Opinion-based Analysis of Structural Patterns in Online Viral Diffusion

Jun Yang[1, 2], Nan Zhou[1], Yifan Li[1], Xueer Xue[3], Yingfei Dong[4], Yibo Xue[2, *], Jun Li[2]
[1]*Department of Automation, Tsinghua University, China.*
[2]*Research Institute of Information Technology, Tsinghua University, China.*
[3]*Department of Computer Engineering, San Jose State University, USA.*
[4]*Department of Electrical Engineering, University of Hawaii, USA.*
({jun-yang15, zhoun14, liyf14}@mails.tsinghua.edu.cn, cindyxue1405@gmail.com, yingfei@hawaii.edu, {yiboxue, junl}@tsinghua.edu.cn)

***Abstract*—As online information is often disseminated by various cascades, it is critical to understand the underlying mechanisms and influential factors that lead to the structural diversity in online diffusion. In this work, we examine the relations between the public opinion among information propagators and the structural patterns of online diffusion based on a Weibo dataset. To complement the existing structural measures of online diffusion cascades, we first propose six new structural patterns based on the analysis of multi-dimensional structural metrics. We then propose a set of new public opinion metrics by a hybrid sentiment analyzer to reveal the correlation between public opinion and diffusion structures. We further apply our understanding to predict the diffusion structure of diffusion events based on public opinion. The results of our experiments show the effectiveness of the proposed opinion-based analysis in the prediction of both structural virality and our proposed structural patterns.**

Keywords—Information diffusion; sentiment analysis; structural virality; structural patterns; online social networks

I. INTRODUCTION

Online social networks (OSNs) have enabled millions of users sharing their interests almost instantly, such as news, ideas, and personal opinions. Twitter had 330 million monthly active users globally in October 2017, while one of the most popular and representative online social medias in China, Sina Weibo, had 376 million monthly active users at the same time. As millions of users produce, share, and comment on various information through OSNs in their daily lives, the mechanisms of online information diffusion have also attracted enormous attentions from researchers due to its growing importance in various fields, such as sociology, economics, and cyber security.

Online diffusion is mostly via peer-to-peer spreading. Users can decide whether to share certain received information with their friends. Such mechanisms usually spread information in tree-like cascades, and the characterization and prediction of those cascade structures have become an important topic in social network research, which can be applied in precision marketing, viral advertising, rumor control, anti-terrorist campaigns, etc.

Early research by Kempe et al. [1], Aral et al. [2], and Budak et al. [3] introduced the structural diversity in online diffusion, usually caused by different user relations and underlying networks. Goel et al. [4] introduced a definition of structural virality and related numerical metrics that quantified the intuitive distinctions between broadcast-dominated cascades and viral diffusions, which is rigorously defined by observed diffusion cascade structures and without using any underlying network knowledge or models. Using the same structural virality metrics [4], Cheng et al. [5] presented their approach to predict the eventual structural virality of diffusion cascades based on early observations. They utilized user features, shared contents, cascade structure and temporality in their prediction experiments to show the predictability of the eventual diffusion "shapes".

Apart from the previous work (e.g., [5]), to the best of our knowledge, other potential influential factors have not yet been considered for diffusion structure prediction, in particular, *public opinion*. In certain social networks like Twitter and Weibo, users are allowed to share information with their own comments, which often express their sentiments and attitudes, and thus reflect the public opinion of certain diffusion events. Although public opinion in OSNs has been discussed in some previous works [6-9], no research has examined its *relation with diffusion structures*. We will focus on this topic.

Another challenge in diffusion structure prediction is proper structural metrics and prediction targets. Goel et al. [4] proposed a widely-accepted one-dimensional value to measure structural virality. Cheng et al. [5] used the statistical median of structural virality as the threshold in prediction experiments to achieve a balanced classification problem. In some application scenarios, more refined prediction results are expected, and the existing approaches are not sufficient to describe the detailed structural characteristics. Thus, it will be helpful to have high-dimensional observations in multiple structural measures, which complement the existing one-dimensional metrics.

