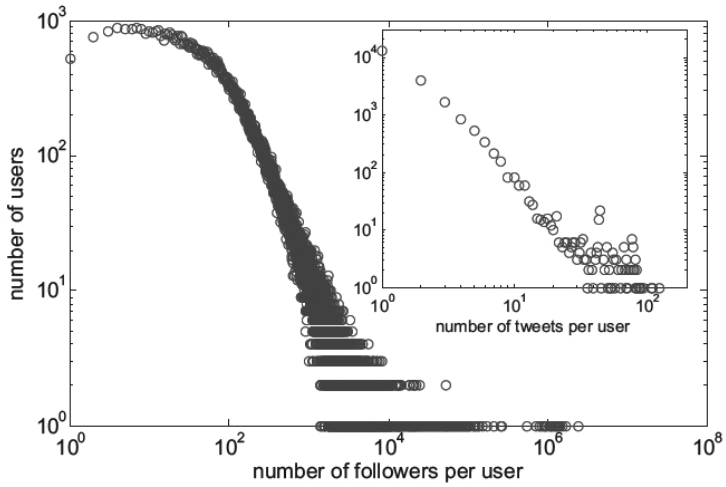


FIGURE 3.2 The long tail shape from Twitter data set [25].

who have a small number of followers are the majority, while it is rare to have a very large number of followers. So is the number of tweets [5]. Users who post less than 10 tweets are often to see, but those who post hundreds of tweets per day are rare.

3.2 CLASSIFICATION OF SOCIAL NETWORK

Now that we have taken a general look at what is a social network, let us learn more about the classification of social networks and the contents we are going to predict.

3.2.1 Narrow-Sensed Social Network

When it comes to social network, Facebook, Twitter, and Weibo are the most famous and frequently used ones [3]. These types can be referred to as narrow-sensed social network. The very characteristic of these social networks is the focus on communication. Users in these social networks share their own lives or interesting topics via fresh news, photos, videos, logs, etc. Their friends and followers read these textual contents and show their positive or negative opinions on them. In general, these social networks are widely used by people who are real-world friends or who are eager to understand more about celebrity news.

3.2.2 News-Based Social Network

News-based social network concerns news. News is packaged information about new events happening somewhere else. Nowadays, news is no longer

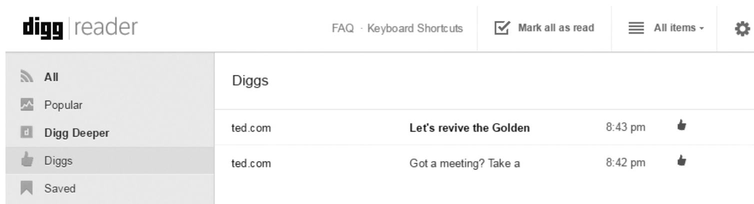


FIGURE 3.3 The user interface of Digg.

sold in the streets. They are in electronic versions, distributed to websites. Despite large amounts of news websites over the world, few of them can be called social networks.

Digg is a non-intuitive example (changes a lot in the past few years, the GUI is in Figure 3.3). Digg is a news aggregator with an editorially driven front page, aiming to select stories specifically for the Internet audience such as science, trending political issues, and viral Internet issues. News in Digg cannot only be created by professional journalists, but also by everyone who uses Digg by uploading news around them. This is why Digg can be defined as a social network. Users in Digg have friends (in specific, followings and followers). They can read news created by their friends, digg it (vote for it), and comment on it. That is where the interactions happen, which is an important standard to judge whether a news website can be identified as a social network. There are some studies carried out on Digg [25, 35, 40].

Reddit and Hacker News are other vivid examples of news-based social networks. Reddit is quite similar to Digg. Registered users of this website, usually referred to as "redditors", can vote "up" and "down" for the news and decide their positions on the site's page. Hacker news is a little bit different. We can easily tell from the name of Hacker that this social news website focuses on computer science and entrepreneurship. What is more, there is no option to down-vote the post which means users can only up-vote it or not vote on it. In general, content that can be submitted on Hacker News is defined as "anything that gratifies one's intellectual curiosity". No matter how distinct these websites are, the elements of interaction and communication make them the news-based social networks.

3.2.3 Major-Based Social Network

When it comes to major-based social networks, LinkedIn must be the most well-known one. LinkedIn (in Figure 3.4) is a business-oriented social network. The basic functionality of LinkedIn allows users to create profiles and establish relationships with each other in an online social network, mapping out real-world professional relationships. Compared with other social networks, the relationship in this special social network is more important in that users can seek jobs, people, and business opportunities recommended by someone

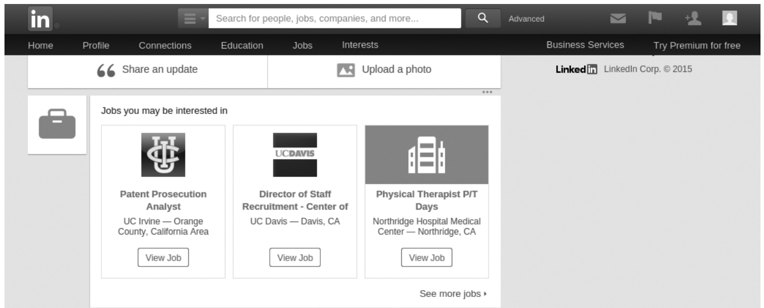


FIGURE 3.4 The user interface of LinkedIn.

in one's contact network. LinkedIn is called a major-based social network in that it works mainly for employers and job seekers. By using this social network, employers can list jobs and search for potential candidates by browsing candidates' profiles, which act like resumes in some way. Of course, job seekers should fill their profiles with their professional skills ahead of time and seek appropriate jobs.

Another kind of major-based social network is non-intuitive, which is used for single men and women to seek a mate. This kind of social network is really popular in China, such as Zhenai Net, Baihe Net, Shijijiayuan Net, etc.

Various kinds of social networks are used in research papers in experiment section. We give a statistic graph (Figure 3.5) after our survey on social networks used from year 2010 to 2013.

It should be mentioned that there are probably more than one social network used in one paper.

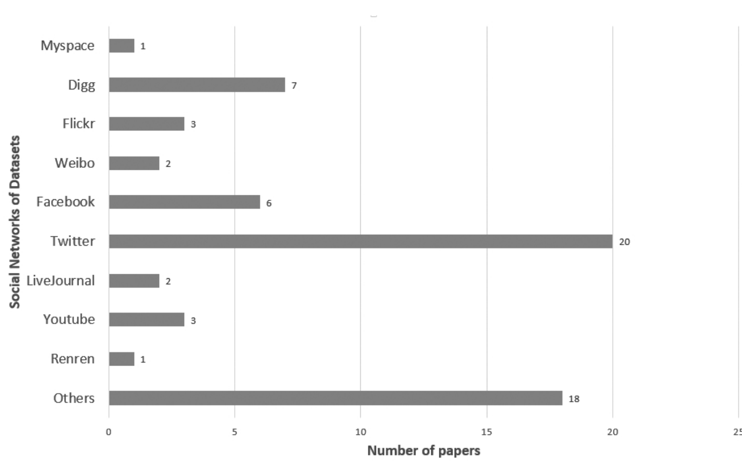


FIGURE 3.5 Social network used in research papers in recent years.

