# Postprint: Research on Opinion Leader-Based Weibo Life Cycle Prediction Model

**Authors:** Huang Xianying, Yang Linfeng, Liu Xiaoyang, Gao Mindong

**Date:** 2018-05-18T00:00:00+00:00

## Abstract

To effectively investigate the role of opinion leaders in Sina Weibo propagation within social networks, as well as the lifecycle and propagation patterns of Weibo posts, an OLB Weibo propagation prediction model is proposed. First, Weibo data is crawled and analyzed; second, mathematical expressions for four factors related to influence are fitted, and a weight calculation method is provided through the Analytic Hierarchy Process; finally, the OLB model is constructed by utilizing the relationship between the calculated influence, repost counts, and related factors, thereby enabling experimental prediction and analysis of opinion leader propagation effects and Weibo lifecycle. Simulation results demonstrate that opinion leader influence is directly proportional to the propagation effect of their Weibo posts in information dissemination. Through error analysis, the average error values for four datasets are obtained as 1.0%, 5.0%, 2.4%, and 5.1%, respectively. The proposed OLB model is reasonable and effective for predicting Weibo propagation patterns.

## Full Text

## Preamble

### Research on a Microblog Life Cycle Prediction Model Based on Opinion Leaders

*Huang Xianying, Yang Linfeng?, Liu Xiaoyang, Gao Mindong*
(College of Computer Science & Engineering, Chongqing University of Technology, Chongqing 400054, China)

**Abstract:** To effectively investigate the role of opinion leaders in Sina Weibo dissemination and the life cycle and propagation patterns of microblogs in social networks, this paper proposes an OLB (Opinion Leader-Based) microblog propagation prediction model. First, we crawled microblog data and conducted data analysis. Second, we derived mathematical expressions for four factors related to influence and calculated their weights using the Analytic Hierarchy Process (AHP). Finally, we constructed the OLB model by leveraging the computed influence and the relationship between forwarding counts and relevant factors to experimentally predict and analyze the communication role of opinion leaders and microblog life cycles. Simulation results demonstrate that opinion leader influence is directly proportional to microblog propagation effectiveness in Weibo information dissemination. Through error analysis, the average error values for four datasets were 1.0%, 5.0%, 2.4%, and 5.1%, respectively, indicating that the proposed OLB model is reasonable and effective for predicting microblog propagation patterns.

**Keywords:** propagation model; opinion leader; influence; microblog life cycle

## 0 Introduction

The rise of microblogging has transformed people' s lives. As an emerging social network medium characterized by concise text, refined content, and rapid information dissemination, Weibo has gradually become the primary channel through which people obtain news. Microblogs play a crucial role in social networks, and opinion leaders, as intermediaries in information propagation, create a two-step flow of communication that significantly shapes public opinion. This has attracted considerable research attention, making the modeling and analysis of opinion leader influence and microblog life cycles one of the most cutting-edge issues in online social network research.

Recent studies by domestic and international scholars have focused on online social networks and complex networks. Regarding opinion leader identification, Wu et al. [1] proposed a novel PageRank-based method for identifying opinion leaders based on user influence, finding that opinion leaders change over time and only a small number of users remain opinion leaders across different periods. Zhang and Liang [2] categorized social network roles into predefined and undefined roles, discovering that social network role identification requires a "combined approach" to identify social roles. Li et al. [3] proposed a framework to identify signature opinion leaders and maximize information dissemination through content analysis of microblog factors. Wu et al. [4] developed a microblog opinion leader mining algorithm based on user behavior networks, utilizing both microblog content and social attributes. These studies have gradually matured the identification of opinion leaders, yet research on their propagation patterns remains limited.