In this paper, we examined the structures of online diffusion cascades and empirically explored their relations with public opinion.

II. RELATED WORK

We briefly discuss related work here, and leave more details of related methods in the following sections when we analyze related methods and compare them with our methods.

Sentiment analysis. We propose sentiment analysis on user-generated content to quantify the public opinion of diffusion events. According to Schouten et al. [10], the existing method of sentiment analysis are divided into three categories, including lexicon-based ones, supervised machine learning, and unsupervised machine learning. Ekman et al. [11]

978-1-5386-4485-0/18/$31.00 ©2018 IEEE

proposed six basic emotional states. Xu et al. [12] introduced a weighted Chinese fine-grained emotional lexicon. In our paper, referring to previous work by Zhao et al. [13], we propose a hybrid approach combining vocabulary knowledge with supervised machine learning algorithms.

Online diffusion prediction. Earlier work investigated the internal mechanisms and influential factors of individual diffusion behaviors and diffusion popularity [14-18]. Several model-based theories were proposed [19-21]. Different from these approaches, we focus on macro-scale diffusion structures rather than diffusion sizes or micro-scale diffusion events.

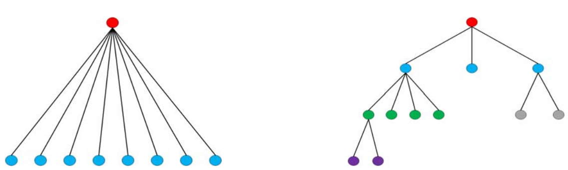
Cascade structure analysis. A few projects indirectly examined the diffusion structures by analyzing time-series information [22-25]. Statistical analysis for structural characteristics had also been conducted [26-30]. Recent work by Han et al. [31] tried to study the relations between emotion factors and the structural virality. Different with their work, we use structural patterns rather than their one-dimensional measure to quantify detailed diffusion structures. Furthermore, the structural knowledge is considered in our public opinion metrics, which is proven critical in our experiments.

III. Limitations of Existing Structural Metrics and Definitions of Proposed New Structural Patterns

The structures of diffusion cascades have been considered to reflect the underlying mechanisms of information diffusion, and may be influenced by various factors. Fig. 1 illustrates typical examples with varying diffusion structures. The top red root node represents the source publisher, and each edge represents an individual diffusion behavior for a user pair.

The diffusion in Fig. 1.(a) is dominated by a large broadcast from a single parent (the root node), while Fig. 1.(b) shows a multigenerational branching process, in which the diffusion is driven by several different influencers. These diffusion structures look obviously different even if they might be similar in diffusion popularity. Diffusions with structures like Fig. 1.(b) are usually considered to be more viral according to the previous studies, since longer diffusion paths may cause a bigger and more unpredictable spreading scope. Although the difference is intuitive, systematic metrics are required, since the complete description of graphs is usually complicated and too difficult to study.

In this section, we first analyze existing metrics and then illustrate the necessity of new structural patterns. We then propose six new structural patterns with precise definitions.



(a) Broadcast-dominated. (b) Multigenerational.

Fig. 1. Typical diffusion structures.

A. Our dataset

To study the structure and public opinion of online diffusion, we randomly collect samples of Weibo popular diffusion events from December 30, 2016 to May 24, 2017. We collect personal information of users, tweet contents and retweeting activities. In the dataset, 28,307,541 messages and 11,225,257 unique user profiles are involved.

Weibo provides the complete reposting list and explicit information of parent node for every tweet and retweet in their API[1]. With the algorithm introduced by Yi et al. [32], a total of 60,551 independent diffusion events are reconstructed as corresponding diffusion trees[2]. We select a total of 31,978 events that have at least 100 nodes and construct the diffusion dataset for our structural research.