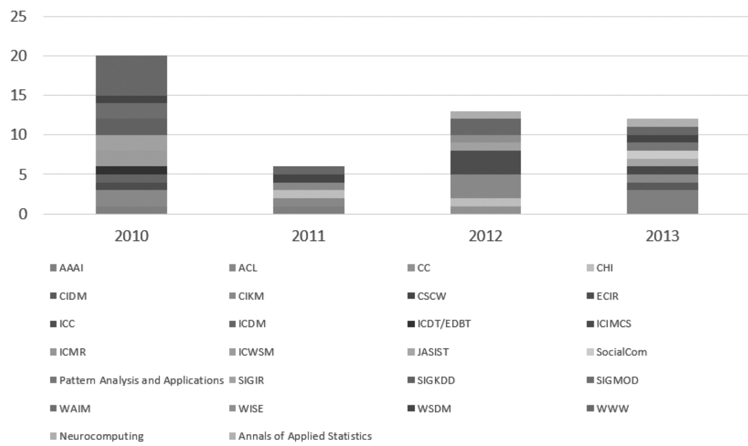


FIGURE 3.6 Number of papers in related conferences every years.

Furthermore, in Figure 3.6, we analyze the number of papers in conferences which pay attention to predict popularity of online contents every year.

3.3 PREDICTION MODEL

Our goal is to research the popularity prediction of textual content, so we shall investigate prediction models involved in this area. In this section, it first gives a brief description of the feature selection. After that three aspects worth considering in prediction models are followed.

3.3.1 Feature Selection

When establishing a prediction model, usually we need to find some elements or aspects based on which we can make a reliable prediction. We refer to these elements or aspects as features. There can be various features for prediction but not all of them contribute to the prediction result. Thus, selecting relevant and supporting features is crucial to the accuracy of prediction. Since the piece of datum in a social network usually contains many redundant or irrelevant features, we should implement some feature selection technique. At the very beginning, we may enumerate several features from different aspects. The most common features that we can think of in social network prediction includes length of a tweet, the number of words, etc. To filter useful features, different scenarios are proposed. We will explore them successively and conclude at the end of this subsection.

Before the overture, we should realize that feature selection technique is to be distinguished from feature extraction. Feature extraction creates new features from functions of the original features, whereas feature selection returns a subset of the features. The typical method that feature extraction applies would be Principal Components Analysis (PCA). Both selection and extraction are recommended, but we prefer selection because our features are mostly independent.

3.3.1.1 *Mature Tool*

We may first try some general methods to select features. Arunee Ratikan [33] uses Weka module to select attributes. Weka is a statistical tool that provides multiple functions. All of Weka's techniques are predicated on the assumption that the data is available as a single flat file or relation, where each data point is described by a fixed number of attributes. Weka builds an attribute selection function based on [4]. You can use it directly as a blackbox for its effectiveness and convenience.

3.3.1.2 *Correlation-based Method*

Other than mature tools, the most natural way to do feature reduction is correlation-based method. There are several evaluation measurements and the detailed processes are varied.

The main point of correlation is to score features via experiments. For instance, Marilyn A.Walker [29] calculates Pearson's correlation coefficients between Linguistic Inquiry and Word Count (LIWC) features and personality ratings. It can be acquired by Equation (3.1) given two variables X and Y as follows:

$$\rho = Cor(X, Y) = \frac{Cov(X, Y)}{\sigma_X \sigma_Y} \tag{3.1}$$

where Cov is the covariance and σ_X is the standard deviation of X. The result falls between +1 and -1 inclusive, where 1 is total positive correlation, 0 is no correlation, and 1 is total negative correlation.

Besides the algorithm, we may employ some tactics for evaluation. The simple examples can be found in Rushi Bhatt's work [7]. Their models effectively combine user features and social features to predict adoption, which is better than using either user features or social features individually. A more complex instance is from Shoubin Kong's work [21]. It presents the evaluation method as F_1-score and gains the importance of features by dropping them one by one. The most influential one is obtained when the result is terrible without it.

3.3.1.3 *Unique Method*

It is also encouraged to cope with data differently under various circumstances. One intuitive case is Tim Paek's work [30], whose features are enormous. They reduced the number of features to the top $3K$ features in terms of log likelihood ratios as determined on the training set. A likelihood ratio test is based on the likelihood ratio, which expresses how many times more likely the data are under one model than the other. This likelihood ratio, or equivalently its logarithm, can then be used to compute a p-value. When it comes to the feature selection, Dunning first combined them together [11] and it has been widely used ever since.

Moreover, Hila Becker [6] involves an ensemble algorithm, which considers each feature as a weak indication of social media document similarity, and combines all features using a weighted similarity consensus function. Ensemble clustering is an approach that combines multiple clustering solutions for a document set. It enhances its ability to account for different similarity metrics.

Selecting features is not a necessary process, but it could improve the accuracy of the prediction result and reduce the computation complexity. When we encounter this problem, the suggested methods are mature tools and correlation-based method. Meanwhile, when we want to achieve some specific aims like choosing the most important feature, some special methods are also promising if designed properly. Also, several special cases illustrate the diversity of this issue.

3.3.2 Text Content

Before we turn to predicting models, there is a classification about the content we predict that should be clarified. Online contents that we would like to predict can be categorized into three main kinds: something, news, topics, events.

Something is a single message user post on social network. It's something new that a user wants to share with his social friends. *Something* can be either a tweet in Twitter, or a story in Digg.

News narrowly refers to news that is broadcast on TV or online newspapers.

Topics usually has some key words. And many *something* which all contain these key words consist of a topic. The contents of a topic could change from time to time. For example, we talked about the famous singer Taylor Swift yesterday because of her new song written to another ex-boyfriend, but we talk about her new song today because it wins a Grammy Award.

Event is defined as an activity or action with a clear, finite duration given a particular target entity which plays a key role in the duration [32].

3.3.3 Predicting Models

Predicting model is the soul of prediction. It is the most important element in this research area. A good model will not only emerge the inner characteristic of training data, but also predict what is popular in the future precisely and timely. The fundamental idea of constructing a model is to take possible and significant elements into account by combining them linearly with static or dynamic weights so that it can make the prediction with history information.

The difficult part of creating a prediction model comes when deciding what factors to include and how to represent them.

In this section, we are going to introduce several factors and methods of various models taken for prediction. We classify them into several types: user behaviors, life cycles of textual contents, and network topology. Notice that these methods are not totally independent. Actually you can find some of them together in one prediction model.

3.3.3.1 Prediction Based on User Behaviors

User behavior in social networks covers various activities that users can do online, including expressing new opinions, commenting, sharing, browsing and so on [18]. For instance, a typical user of Twitter may deliver a new tweet about the good weather, thumb's up a friend's tweet, share a funny story, and do not forget to update a new selfie. Every new update and every interaction with people in the social network makes up a series of user behaviors.