For data acquisition, Deng et al. [5] constructed a microblog crawling analysis tool based on Sina Weibo's API, employing a novel MapReduce approach for microblogging to detect core opinion leaders with improved accuracy and reduced runtime. Li et al. [6] proposed three metrics—activity, propagation power, and coverage—to evaluate the influence of Weibo opinion leaders. Zhao et al. [7] constructed an influence calculation method using direct and indirect followers to predict microblog forwarding volume. Zhang et al. [8] employed Analytic Hierarchy Process to propose a user initial influence calculation model and, drawing on PageRank algorithm principles, proposed a method for calculating the diffusion of user initial influence. Feng et al. [9] proposed a short text subtopic classification algorithm based on LDA semantic information and HowNet knowledge base, comprehensively measuring microblog influence from explicit, implicit, and user perspectives. Ding et al. [10] proposed a MultiRank random walk model based on multi-relational networks, dividing influential individuals according to their influence attributes, with results showing that multi-topic hierarchical influential individuals constitute only a small portion of all influential individuals.

Regarding microblog content analysis, Dong and Liu [11] utilized community modularity, average shortest path, and network diameter to measure network tightness, classifying microblogs based on support vector machines to identify hype microblogs. Integrating machine learning algorithms with microblog classification enables effective categorization [9][11]. These studies serve as valuable preprocessing tools for analyzing the life cycle of emergency microblogs.

Research on microblog propagation patterns and prediction is equally critical. Wei et al. [12] proposed three information release patterns and constructed three different diffusion models. Tang et al. [13] proposed a dynamic interest-based friend network and built a diffusion model. Zhang et al. [14] developed a novel method to detect emergency events and predict their future development. Jalali et al. [15] proposed a DY dynamic model based on empirical data, utilizing maximum likelihood estimation for model testing and calibration. Long et al. [16] proposed a new overlapping community identification algorithm for directed networks. However, these propagation pattern studies have not employed mathematical modeling to consider microblog life cycles, nor have they examined the role of opinion leaders in the propagation process.

Studies by domestic and international scholars demonstrate that opinion leaders play a significant role in online social network propagation. With Weibo’ s rapid development and the massive volume of information spread across various social network platforms, analyzing social network behavior holds great significance. Currently, Sina Weibo has reached an average of 300 million monthly active users and over 100 million daily active users. This paper will analyze various factors in Sina Weibo, calculate relevant factor weights, analyze opinion leader influence, and construct a mathematical model to analyze microblog life cycles.

## 1 Data Collection

As China’ s largest online media platform and one of the world’ s most widely used microblog service providers, Sina Weibo generates massive data volumes, making it ideal for analyzing online social network behavior. To analyze the life cycle of original or forwarded microblogs by opinion leaders, we must first construct metrics for opinion leader propagation power and influence factors.

The data used in this study was collected through web crawling. Since Weibo protects user information, accessing too many pages within a short timeframe may result in account suspension. Therefore, we employed three accounts to crawl relevant information separately. We first crawled 200 users from Weibo' s Hall of Fame, collecting each user' s total microblog count, original microblog count, following list size, influence magnitude, and follower count. We also crawled 50 microblogs per user to calculate average comment and forwarding volumes. Activity level was derived from total microblog count, average daily microblogs, and active days. After initial crawling, we filtered out users with extremely low comment and forwarding counts and removed duplicates, resulting in 172 users.

Second, we crawled trending topic data from March 20, 2017, to June 30, 2017, selecting two time points daily. At each time point, we collected eight trending topics, yielding 508 unique topics over two months after removing duplicates. Since the number of microblogs per topic varied significantly and some ordinary users posted irrelevant microblogs to gain attention, we only analyzed 1-3 microblogs with the highest comment and forwarding counts per topic, ultimately obtaining 822 microblogs involving 1,183 opinion leaders. The following sections analyze the collected factor data and conduct modeling research on trending microblog data.

## 2 Opinion Leader Influence Analysis

After data collection, we categorized microblog factor data and investigated opinion leader influence. First, we analyzed microblog propagation patterns. Second, we examined various metrics related to opinion leader influence. Finally, we constructed a weight calculation method for opinion leader influence. The metric classification is shown in Table 1 .