The average tree size of the qualified events is 849, and their size distribution is also calculated. We notice that the tree size in our diffusion dataset follows a heavy-tailed distribution, similar to existing results ([4, 5]). Moreover, since large diffusions rarely occur on OSNs (only 0.025% of all tweets were with at least 100 retweets, noted by Goel et al. [4]), our dataset is a considerable sampling from all Weibo popular diffusion events during this period. Thus, we believe that our diffusion dataset can be considered as an unbiased sampling with respect to Weibo popular diffusions.

B. Existing Structural Metrics and Their Limitations

While different structural metrics have been proposed, we mainly consider the following ones in our work.

Wiener Number (WN) [33] is originally proposed in mathematical chemistry to represent the average distance between all pairs of nodes in a graph. It was introduced as a diffusion structural metric by Goel et al. [4], and has been widely accepted as a state-of-art measure for structural virality.
Modularity (MOD) is a measure in community detection proposed by Blondel et al. [34]. It is first used to measure diffusion structures by Yi et al. [32].
Average Depth (AD) is the average depth of nodes in the tree.
Maximum Sub-Tree Proportion (MST) is the fraction of nodes that are in the largest sub-tree, over the total number of nodes in the tree. It has been analyzed by Rodrigues et al. [29].
Non-Maximum Broadcast (NMB) is the fraction of nodes that are not the children of the largest broadcaster (the node with the largest number of children in the diffusion tree), over the total number of nodes in the tree.
Distinct Parent (DP) is the probability that two randomly selected nodes from the tree have different parent nodes.

All these metrics are continuous variables, and can distinguish broadcast-dominated and multigenerational diffusions by different values. Among which, WN is one of the most widely accepted measures for diffusion structural virality. However, one existing study has shown the relatively low isomorphism-discriminating power of WN [35]: some heterogeneous structures may not be distinguished perfectly by WN, especially as the tree size increases. As these metrics are usually used separately in existing methods, it is natural to consider combining some of them to describe diffusion

[1] https://m.weibo.cn/api/.

[2] A list file with all unique ids of original tweets in our dataset could be downloaded at http://pan.baidu.com/s/1geWI9Bd/.

structures more elaborately and discriminately. Thus, a series of analysis has been carried out.

Goel et al. [4] computed the Spearman rank correlation [36] between some structural metrics mentioned in the above (WN, AD, NMB and DP), and all metrics were observed to be highly correlated with each other. Thus, those metrics were considered to be equivalent and interchangeable, to a certain extent. However, we notice that in the case when the diffusion is not so viral, all those metrics are almost strictly equivalent since all of them are very close to their lower limits. In addition, according to both the previous work ([4]) and our observation, most of the diffusion events have little virality. The statistical median value of WN is 2.03 in our dataset, e.g., which is just slightly higher than the out-and-out non-viral cases. Thus, we think that the large non-viral portion may be the main cause of observed high correlation between structural metrics. We like to examine whether such correlation maintains among the highly-viral events.

The structural virality is measured by the widely accepted metric WN, and the Spearman rank correlation matrix is generated among viral events. As a specific example, we only consider the metric correlation among the diffusion events that have the top 10% highest WN values. Table I shows the results computed over the entire dataset and the results among the top 10% most viral events. We can see the high correlation over the entire set, which agree with the conclusions in [4]. However, the correlation significantly declines among the highly viral events, which confirms our suspicion.