Notably, user behaviors, in different social networks, can be different. The most common behavior in Twitter is creating new tweets and retweeting others, but on other news and story websites like Digg and Flickr, page browsing, commenting, sharing, and other post-read actions [2] happen more often. User behaviors are dynamic. Users change their behaviors from time to time so user behaviors are usually represented in a function with time t as a variable because that is how these data are collected. For example, Kong et al. [21] monitored the time series of all hashtags and predicted the burst and popularity of them. We then introduce several examples to illustrate how to use user behaviors for prediction.

Box-Office Revenue Prediction: Asur and Huberman [4] tried to predict the box-office revenues by studying the tweets on Twitter. This study defines a Tweet-rate to reflect the user behavior of posting tweets of a movie prior its release. The definition of Tweet-rate is shown in Equation (3.2):

$$Tweet - rate(mov) = \frac{|tweets(mov)|}{|hours|} \tag{3.2}$$

This definition means the number of tweets referring to a particular movie per hour. Then, the revenue is predicted based on the defined Tweet-rate in Equation (3.3).

$$Rev(mov) = \beta_0 + \sum_{i=1}^{7} \beta_i * Tweet\text{-}rate_i(mov) + \beta_{th} * thcnt, \tag{3.3}$$

where *thcnt* means the number of theatres the movies are released in and β values represent the regression coefficients which is learned from historical records. Let us take a deeper look into this model. This model is a linear regression of tweet-rates in 7 days before the release of a particular movie. It collects tweets about a movie and views them as user behaviors on social network and uses them to predict the revenues. Different β_i are included so as to find out the different impacts tweet-rates on each day have on the result. This is one simple example of applying user behaviors on prediction.

News Popularity Prediction: Here is another example using the number of tweets to make the prediction for news popularity [5]. Bandari et al. chose 4 types of predictors to predict the popularity of news. They first defined a feature, named *t*-density, to represent the "popularity" of news per link using the number of tweets, as shown in Equation (3.4).

$$t\text{-}density = \frac{\texttt{Number of Tweets}}{\texttt{Number of Links}}, \tag{3.4}$$

where the "Number of Links" represents the links in a news category. The authors considered 4 types of predictors, *Category Score* represented by *t*-density of a category, *Subjectivity* obtained from an existing subjectivity classifier, *Name Entities* (a known place, person, or organization) obtained from an existing extraction tool, and *Source Score* represented by *t*-density of each source, to predict the popularity of the news with regression algorithms and the result using linear regression as shown in Equation (3.5):

$$ln(T) = 1.24ln(S) + 0.45ln(C) + 0.1Ent_{max} - 3 \tag{3.5}$$

where T represents the number of tweets, S represents the source *t*-density score, C represents the category *t*-density score, and Ent_{max} is the maximum *t*-density of all entities found in the paper.

User Behavior Prediction: The previous example takes advantage of user behaviors to make prediction of something outside of social network, while user behaviors can directly be used to predict user behaviors. Lee et al. [23] predicted total comments from early comments and Castillo et al. [9] predicted the visits from early visits, to name just a few.

Among numbers of researches, Huang et al. [15] is a typical example. They applied their model, the *Parameterized Social Activity Model* (PSAM), to predict the tendency of social activities (denoted as N_t). Here social activities can be something like becoming a member, posting a message, adding a comment, inviting a friend, etc., which just matches the user behaviors that we are discussing.

To predict the trend of the evolution of a social activity, this model uses a continuous-time stochastic process. The trend of N_t consists of two parts, one is the tendency of evolution, and the other is the random impact from the environment. The authors derived a parameterized stochastic process with a drift and diffusion as in Equation (3.6).

$$dN_t = \gamma(\overrightarrow{\lambda_t})N_t dt + \sigma(\overrightarrow{\lambda_t})N_t dW(t) \tag{3.6}$$

where t is the continuous time, $\overrightarrow{\lambda_t}$ is the time-involved activity features vector. The first part of Equation (3.6) indicates the growth or shrinkage of activities in a social network, while the second part is the random impact, which we emphasize here. In this part, the Wiener Process (WP, which is also called Brownian Motion) is used. WP has a good performance on predicting a near future using data from the present. (However, it cannot predict long time future because the expectation of the future is exactly the same as the expectation of the present. That is to say, WP has no tendency to increase or decrease in the future.)

Finally, for this model the authors evaluated three publicly available datasets: Facebook-wallpost dataset [39], Facebook friend-request dataset [39], and Citeseer co-authorship dataset [14]. The experimental result shows that the model has more than 0.8 accuracy which indicates the effectiveness of it.

Story Popularity Prediction: Apart from some easily recognized user behaviors, page browsing, as another kind of major user behavior in social news portal, can also be considered as part of the predicting model. According to [16], the distribution of the number of pages a user visits before leaving the web site can be modeled by a two-parameter inverse Gaussian distribution in Equation (3.7):

$$P(L) = \sqrt{\frac{\lambda}{2\pi L^3}} \exp\left[\frac{-\lambda(L-\mu)^2}{2\mu^2 L}\right] \tag{3.7}$$

where L is the random value of the number of links a user follows before the page value first reaches the stopping threshold. And mean $E(L) = \mu$, variance $Var(L) = \mu^3/\lambda$, where λ is a scale parameter.

Using this theorem, Lerman et al. [26] proposed a model of social dynamics to predict popularity of news based on the social network Digg. The standard for popularity is the number of votes a story in Digg gets. Moreover, the number of votes a story receives depends on the combination of its visibility and interest, with visibility coming from different parts of the Digg user interface (the friends interface $v_{friends}$, upcoming v_u, and front pages lists v_f) and the position within each list. Therefore, the Rate Equation for the number of votes a story gets N_{vote} is shown in Equation (3.8):

$$\frac{dN_{vote}(t)}{dt} = r(v_f(t) + v_u(t) + v_{friends}(t)) \tag{3.8}$$

where r is the probability a user seeing the story will vote on it. To represent v_f and v_u, we use the inverse Gaussian distribution in Equation (3.7), and get Equation (3.9):

$$v_f = v f_{page}(p(t))\Theta(N_{vote}(t) - h) \tag{3.9}$$

where f_{page} models user behaviors of browsing web pages, that is a decreasing visibility of stories on a page from upper to under, which refers to the inverse Gaussian distribution. And $p(t)$ is the position of a story in page p at time t.

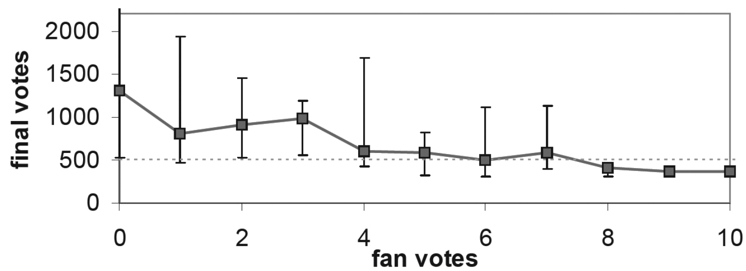


FIGURE 3.7 Number of fan votes within the first 10 votes VS final votes received by front page stories [26].