### 2.1 Analysis of Opinion Leader Related Factors

We normalized the influence of crawled data using Equation (1). We plotted relationships between opinion leader influence and: total forwarded microblogs, following list count, follower count, average comment count, average forwarding count, and activity level to identify correlations.

Figure 2 [Figure 2: see original paper] shows that the number of microblogs forwarded by opinion leaders and their following list counts exhibit no functional relationship with influence. For the remaining four factors, we normalized their values to the interval [0,1] using Equation (1), then performed function fitting to analyze their relationship with influence.

Figure 3 [Figure 3: see original paper] reveals functional relationships between all four factors and influence. Through fitting four linear lines, we derived Equations (2)-(5). Table 2 presents the fitting curve metrics. The analysis shows that all four factors have linear relationships with influence, with forwarding volume showing the strongest relationship and follower count the weakest.

### 2.2 Construction and Analysis of Opinion Leader Influence

During microblog propagation through opinion leaders, influence most significantly impacts communication effectiveness. Since influence comprises multiple factors, it plays a critical role in information dissemination—greater opinion leader influence yields broader network propagation and greater social impact. We calculate opinion leader influence using follower count, comment count, forwarding count, and activity level combined with Equation (6). Activity level is obtained by normalizing and averaging three factors: total microblog count, average daily microblogs, and active days.

To determine factor weights in influence calculation, we normalized attribute matrix data for comparison within [0,1], but attribute importance remained uncertain. Therefore, we employed Saaty' s 9-point importance scale and pairwise comparison judgment matrix expert method to calculate relative importance. For instance, if evaluators consider follower count equally to moderately more important than forwarding count, we assign values of 4-5; if significantly more important, we assign 7; values of 6 indicate importance between "moderately more" and "significantly more." After comparing all metrics, we applied AHP to obtain the attribute relative importance judgment matrix. Finally, we determined factor weights through consistency testing, as shown in Table 3 .

Using weights from Table 3, we calculated opinion leader influence to better analyze microblog life cycles. Next, we examined the functional relationship between computed opinion leader influence and corresponding forwarding counts (microblogs with forwarding counts exceeding 100,000 are considered to involve 水军 (paid posters) and are capped at 100,000 for analysis).

## 3 OLB Model Establishment

### 3.1 Microblog Life Cycle Analysis

Randomly selecting 10% (82 microblogs) from the 822 collected microblogs reveals two primary propagation patterns for trending microblogs, as shown in Figure 4 [Figure 4: see original paper]. The life cycle differs from Figure 1 [Figure 1: see original paper] mainly because Sina Weibo' s high real-time nature enables immediate propagation by ordinary users and opinion leaders upon publication, resulting in rarely observed latency periods.

A microblog life cycle consists of latency, surge, secondary growth, and decline periods. The black shaded area in Figure 4 plots hourly forwarding volumes, while the upper-right curve shows cumulative hourly forwarding counts, describing two typical propagation patterns. Figure 4(a) shows a process from surge to decline after forwarding, clearly displaying surge and decline phases. Figure 4(b) includes an additional secondary growth phase, occurring when a microblog with stable forwarding counts is forwarded by a high-influence opinion leader, triggering secondary growth.

### 3.2 OLB Model Construction

Opinion leader influence gradually decays during information propagation. This study constructs influence using four factors from Sections 2.1 and 2.2, assuming follower count, forwarding count, and activity level remain constant during propagation.

When $t < t_c$, the microblog is in the surge phase, where forwarding count increases rapidly over time, modeled by Equation (9) with $t_c$ as the time control parameter. When forwarding reaches stability, we introduce amplification factor $k$ to the total forwarding count in the model. The amplification factor scales mathematically calculated forwarding counts to actual values and relates to opinion leader influence—greater influence yields larger $k$, as shown in Equation (10).