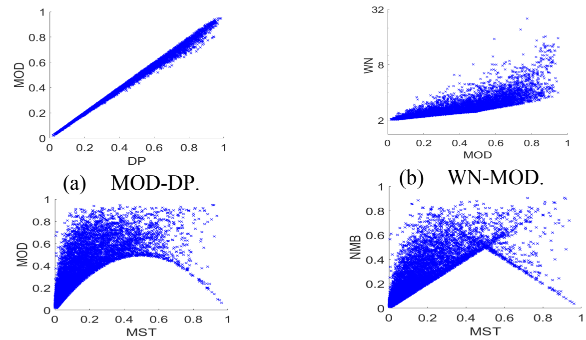
Goel et al. [4] also mentioned that one of the benefits to use a continuous structural metric (rather than category-based one) is that the structural virality between any two events can be compared. We like to examine the isotonicity among different metrics, especially for the highly viral events. Specifically, we impose different thresholds as the "viral proportion" (e.g., top 50% or 10% events with the highest values of chosen metrics). We then calculate the overlap ratio between the viral event sets, defined by different metrics. We notice that the overlap ratio significantly declines when the threshold varies from 50% to 10%. For example, about 37% of the viral events that classified by WN are considered non-viral by MST, when the threshold is set to top 10%.

From the above analysis, we see the independence and divergence between various structural metrics, especially among the highly viral events, which may indicate the necessity to combine multiple metrics for better measurements. Fig. 2 shows several typical value distributions on a two-dimensional plane with different metric combination. Some metrics show high consistency such as MOD and DP in Fig. 2.(a), while other distributions seems relatively chaotic.

Note that the interesting distribution in Fig. 2.(d), where many points concentrate around two diagonals while the bottom quarter has no points. Since the metrics in Fig. 2.(d) (MST and NMB) have clear physical meanings, the corresponding diffusion structure can be roughly inferred by the point position. Thus, we define several new structural patterns and illustrate them by the division of the two-dimensional plane, as introduced in the following subsection.

TABLE I. RANK CORRELATION BETWEEN STRUCTURAL MECTRICS

	WN	**MOD**	**MST**	**NMB**
Entire Set				
WN	1.0000	0.9828	0.8852	0.9820
MOD	0.9828	1.0000	0.9268	0.9999
MST	0.8852	0.9268	1.0000	0.9279
NMB	0.9820	0.9999	0.9279	1.0000
10% Viral				
WN	1.0000	0.6095	0.2962	0.5346
MOD	0.6095	1.0000	0.5361	0.9678
MST	0.2962	0.5361	1.0000	0.6437
NMB	0.5346	0.9678	0.6437	1.0000



(a) MOD-DP. (b) WN-MOD. (c) MOD-MST. (d) NMB-MST.

Fig. 2. Structural metrics value distribution.

C. *Proposed Structural Patterns*

By examining the two-dimensional distributions between multiple metrics, we propose six structural patterns by the division of the plane. It is necessary to emphasize that our notion of structural patterns is to complement rather than substitute for the existing structural virality metrics [4].

Each pattern is named according to its characteristics. We empirically divide the plane into different areas based on the observation of real-world sample distribution and our understanding of the structural metrics. Specific definition and description is as follows. Our partition fully covers the entire feature space and does not depend on different dataset.

Flare. Structures with metric values close to the origin and left-bottom diagonal in Fig. 2.(d) (NMB<0.5, NMB≈MST, MST<1/11), which represent that the broadcast from the root node dominates all diffusions, and the influence of any other nodes is a magnitude weaker (less than 1/10 of the root node's). Such a diffusion structure contains just one main exposure by the original publisher and is similar to a flare.

Echo. Structures around the left-bottom diagonal but further from the origin compared to *Flare* (NMB<0.5, NMB≈MST, MST≥1/11). The root broadcast dominates the diffusion with a majority, while another secondary node leads almost all other diffusions. The root node gets support from a weaker but important opinion leader as a single echo.

Detonation. Structural metric values distribute around the right-bottom diagonal of Fig. 2.(d) (NMB<0.5, NMB+MST≈1), which means that a secondary node leads the diffusion with a majority, while the root node broadcasts to most of the remaining nodes. The first broadcast causes a larger explosion, like a detonation.

Colony. Structures that at the area of the top and the right quarter in Fig. 2.(d) (NMB+MST≥1), which means some secondary nodes are stronger influencers than the root node. Unlike *Detonation*, the diffusion process is driven by multiple secondary influencers, whose diffusion topology looks like a bacteria colony.