This model is extremely intuitive in sync with common sense. It derives human activity with precise mathematical models. Additionally, the result the model gets is similarly logical and reasonable. In another paper of Lerman's work [24], it derives that votes from fans of the submitter or previous voters ultimately go on to accumulate fewer votes than stories that initially receive few fan votes (Figure 3.7). That is because a story that is of interest to a narrow community will spread within that community only, while a generally interesting story will spread from many independent sites as users unconnected to previous voters discover it with some small probability and propagate it to their own fans.

3.3.3.2 Prediction Based on Life Cycles

Besides studying user behaviors alone, much research studies how one piece of information spreads in the social network, how it starts to be popular, and how soon it disappears. We describe the duration of a piece of information in the social network as life cycle. We believe that majority of social network users have such an experience: a topic suddenly appears out of nowhere and everyone talks about it immediately, but just two or three hours later, it has been forgotten and another new issue takes its place. In other words, some information spreading in social network bursts out suddenly and disappears quickly. If we can determine the pattern beforehand, we can distinguish among topics that will be popular and those that will not. By studying the life cycle, approximate prediction can be made. Combined with other factors like user behaviors, the prediction could be accurate.

Discussion on the Types of Life Cycles: Research on life cycles shows that different types of textual contents have every distinct life cycle pattern. In [9], the researches come up with a qualitative and quantitative analysis of the life cycle of online news stories. It probes into the Al Jazeera

English, a large international news network, finding out two life cycle patterns of two types of stories, News and In-Depth. The pattern for the News article can be roughly described by an "80:10:10" rule. To be specific, 78% of News articles decrease either at once or after a short delay when published. About 12% of News articles stay steadily or increase during the first 12 hours. The other 10% of the articles initially decline in visits per minute, until a special moment when the decline reverses. The authors called the phenomena *rebounding*. On the other hand, In-Depth articles sustain a level of visits during several hours and are not as time-sensitive as the News articles. Seeing the life cycle patterns, the authors designed a linear regression model using several social media features. The model can be represented as Equation (3.10) and Equation (3.11):

$$lm(vis7d \sim v) \tag{3.10}$$

$$lm(vis7d \sim (v + vr + vd + f + t + foll + ent + uni + unip + cp)^2) \tag{3.11}$$

where $vis7d$ means the predicted total visits after 7 days, and the features include number of visit (v), number of visits from link referrals (vr), email/IM (vd), shares on Facebook (f), Twitter (t), the mean number of followers of people sharing on Twitter ($foll$), the entropy of tweets (ent), the number and fraction of unique tweets (uni, $unip$), and the fraction of corporate retweets (cp). This model considers second-order interactions because some of the variables are interdependent.

Life Cycle with Influence Decay: Another work [20] focuses on the decay of information and uses the build novel model on it. It reranks the news from three datasets, Google News Archives (SES), dataset from web portals in China (WPS), and earlier data with the same resource as WPS (TS). The core idea is to include the influence decay factor so that the life cycle of the news can be considered into the model. Together with features from media focus and user attention, the authors used this model to rank the news again. They adopted a sigmoid function to simulate the influence decay, as shown in Equation (3.12):

$$f(x) = \begin{cases} \frac{\delta x}{1+\delta x}, & x > 0; \\ 0, & otherwise. \end{cases} \tag{3.12}$$

where δ is a smoothing parameter. The simplest model appears as:

$$I_d^{\varepsilon} = f(r_d^{\varepsilon}), \tag{3.13}$$

where I_d^{ε} represents the influence value of topic ε and r_d^{ε} is the report-frequency of topic ε during the dth day. Of course this model is too simple, so the authors modified it bit by bit. They remodeled it by calculating the accumulative effect of influence decay by adding continuous d days report-frequency together. What is more, they found that the decay itself can be dynamic from day to day so they tried to deride the influence decay factor from current situation and from Ebbinghaus Curve in Equation (3.14).

$$decay(d) = (1 - \beta_d^{\varepsilon})^{\lambda}. \tag{3.14}$$

Equation (3.14) is a decay function where $decay(d)$ represents the decay function on the day d and β_d^{ε} is the decay rate which can be obtained from sigmoid function with report-frequencies of previous days, as shown in Equation (3.15):

$$s_d^{\varepsilon} = (\sum_{i=1}^{d-1} \alpha_e r_i^{\varepsilon} ln(\frac{\beta_e}{d-i}) + r_d^{\varepsilon}) \tag{3.15}$$

where the $ln(...)$ part is learned from the famous Ebbinghaus forgetting decline, just like memory loss.

Life Cycles for Hashtag: An Example. The study of life cycle proves to be very useful in predicting bursts and popularity. Hashtag is the very example that can be modeled upon the analysis of life cycle. Hashtags, "♯" followed by a word [22], reflect the hidden trend of Twitter. If the next burst of Hashtag can be predicted beforehand, we can make good use of it. Kong et al. [21] studied exactly the burst, as well as the popularity of hashtags, by analyzing the life cycle of hashtags.

The authors of [21] provided formal definitions of four states in the life cycle of a bursting hashtag.

- **State 1, Active**. If all tweets containing the hashtag reach a threshold ϕ posted within four continuous time intervals after a time point t_a, it is defined that the hashtag becomes active since t_a.
- **State 2, Bursting**. Within 24 hours after a hashtag becomes active, if the number of tweets containing the hashtag during a time interval t becomes greater than the $\max(C_1 + \delta, 1.5C_1)$ where C_1 is the number of tweets containing the hashtag in the first time interval after the hashtag appears, it is defined that the hashtag bursts.
- **State 3, Off-Burst**. For the burst hashtags, if the number of tweets containing the hashtag drops bellow $\max(C_1 + \delta, 1.5C_1)$ in the following 24 hours after time interval t', it is defined that the hashtag becomes off-burst.
- **State 4, Inactive**. If the hashtag cannot trigger another active state, then it is defined as inactive. In [21] it was determined that among all hashtags, 95% burst within 6 hours since they become active; 96% become off-burst within 24 hours since active and about 98% are inactive within 48 hours since active.

With the findings of the life cycle of bursting hashtags, [21] makes attempts on five regression models including Linear Regression, Classification and Regression Tree, Gaussian Process Regression, Support Vector Regression, and Neural Network to make the prediction of whether a hashtag will burst and how long it will last if the hashtag bursts. Its result shows that Gaussian Process Regression appears to have the best performance and the correct prediction of bursting hashtags can be made 55 minutes on average earlier than the start of the burst.

The analysis of life cycle is fit to make predictions in social network because the message spreading in the network has its unique features which should be thought of when building the prediction model.

3.3.3.3 Prediction Based on Network Topology

The common way to solve problems within a network is to study its topology and the social network is not an exception. Indeed, the study of prediction based on network topology is traditional compared with user behaviors and life cycles of a piece of information because the analysis of network topology is widely used in other fields of study if a network is involved.