When $t < t_c$, the microblog remains in the surge phase with rapidly growing forwarding count, described by Equation (11). When $t = t_c$, the surge phase ends and growth slows to stability. The final forwarding count can be calculated using Equation (11). For computed forwarding count $forwardN$, we perform error analysis using Equations (12)-(13).

## 4 Simulation Results and Analysis

This simulation uses trending microblog data. The proposed prediction model applies to original microblogs posted by opinion leaders. Since comment and forwarding volumes are generally random, the model is suitable for trending microblogs by opinion leaders.

### 4.1 Analysis of Trending Topic Popularity

We selected four different categories of high-forwarding topics from the crawled data to fit microblog event popularity curves, analyzing how a topic' s attention evolves from emergence. Values were normalized before simulation.

Figure 6 [Figure 6: see original paper] shows that after a topic' s release, public attention peaks initially but gradually fades over time. When media releases subsidiary events as the topic fades, public attention resurges, creating cyclical patterns until popularity eventually diminishes.

### 4.2 Comparative Analysis of Real Data and OLB Model

We analyzed four microblogs from different topics in the popularity analysis. First, we simulated opinion leader follower counts and activity levels. Second, we identified opinion leaders among forwarders and calculated their influence using one-year data. Finally, we compared real data, traditional models, and the OLB model (using different colors for different opinion leaders). We adopted the model from Reference [7], which is widely used, and accumulated forwarding counts for more intuitive comparison.

Table 4 presents parameters for a movie promotion microblog posted by Kevin Tsai, forwarded by celebrity opinion leaders including Yang Mi, Wang Jia’ er, Fan Bingbing, and Zhang Dada for promotional purposes. Figure 7 [Figure 7: see original paper] shows that opinion leaders forward microblogs at irregular time points, all exhibiting surge-then-decline patterns. Information digestion occurs rapidly—public attention peaks quickly then drops to zero, reflected in cumulative forwarding counts that stabilize after reaching a maximum. Yang Mi demonstrated the strongest information propagation effect due to her substantial entertainment industry influence. The four opinion leaders’ forwarding volume accounted for 57% of total forwarding, highlighting opinion leaders’ significant role in Weibo propagation. The average error between real data and the OLB model was 1.8%, with the OLB model outperforming traditional methods for two opinion leaders.

Table 5 and Figure 8 [Figure 8: see original paper] show the “Lotte THAAD” incident microblog, involving mostly self-media opinion leaders (as public figures typically avoid stating positions on political issues). The three self-media accounts contributed 59% of total forwarding volume, demonstrating substantial influence. The average error was 5.0%, with the OLB model performing better for two opinion leaders compared to the traditional model.

Table 6 and Figure 9 [Figure 9: see original paper] address the “Bitcoin Ransomware” incident, with most opinion leaders from the IT industry. The three opinion leaders contributed 54% of total propagation volume. The average error was 2.4%, with the OLB model outperforming the traditional model for two opinion leaders.

Table 7 and Figure 10 [Figure 10: see original paper] present the “Hangzhou Arson Case” —a shocking June incident with numerous involved opinion leaders. The microblog from the victim’ s family head attracted participation from self-media, celebrities, and news media opinion leaders. The relative error was more pronounced due to multi-level propagation, where a low-influence opinion leader’ s microblog being forwarded by a high-influence leader creates disproportionate forwarding effects. The average error was 5.1%, with the OLB model showing smaller errors for all three opinion leaders compared to the traditional model.

Figures 7-10 demonstrate that the OLB model outperforms traditional models for opinion leader-based microblog prediction, proving its rationality and effectiveness.

### 4.3 Error Analysis

After simulating four microblog datasets, the OLB model demonstrated clear effectiveness in predicting life cycles after opinion leader forwarding. The average error values were 1.0%, 5.0%, 2.4%, and 5.1%, respectively. Figure 11 [Figure 11: see original paper] compares individual errors with the Reference [7] model.