Firework. Structures that at the top half of the left quarter (NMB≥0.5, NMB+MST<1). Metrics in this area show that the root broadcast dominates the largest influence but is not yet an absolute majority. More than half of the retweets happen in secondary nodes.

Galaxy. Structural metric values in the bottom half of the left quarter in Fig. 2.(d) (NMB<0.5, NMB+MST<1). Different from *Firework*, the majority (more than half, to be accurate) of the diffusion is directly by the root broadcast. On the other hand, multigenerational diffusion exists dispersedly compared to *Flare* or *Echo*.

The relation between our structural patterns and the structural virality (measured by WN) is also analyzed, as shown in Table II. We notice that if we use a balanced viral threshold, which means the events with the top 50% highest WN values are considered as viral diffusions, the *Flare*, *Echo* and *Galaxy* patterns account for the majority. However, when the threshold is set to the top 10%, hardly any *Flare* events are included. On the other hand, almost all *Colony* and *Firework* events are with the top 10% highest structural virality.

These new definition of structural patterns complement the existing structural metrics, and help us describe the detailed diffusion structural characteristics, which will be used in our following study on opinion effects.

TABLE II. STRUCTURAL PATTERNS IN VARYING VIRAL THRESHOLD

Viral Ratio	**Fla.**	**Ech.**	**Det.**	**Clo.**	**Fir.**	**Gal.**
50%	8235	1996	355	469	563	4371
20%	259	1486	336	469	563	3283
10%	16	442	155	460	552	1573

IV. NEW PUBLIC OPINION FEATURES AND STRUCTURAL PREDICTION

To study the relations between public opinion and the proposed diffusion structures, we first need to label *individual opinions* expressed by users in diffusion events. Referring to Ekman's basic emotion theory [11], one of the most representative emotion theory, we definite six emotion states. The six basic emotion states include *happiness*, *surprise*, *fear*, *sadness*, *anger,* and *disgust*. The *individual opinions* of users are annotated according to the source contents and comments published by users when they retweet.

In this section, we first introduce the method to categorize *individual opinions*, and then propose a set of public opinion features for diffusion structure prediction. Hybrid Sentiment Analysis and New Opinion Features

1) A Hybrid Sentiment Analysis Method

Refer to existing work as [13, 37], we use a hybrid method that combines vocabulary knowledge and supervised machine learning algorithms. This method has been proven effective for handling short text without complex grammars. The lexical knowledge includes an open-source Chinese emotion lexicon [12] and manually annotated Weibo emojis. By which, we first build an emotion-labeled dataset (about 4.2 million tweets) from text contents in our diffusion dataset introduced in Sec. III.A. The labeled dataset is then used to train our Naïve Bayes algorithms. Our classifier analyzes the contents published by a user, and determine the *individual opinion* of this user for a specific diffusion event. Our classifier achieves Precision, Recall and F1-score over 80% empirically, which shows similar performance compared to the existing methods. Furthermore, our method also handles users that published no comment on a tweet and have not be categorized by our classifier yet. Those users will be considered to imply the same opinions as their parent nodes, referring to the emotion contagion theory by Guillory et al. [38].

Using our method, the *individual opinion* of every user in a diffusion event can be successfully distinguished. By labeling *individual opinion* of every user, we are now able to describe the public opinion characteristics of a diffusion event in the following.

2) Public Opinion Features

Now we focus on the public opinion features for diffusion events. With the above classifier, we label *individual opinion* of every user in a diffusion. To conduct our analysis in a macro scale, we need recapitulative metrics to describe the overall public opinion based on particular information on diffusion. According to Yang et al. [40], we use six public opinion features including ***source opinion*** (***SO***), ***opinion composition***, ***weighted opinion composition***, ***opinion distribution entropy***, ***objection proportion***, and ***within-group opinion distance*** (***WOD***). Based on these new public opinion features, we examine the influence of public opinion in diffusion structures and perform further analysis and evaluation.