Another characteristic of network topology different from user behaviors and life cycle is the stableness. The topology is more static. The network has already been built before an event happens or a piece of news bursts, thus the topology features can be extracted using existing user profiles, following relationships or other hidden but existing qualities. First, let us take a look at the widely accepted formal concepts that work in the topology of social network and then view some examples.

The whole social network can be represented as a graph $G(V, E)$ with nodes V and edges E.

- **Node:** Every user in the social network is denoted as a single node. The number of nodes is usually denoted as $|V|$.
- **Edge:** If user A follows user B, there is a directed edge from user A pointing to user B. Since the following relationship is bidirectional, user B has an edge pointing to user A if user B follows user A back. The number of edges is usually denoted as $|E|$.
- **In-Degree:** The In-Degree of user i is the number of edges pointing from other users to user i. In other words, the number of users who follow user i.
- **Out-Degree:** The Out-Degree of user i is the number of edges pointing from user i to other users. In other words, the number of users who user i follows.

Discussion on Node Attribute. Among the topological attributes, the node attribute is the most frequently used in various models because of its generality and applicability. The In-Degree and Out-Degree can be straightforwardly included into the qualities of nodes and the qualities of following relationships somehow can also be covered with the qualities of nodes with some ideas like density or similarity. What is more, the qualities of nodes can be applied in the establishment of models.

For example, Hu et al. [13] found out that celebrities have a great influence on spreading Osama Bin Laden's death in Twitter which helps explain that the qualities of some specific nodes can indeed be used in prediction and the life experience tells us that people are interested to read a tweet that is retweeted by President Obama.

Similarity between a user and a topic and user influence are concrete examples of using the qualities of nodes to make fine predictions in social network. We can see both qualities in Zhang's work [45].

Zhang et al. [45] provided a model, based on user influence and user's interest in a certain topic, to predict popularity of a burst event in Weibo, a Chinese social network. The volume of micro-blogs that discuss the event at time $t+1$ is modeled as a linear function (3.16):

$$V_e(t+1) = \sum_{u \in In(t)} Sim(u,e) * F_u(t - t_u) + \chi_{t+1} H_e(t+1), \tag{3.16}$$

where $Sim(u,e)$ represents the user u's interest in event e and $F_u(t-t_u)$ denotes user u's influence power after he is infected. The $Sim(u,e)$ is the underlying qualities that can be obtained from the network topology. In [45], the authors took advantage of Latent Dirichlet Allocation (LDA) model [8], an unsupervised machine learning technique to identify the topic information from a large set of documentation. In this example, the user u's history profiles and all collected tweets are about the event e. The LDA model produces a probability distribution over latent topics of the user and the distribution over each topic e. By taking the cosine similarity, the user's interest in each topic is reflected. The key point to Equation (3.16) is to train user influence so that it can be used to make prediction of the volume.

From all kinds of models we have discussed, we have a key observation that plenty of models are derived from search engines and blogs. It is reasonable, because social network is kind of evolved from these two branches. Whenever we long for innovation, never forget to look back for previous precious experience.

At the end of this section, we are able to summarize and category state-of-the-art papers into Table 3.1. This may give a general idea of predicting popular online content.

TABLE 3.1 State-of-the-art of predicting popular online content

Reference	Predict Model	Dataset	Content Class	Social Network Level	Details
[19]	User behavior	3500 stories with 2.1 million votes and 1.7 million social links in Digg	Something	Meso	Concentrate on user's attention
[10]	Life cycle	10G data from SNAP and Inforplease	Topic	Macro	Probabilistic prediction using Bayesian likelihood function
[13]	Life cycle	614,976 tweets containing string "laden"	News	Macro	Media people, mass media and celebrities help the propagation of news in Twitter
[32]	User behavior	104,713 celebrities from Twitter	Event	Macro	Use 3 models and gold standards for classification
[14]	User behavior	35,809 posts and their associated comments from Facebook	Something	Macro	Combine content specific features, author specific features and temporal activity features to predict
[28]	Life cycle	12 million Singapore Twitter users with 31 million tweets	Topic	Macro	Regards hashtag as a sort of classification problem and evaluates 7 content features as well as 11 contextual features
[43]	Life cycle	Through Twitter Search API	Something	Macro	Set up a Bayesian probabilistic model to analyze the trend of retweet
[44]	User behavior	16 million tweets, 2.6 million unique words and 148 popular events	Event	Micro	Detecting burst events and clustering burst words
[21]	Life cycle	Not known	Topic	Macro	Divides a hashtag into several statuses
[34]	User behavior	Over 3 million tweets from 250,000 in California, New York and Texas	News	Micro	Analyze the preference for news in different area in order to rerank the top 10 news
[26]	User behavior	510 stories on Digg	Something	Micro	A statistical model with probabilistic user online behavior

3.4 EVALUATION

3.4.1 The Importance of Evaluation

Once a novel prediction model is created, no one can guarantee that it will make the perfect prediction. Intuitively, we want to know whether the predicted result is correct, to what extent and whether it outperforms previous models if there exist some. At this moment, particular methods should be utilized to help scholars statistically analyze the effectiveness and performance of the model. This step is called evaluation. Without it, all the discussion about the model is impractical and not convincing. A good evaluation method reflects the abstract aspects of the model in concrete statistic figures and thus helps analyze it.

3.4.2 Evaluation Metrics

There does not exist one single evaluation metric applicable to all models but generally, some metrics are widely noticed in predictions in social network. Before we talk about specific evaluation metrics, we first distinguish three prediction goals:

1. Ranking prediction — predict the result in the form of a ranking list.
2. Classification prediction — predict the classification.
3. Numerical prediction — predict the exact value.

For each prediction goal, we sort out respective evaluation metrics that are widely used in predictions in social network.

3.4.2.1 Ranking Prediction

Some papers predict the influence or popularity in the form of ranking. [20], [37] and [34] predict the popularity of news events. [42] predicts the users who will be influenced by a focal user. To evaluate the quality of a ranking list, we can apply methods in information retrieval. $nDCG$ and AP are good choices.

Discounted Cumulative Gain (DCG): Discounted Cumulative Gain, abbreviated as DCG, is often used to measure the quality of web search engine algorithms but basically, it measures ranking quality.

Given a ranking list L, each item in L has a graded relevance which represents the true value of the item. Denote the graded relevance at position i in L as rel_i. The DCG at a particular position p is defined as:

$$DCG_p = rel_1 + \sum_{2}^{p} \frac{rel_i}{log_2(i)} \tag{3.17}$$

The equation above takes into account the penalty of the higher graded relevance appearing in a lower position as a reduced graded relevance logarithmically proportional to its position.

An alternative formulation of DCG appears as:

$$DCG_p = \sum_{1}^{p} \frac{2^{rel_i} - 1}{log_2(i+1)} \tag{3.18}$$

$nDCG$, the normalized DCG, qualifies the measurement within 0 and 1 by comparing the DCG of predicted ranks with the DCG of the ideal one. It's defined as:

$$nDCG_p = \frac{DCG_p}{IDCG_p} \tag{3.19}$$

where $IDCG_p$ is the maximum DCG_p provided the graded relevance.