Figure 11 shows that larger opinion leader influence correlates with smaller OLB model error values. As influence decreases, OLB model errors become less stable. Compared to traditional models, the OLB model shows significant improvement for most opinion leaders. Overall, OLB simulation data exhibits small errors compared to real data, indicating the model' s practical predictive capability.

## 5 Conclusion

Existing research lacks definitive mathematical models for predicting microblog life cycles in online social networks. This study analyzes characteristics of Sina Weibo microblogs, finding that propagation through opinion leaders achieves greater effects, with forwarding count being the key factor in diffusion capacity. Function fitting of influence-related factors reveals linear relationships between opinion leader influence and four factors: follower count, average comment count, average forwarding count, and activity level. Using AHP to assign corresponding weights, we calculated opinion leader influence magnitude. Building upon analysis of relationships among microblog factors, we constructed the OLB mathematical model. Simulation and error analysis validated the OLB model' s practical effectiveness. Future research will focus on microblog content classification and defining the duration of opinion leader influence in the model.

## References

[1] Wu Y, Ma L, Lin M. Opinion leader discovery algorithm based on user influence [J]. Journal of Chinese Computer Systems, 2015, 36(3): 561-565.

[2] Zhang S, Liang X, Qi J. Survey on social network role identification methods [J]. Chinese Journal of Computers, 2017, 40(3): 649-673.

[3] Li F, Du T C. Maximizing micro-blog influence in online promotion [J]. Expert Systems with Applications, 2017, 70: 52-66.

[4] Wu X, Zhang H, Zhao X, et al. Microblog opinion leader mining algorithm based on user behavior network [J]. Application Research of Computers, 2015, 32(9): 2678-2683.

[5] Deng X, Li Y, Lin S. Parallel Micro Blog Crawler Construction for Effective Opinion Leader Approximation [J]. AASRI Procedia, 2013, 5(3): 170-176.

[6] Li Y, Hu Y, Xiong X. Evaluation model of microblog opinion leaders [J]. Information Security and Communications Privacy, 2013(2): 79-81.

[7] Zhao H, Liu G, Shi C, et al. Microblog forwarding prediction based on forwarding propagation process [J]. Acta Electronica Sinica, 2016, 44(12): 2989-2996.

[8] Zhang Y, Jiang Y, Chen R. Analysis and mining of opinion leaders in microblog user relationship networks [J]. Journal of Beijing Information Science & Technology University, 2015(4): 7-14.

[9] Feng S, Jing S, Yang Z. Chinese microblog topic opinion leader mining based on LDA model [J]. Journal of Northeastern University, 2013, 34(4): 490-493.

[10] Ding Z, Zhou B, Jia Y. Topic hierarchical influence analysis based on multi-relational networks in microblogging [J]. Journal of Computer Research and Development, 2013, 50(10): 2155-2175.

[11] Dong Y, Liu Y, Luo J. Hype microblog identification method based on support vector machine [J]. Computer Engineering, 2015, 41(3): 7-14.

[12] Wei J, Bu B, Liang L. Estimating the diffusion models of crisis information in micro blog [J]. Journal of Informetrics, 2012, 6(4): 600-610.

[13] Tang M, Mao X, Guessoum Z, et al. Rumor diffusion in an interests-based dynamic social network [J]. The Scientific World Journal, 2013, 3(8): 2-9.

[14] Zhang X, Chen X, Chen Y, et al. Event detection and popularity prediction in microblogging [J]. Neurocomputing, 2015, 149(8): 1469-1480.

[15] Jalali M S, Ashouri A, Herrera-Restrepo O, et al. Information diffusion through social networks: the case of an online petition [J]. Expert Systems with Applications, 2016, 44(9): 187-197.

[16] Long H, Li B. Overlapping Community Identification Algorithm in Directed Network [J]. Procedia Computer Science, 2017, 107(3): 527-532.