B. Relations of Opinion and Diffusion Structures

We now use statistical tests to reveal inherent relations between public opinion and structure exists, based on the proposed public opinion features. We conducted a series of tests and the results indicate the significant impacts of public opinion in structural patterns. The composition of events grouped by source opinons is presented in Table III.

We notice that the pattern composition among various source opinions seems quite different. For instance, *anger* and *happiness* tweets show smaller probabilities to have *Flare* structures; the probability that a *surprise* tweet has an *Echo* diffusion is two times greater than that of an *anger* tweet; the *Colony* proportions between *happiness* and *sadness* tweets show obvious differences; an *anger* tweet is much more likely to have a *Firework* or *Galaxy* structure compared to a *surprise* tweet. We performed the Pearson's chi-squared test and the composition differences of all dominant opinions are proved statistically significant. Thus, we confirm the existence of relations between diffusion source opinions and their structural patterns.

TABLE III. PATTERN COMPOSITION OF DIFFERENT OPINION

	Fla.	Ech.	Det.	Col.	Fir.	Gal.
Anger	0.5042	0.1041	0.0183	0.0261	0.0424	0.3049
Disgust	0.5635	0.1469	0.0231	0.0304	0.0294	0.2067
Fear	0.5432	0.1494	0.0238	0.0187	0.0247	0.2403
Happiness	0.5044	0.1269	0.0268	0.0380	0.0346	0.2692
Sadness	0.5539	0.1778	0.0233	0.0146	0.0146	0.2157
Surprise	0.5541	0.1991	0.0152	0.0303	0.0108	0.1905

C. Diffusion Structure Prediction

To examine effectiveness of public opinion features in the structure prediction, we first propose our prediction algorithm. We use feature sets including both our new opinion-based features and baseline features from the previous methods. We present the prediction performance and the impacts of public opinion from different perspectives.

1) Prediction Algorithm

Base on above analysis, we propose an algorithm for diffusion structure prediction. Our algorithm combines public opinion features with other known influential factors in structure prediction.

Similar with the previous work in [5], we define the prediction tasks as classification problems. By an early observation on a small initial portion of a diffusion event, we try to predict its eventual structure. In our prediction tasks, we try to predict their eventual structural patterns according to early observed information.

The window for early observation is set as the first 15 minutes after the original tweet is published. By analyzing our dataset, only about 4.8% retweets happened in such a window. All information in this period is available for us to perform predictions. It would be a great help if such early detection is valid.

We first like to show that how the public opinion features help us in diffusion structure prediction. Moreover, we are also interested in understanding how various aspects of prediction tasks influence the prediction performance, especially the impacts of public opinion features.

We use the logistic regression classifier in all our prediction tasks, and primarily evaluate the performance by the *F1-score* and the *area under the ROC curve (AUC)*. We performed 10-fold cross validation, and all performance improvement is confirmed statistically significant by ANOVA.

2) Prediction Feature Selection

We here introduce the factors considered in our structure prediction. We first use a series of features proposed by Cheng et al. [5] as the baseline features for comparison. Because some information they used is unavailable in Weibo, we have to do some adjustments. Furthermore, we propose our public opinion features that have been introduced in Sec. IV.A.(2).

Contents features include *the number of mentions, hashtags, emojis, and text lengths* for the original tweet contents.
Original poster features include *the numbers of followers, followees, and past tweets; gender; or whether the user is verified.*
Resharer features include *proportion of male, verified, repeatedly participating, and retweet with contents among all nodes; 90th percentile follower numbers of all nodes; potential viewer (the summary of follower number of all nodes).*
Structural features include *the size of the observed cascade*, as well as all metrics mentioned in the above (*WN, MOD, AD, MST, NMB, DP*).
Temporal features *include the publishing time (in the time of a day) of the original tweet; average reposting time interval since the original tweet and a previous tweet.*

The proposed **public opinion features** are our focus, including ***SO***, ***opinion composition***, ***weighted opinion composition***, ***opinion distribution entropy***, ***objection proportion,*** and ***WOD***.