With $nDCG$, we can compare the ranking lists of different grading standards. When evaluating the ranking models, usually $nDCG$s at different positions are calculated so as to see the consistency and applicability of the model. For example, in [37], $nDCG@1(nDCG_1)$, $nDCG@5$, $nDCG@10$, $nDCG@20$, $nDCG@100$ are calculated and compared.

Average Precision (AP): Apart from DCG and $nDCG$, AP (average precision) is the other useful evaluation metric for rankings. For example, [32] uses AP to evaluate the ranking result of controversial events.

Before the presentation of AP, we first define the precision at position p.

$$Precision(k) = \frac{relavant\ \ documents\ \ in\ \ top\ \ k}{all\ \ documents\ \ in\ \ top\ \ k} \tag{3.20}$$

Then, AP can be defined as:

$$AP = \frac{\sum_{k}^{n} = 1(P(k) \times rel(k))}{number\ \ of\ \ relevant\ \ documents} \tag{3.21}$$

where $rel(k)$ is an indicator function equalling to 1 if the item at position k is relevant, zero otherwise.

AP evaluation metric is especially useful when ranking the extent of binary classification items like controversial events in [32] because it is quite difficult to give a concrete number to represent how controversial the event is.

3.4.2.2 Classification Prediction

There are some classification tasks in prediction in the field of social network like the binary classification problems of users' interest in news recommendation [17], and other classification of personality prediction [27]. The evaluation metrics used for classification is quite limited and confusion matrix is mostly seen.

TABLE 3.2 An example of confusion matrix

	Predicted class		
Actual class	A	B	C
A	5	3	0
B	1	6	0
C	1	2	5

Confusion Matrix. Confusion matrix divides the result for each class in the classification into four categories:

1. True positives(TP): correctly identified
2. False positives(FP): incorrectly identified
3. False negatives(FN): incorrectly rejected
4. True negatives(TN): correctly rejected

For example, we now have a classification result as Table 3.2. Samples are classified into three classes, A, B and C. We can calculate the TP, FP, FN and TN for each class. Take class A as an instance, the TP, FP, FN and TN for class A are: $TP = 5$ (A classified as A), $FP = 2$($\overline{A}$ classified as A), $FN = 3$ (A classified as $\overline{A}$) and $TN = 13$ ($\overline{A}$ classified as $\overline{A}$).

For each of the binary classification (like A and not $\overline{A}$), define the five frequently used indexes to evaluate the performance of classification prediction.

$$accuracy = \frac{TP + TF}{TP + TF + FP + FN} \tag{3.22}$$

$$precision = \frac{TP}{TP + FP} \tag{3.23}$$

$$sensitivity(recall) = \frac{TP}{TP + FN} \tag{3.24}$$

$$specificity = \frac{TN}{FP + TN} \tag{3.25}$$

$$F_1 = 2\frac{precision \cdot recall}{precision + recall} \tag{3.26}$$

Accuracy is used to express the percentage of correctly classified from all instances. Precision reflects the percentage of correctly classified from all positive instance. Precision has a significant effect when dealing with imbalanced classification where most of the instances belonging to the same class leads to high accuracy even without classification. Sensitivity, also called "recall", expresses the percentage of correctly identified actual positive. The F_1 score (also F-score or F-measure) is a measure of a test's accuracy using precision and recall. An F_1 score reaches its best value at 1 and worst score at 0. In [36], precision and F_1 score are both applied to test the prediction result of negative links in social media. Specificity shows the percentage of correctly identified actual negative. The application of confusion matrix and these five indexes are very classical in evaluating classification problems.

3.4.2.3 Numerical Prediction

Numerical prediction is mostly used compared to the other two because when predicting the popularity or influence in social network, a certain value is predicted to represent the popularity or influence, such as the box-office revenue [4], the number of votes for a piece of story [26] or scores for 5 aspects of personality [12]. Basically, the evaluation metrics for the numerical evaluation can be divided into two kinds: correlation-based method and error-based method.

Correlation-based Method: Correlation-based method tries to find the relationship between two arrays of variables.

1. The Pearson product-moment correlation coefficient

   For linear relation, the Pearson product-moment correlation coefficient is the most widely used.

   The Pearson product-moment correlation coefficient(r) measures the linear dependence between two variables X and Y. r ranges from -1 to 1, where 1 means positive correlation, 0 means no correlation and -1 means negative correlation. In practice, X can be the predicted value vector and Y can be the real value vector.

   The Pearson correlation coefficient is defined as:

   $$Corr(X,Y) = \frac{\sum_{i=1}^{n}(x_i - \overline{x})(y_i - \overline{y})}{\sqrt{\sum_{i=1}^{n}(x_i - \overline{x})^2 \sum_{i=1}^{n}(y_i - \overline{y})^2}} \tag{3.27}$$

   where n means the total number of elements in X and Y, $\overline{x} = \frac{1}{n}\sum_{i=1}^{n} x_i$ and $\overline{y} = \frac{1}{n}\sum_{i=1}^{n} y_i$.

   For example, in [26], it uses the Pearson correlation to estimate the relations between predicted votes and the final actual votes for every piece of story. [12] and [15] are other examples using the Pearson correlation.

2. Coefficient of Determination

   More generally, the coefficient of determination, denoted as R^2 or r^2 when the model is simple linear regression, represents the percent of the data that is the closest to the line of best fit. R^2 ranges from 0 to 1, and the better values fit the model, the closer R^2 is to 1. [4] uses coefficient of determination to find out how the predicted box-office revenue fits the reality. [5] uses this to evaluate the performance of Linear Regression and SVM Regression in predicting the popularity of a piece of news. The R^2 is defined as follows:

$$R^2 = 1 - \frac{\sum (x_i - y_i)^2}{\sum (y_i - \overline{y})^2} \tag{3.28}$$

where x_i can be the predicted value, and y_i is the real value.

Error-based Method: Error-based method evaluates the model by analyzing the error between predicted results and actual values statistically. The closer the calculated figures of error come to zero, the better the performance of the model is. The commonly used error-based metrics include:

$$RMSE = \sqrt{\frac{1}{n} \sum (x_i - y_i)^2} \tag{3.29}$$

$$MAE = \frac{1}{n} \sum_{i=1}^{n} |x_i - y_i| \tag{3.30}$$

RMSE indicates the root of the mean value of squared errors. MAE indicates the mean absolute error. In [38] MAE evaluates the performance of using the number of tweets mentioning a party as a predicator of the election result and other election polls and help prove the usability of using Twitter in election prediction.

3.5 LOOK FORWARD

The popularity prediction problem gets hot for a reason – making precise and accurate prediction brings us much good. Both inside the scope of social media and using similar methods in other fields, the problem proves its value in different scenarios. This chapter introduces several practical usages of the methods and proposes future outlooks that will bring advantages to different people.

Intuitively, we can optimize the web contents organization through predicting their future popularity, putting potential contents at conspicuous places. People love stuff online, while they can only cover a small part compared to all resources available. Nowadays, some websites hire web editors to help them dig up hot topics, events or articles daily from huge resource bonanzas, which usually contain millions of objects. However, this work takes time and money, with no guarantee for the chosen content to be popular in the future. With prediction techniques, websites can exempt themselves from all the expenses and refine the decision faster and better as well. Once the contents are set properly, sites' traffic will increase significantly. Here is a practical example. Agarwal et al. from Yahoo! Inc. succeeded in boosting the number of click counts for Yahoo! news by leveraging automatic selection algorithms which estimate users' interest in new articles [1].

Online marketing also benefits much from the prediction of web contents. We are now living in a society where companies spend around 30% of their budget to do online marketing. If a company pays money advertising proper items and stops those that are not promising in the future, it can then maximize its profit and save the budget. For instance, movies or shows can be studied and predicted, helping investors make a decision whether to put money in it or not. Netflix studied its big data pool and predicted that the show, House of Cards, would be a big hit. Therefore, the company invested in the show and promoted it, which has now proved to be a huge success. Moreover, predicting and scheduling can also be a real-time process. By monitoring the fluctuation along time, a company is able to dynamically change its strategy and adapt its decision according to the changing marketing environment, ensuring that it takes proper measures before the loophole is noticed and the damage is made.

Information predicted and gathered through social media even helps prevent potential social panic. Sometimes, false information permeates the social media and brings much trouble. Those pieces of rumors will be powerless unless they reach a certain amount of attention. If we can predict which event will burst, we hold the initiative. In April 23, 2013, Associated Press Twitter account was hacked by someone who sent the rumor that the president was injured due to explosions at the White House [21]. Though debunked later, the rumor became a widely spread hit topic shortly and thus fiercely shocked the stock market at that time. The Dow Jones Indice dropped by over 100 points, bringing irreversible damage to the market. However, if we had employed prediction skills, we would monitor the topics that were about to burst, and thus made timely measures to stop the false claims. Kong et al. wrote a paper on this topic, using different features to learn and predict whether a hashtag will burst in the future [21].

Prediction is capable of coming out of the virtual world and predicting a user's personality. The social networks provide us with abundant information about a person, from which we can predict a user's personality, with pretty good accuracy. We can thereby properly present different contents or services to different categories of people according to their personality, trying to make everyone happy. Furthermore, we can detect sociopath or psychopath using the same method [31]. These people often show anomalies during personality prediction tests, allowing us to find them by several criteria. By detecting them beforehand, we may prevent the potential hazard they will bring and make our cyberspace a safer place.

The predicting techniques can also extend to other fields. The data amount we are dealing with is simply unprecedented. In order to handle this situation, some methods are utilized to improve the performance of data retrieval and management, including CDN (Content Delivery Network), cache, etc. The current strategy usually prefetches or duplicates data that stood out in the history, say, prioritizing the items with most clicks either in history or in recent times [31]. However, the intention of these cache-like methods is to reduce the future traffic overload, not the past. If we can predict the future hot contents and prioritize them, the miss rate reduces and the delay time decreases, improving the system performance and the user experience.

The advance of predicting technique may help develop a better search engine. Although the state-of-the-art search engines provide a relatively satisfactory result, like Google or Bing search, we still find room to improve. Some people have found that those engines fail to retrieve the latest web contents. They argue that more percentage of new elements shall be brought into the users' horizon. By integrating the predicting algorithm into the previous ones, we then prioritize future potential popular items over the obsolete ones and thus provide the users with the results that will arouse their interests. This prediction can be made according to the overall data or, more precisely but costly, according to each person's historical data.

Considering all the merits that the popularity prediction could bring us, both the academics and the industry shall do further study and research in order to perfect the relevant ideas and techniques. We can either polish the algorithm to improve the accuracy of the prediction, or bring the method to a broader stage where the prediction can provide benefits.

Bibliography

[1] Deepak Agarwal, Bee-Chung Chen, Pradheep Elango, Nitin Motgi, Seung-Taek Park, Raghu Ramakrishnan, Scott Roy, and Joe Zachariah. Online models for content optimization. In *Annual Conference on Neural Information Processing Systems (NIPS)*, pages 17–24, 2008.

[2] Deepak Agarwal, Bee-Chung Chen, and Xuanhui Wang. Multi-faceted ranking of news articles using post-read actions. In *ACM International Conference on Information and Knowledge Management (CIKM)*, pages 694–703, 2012.

[3] Anne Archambault and Jonathan Grudin. A longitudinal study of facebook, linkedin, & twitter use. In *ACM Conference on Human Factors in Computing Systems (CHI)*, pages 2741–2750, 2012.

[4] Sitaram Asur and Bernardo A Huberman. Predicting the future with social media. In *IEEE/WIC/ACM International Conference on Web Intelligence (WI)*, pages 492–499, 2010.

[5] Roja Bandari, Sitaram Asur, and Bernardo A Huberman. The pulse of news in social media: Forecasting popularity. In *International Conference on Weblogs and Social Media (ICWSM)*, pages 26–33, 2012.

[6] Hila Becker, Mor Naaman, and Luis Gravano. Learning similarity metrics for event identification in social media. In *ACM International Conference on Web Search and Data Mining (WSDM)*, pages 291–300, 2010.

[7] Rushi Bhatt, Vineet Chaoji, and Rajesh Parekh. Predicting product adoption in large-scale social networks. In *ACM International Conference on Information and Knowledge Management (CIKM)*, pages 1039–1048, 2010.

[8] David M Blei, Andrew Y Ng, and Michael I Jordan. Latent dirichlet allocation. *Journal of Machine Learning Research*, 3:993–1022, 2003.

[9] Carlos Castillo, Mohammed El-Haddad, Jürgen Pfeffer, and Matt Stempeck. Characterizing the life cycle of online news stories using social media reactions. In *ACM Conference on Computer Supported Cooperative Work and Social Computing (CSCW)*, pages 211–223, 2014.

[10] Zhi-Hong Deng, Xiuwen Gong, Frank Jiang, and Ivor W Tsang. Effectively predicting whether and when a topic will become prevalent in a social network. *AAAI Conference on Artificial Intelligence*, pages 210–216, 2015.

[11] Ted Dunning. Accurate methods for the statistics of surprise and coincidence. *Computational Linguistics*, 19(1):61–74, 1993.

[12] Jennifer Golbeck, Cristina Robles, and Karen Turner. Predicting personality with social media. In *CHI Extended Abstracts on Human Factors in Computing Systems*, pages 253–262. ACM, 2011.

[13] Mengdie Hu, Shixia Liu, Furu Wei, Yingcai Wu, John Stasko, and Kwan-Liu Ma. Breaking news on twitter. In *ACM Conference on Human Factors in Computing Systems (CHI)*, pages 2751–2754, 2012.

[14] Jian Huang, Ziming Zhuang, Jia Li, and C Lee Giles. Collaboration over time: Characterizing and modeling network evolution. In *ACM International Conference on Web Search and Data Mining (WSDM)*, pages 107–116, 2008.

[15] Shu Huang, Min Chen, Bo Luo, and Dongwon Lee. Predicting aggregate social activities using continuous-time stochastic process. In *ACM International Conference on Information and Knowledge Management (CIKM)*, pages 982–991, 2012.

[16] Bernardo A Huberman, Peter LT Pirolli, James E Pitkow, and Rajan M Lukose. Strong regularities in world wide web surfing. *AAAS Science*, 280(5360):95–97, 1998.

[17] Wouter IJntema, Frank Goossen, Flavius Frasincar, and Frederik Hogenboom. Ontology-based news recommendation. In *International Conference on Extending Database Technology (EDBT)/International Conference on Database Theory (ICDT) Workshops*, page 16, 2010.

[18] Long Jin, Yang Chen, Tianyi Wang, Pan Hui, and Athanasios V Vasilakos. Understanding user behavior in online social networks: A survey. *IEEE Communications Magazine*, 51(9):144–150, 2013.

[19] Jeon-Hyung Kang and Kristina Lerman. La-ctr: A limited attention collaborative topic regression for social media. *arXiv preprint arXiv:1311.1247*, 2013.

[20] Liang Kong, Shan Jiang, Rui Yan, Shize Xu, and Yan Zhang. Ranking news events by influence decay and information fusion for media and users. In *ACM International Conference on Information and Knowledge Management (CIKM)*, pages 1849–1853, 2012.

[21] Shoubin Kong, Qiaozhu Mei, Ling Feng, Fei Ye, and Zhe Zhao. Predicting bursts and popularity of hashtags in real-time. In *ACM International Conference on Research on Development in Information Retrieval (SIGIR)*, pages 927–930, 2014.

[22] Haewoon Kwak, Changhyun Lee, Hosung Park, and Sue Moon. What is twitter, a social network or a news media? In *ACM International Conference on World Wide Web (WWW)*, pages 591–600, 2010.

[23] Jong Gun Lee, Sue Moon, and Kave Salamatian. An approach to model and predict the popularity of online contents with explanatory factors. In *IEEE/WIC/ACM International Conference on Web Intelligence and Intelligent Agent Technology (WI-IAT), 2010*, volume 1, pages 623–630, 2010.

[24] Kristina Lerman and Aram Galstyan. Analysis of social voting patterns on digg. In *ACM Workshop on Online Social Networks (WOSN)*, pages 7–12, 2008.

[25] Kristina Lerman and Rumi Ghosh. Information contagion: An empirical study of the spread of news on digg and twitter social networks. In *AAAI International Conference on Weblogs and Social Media (ICWSM)*, volume 10, pages 90–97, 2010.

[26] Kristina Lerman and Tad Hogg. Using a model of social dynamics to predict popularity of news. In *ACM International Conference on World Wide Web (WWW)*, pages 621–630, 2010.

[27] Ana Carolina ES Lima and Leandro Nunes De Castro. A multi-label, semi-supervised classification approach applied to personality prediction in social media. *Neural Networks*, 58:122–130, 2014.

[28] Zongyang Ma, Aixin Sun, and Gao Cong. On predicting the popularity of newly emerging hashtags in twitter. *Journal of the American Society for Information Science and Technology (JASIST)*, 64(7):1399–1410, 2013.

[29] Franffois Mairesse, Marilyn A Walker, Matthias R Mehl, and Roger K Moore. Using linguistic cues for the automatic recognition of personality in conversation and text. *Journal of Artificial Intelligence Research (JAIR)*, 30:457–500, 2007.

[30] Tim Paek, Michael Gamon, Scott Counts, David Maxwell Chickering, and Aman Dhesi. Predicting the importance of newsfeed posts and social network friends. In *AAAI Conference on Artificial Intelligence (AAAI)*, volume 10, pages 1419–1424, 2010.

[31] Stefan Podlipnig and Laszlo Bszrmenyi. A survey of web cache replacement strategies. *ACM Computing Surveys (CSUR)*, 35(4):374–398, 2003.

[32] Ana-Maria Popescu and Marco Pennacchiotti. Detecting controversial events from twitter. In *ACM International Conference on Information and Knowledge Management (CIKM)*, pages 1873–1876, 2010.

[33] Arunee Ratikan and Mikifumi Shikida. Feature selection based on audience's behavior for information filtering in online social networks. In *IEEE International Conference on Knowledge, Information and Creativity Support Systems (KICSS)*, pages 81–88, 2012.

[34] Xin Shuai, Xiaozhong Liu, and Johan Bollen. Improving news ranking by community tweets. In *International Conference Companion on World Wide Web Companion Volume (WWW)*, pages 1227–1232. ACM, 2012.

[35] Gabor Szabo and Bernardo A Huberman. Predicting the popularity of online content. *Communications of the ACM (CACM)*, 53(8):80–88, 2010.

[36] Jiliang Tang, Shiyu Chang, Charu Aggarwal, and Huan Liu. Negative link prediction in social media. In *ACM International Conference on Web Search and Data Mining (WSDM)*, pages 87–96, 2015.

[37] Alexandru Tatar, Panayotis Antoniadis, Marcelo Dias De Amorim, and Serge Fdida. From popularity prediction to ranking online news. *Social Network Analysis and Mining*, 4(1):1–12, 2014.

[38] Andranik Tumasjan, Timm Oliver Sprenger, Philipp G Sandner, and Isabell M Welpe. Predicting elections with twitter: What 140 characters reveal about political sentiment. In *International Conference on Weblogs and Social Media (ICWSM)*, volume 10, pages 178–185, 2010.

[39] Bimal Viswanath, Alan Mislove, Meeyoung Cha, and Krishna P Gummadi. On the evolution of user interaction in facebook. In *ACM Workshop on Online Social Networks (WOSN)*, pages 37–42, 2009.

[40] Chunyan Wang, Mao Ye, and Bernardo A Huberman. From user comments to on-line conversations. In *ACM SIGKDD International Conference on Knowledge Discovery and Data Mining (KDD)*, pages 244–252, 2012.

[41] Stanley Wasserman. *Social Network Analysis: Methods and Applications.*, volume 8. Cambridge University Press, 1994.

[42] Shanchan Wu and Louiqa Raschid. Prediction in a microblog hybrid network using bonacich potential. In *ACM International Conference on Web Search and Data Mining (WSDM)*, pages 383–392, 2014.

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

[44] Xiaoming Zhang, Xiaoming Chen, Yan Chen, Senzhang Wang, Zhoujun Li, and Jiali Xia. Event detection and popularity prediction in microblogging. *Neurocomputing*, 149:1469–1480, 2015.

[45] Xiaoming Zhang, Zhoujun Li, Wenhan Chao, and Jiali Xia. Popularity prediction of burst event in microblogging. In *International Conference on Web Age Information Management (WAIM)*, pages 484–487. Springer, 2014.