3) Prediction Results

We focus on predicting the eventual structural patterns of a viral diffusion. The prediction performance evaluated by AUC for each pattern is reported in Table IV. We first notice that prediction results for different patterns are all significantly improved by the introduction of opinion-based features. Besides, the predictability among different patterns seems to be quite different. For instance, it may be easier to detect a *Firework* diffusion by the early observation, compared to an *Echo* one. Individual analysis is also carried out, and the result indicates that the ***WOD***, ***opinion distribution entropy,*** and ***objection proportion*** are usually effective predictors for structural pattern prediction.

In summary, we find that adding our public opinion features are always helpful by improving the prediction performance compared with the baseline features.

V. CONCLUSION

In this paper, we study the influence of public opinion in online diffusion structures. We rigorously define new structural patterns, and show the strong influence of the proposed public opinion features in both structural virality and structural patterns with statistical analysis and prediction experiments.

We first presented statistical analysis on existing structural metrics, and illustrate the necessity of fine-grained structural measures. We proposed a set of new structural patterns, which complements the depiction for online diffusion structures. We then introduced opinion-related features for the macro-scale public opinion description of diffusion events. We performed statistical analysis to reveal the significant correlation between public opinion and diffusion structures. Finally, we applied our proposed opinion-based model to improve the performance of diffusion structure prediction.

Although we have shown the influence of public opinion features in diffusion events, there are still several open problems. The underlying mechanism that how different opinions drive online diffusion has not yet been fully explained. Contagion models might help, and model-based simulation can be carried out to see whether the observed structural diversity is replicated.

TABLE IV. PERFORMANCE OF PREDICTING STRUCTURAL PATTERNS

	Fla.	Ech.	Det.	Col.	Fir.	Gal.
Opinion	0.629	**0.595**	0.554	**0.613**	0.624	**0.644**
Baseline	**0.641**	0.560	**0.685**	0.585	**0.664**	0.605
All	**0.672**	**0.626**	**0.714**	**0.718**	**0.759**	**0.698**

ACKNOWLEDGMENT

This work was supported by the National Key Technology R&D Program of China under Grant No. 2015BAK34B00.

REFERENCES

[1] Kempe D, et al. Maximizing the spread of influence through a social network[C]//Proceedings of SIGKDD'03, 2003: 137-146.

[2] Aral S, et al. Distinguishing influence-based contagion from homophily-driven diffusion in dynamic networks[J]. Proceedings of the National Academy of Sciences, 2009, 106(51): 21544-21549.

[3] Budak C, et al. Structural trend analysis for online social networks[J]. Proceedings of the VLDB Endowment, 2011, 4(10): 646-656.

[4] Goel S, et al. The structural virality of online diffusion[J]. Management Science, 2015, 62(1): 180-196.

[5] Cheng J, et al. Can cascades be predicted?[C]//Proceedings of WWW'14, 2014: 925-936.

[6] O'Connor B, et al. From tweets to polls: Linking text sentiment to public opinion time series[J]. ICWSM, 2010, 11(122-129): 1-2.

[7] Anstead N, et al. Social media analysis and public opinion: The 2010 UK general election[J]. Journal of Computer - Mediated Communication, 2015, 20(2): 204-220.

[8] Naveed N, et al. Bad news travel fast: A contents-based analysis of interestingness on twitter[C]//Proceedings of Websci'11, 2011: 8.

[9] Stieglitz S, et al. Emotions and information diffusion in social media—sentiment of microblogs and sharing behavior[J]. Journal of management information systems, 2013, 29(4): 217-248.

[10] Schouten K, et al. Survey on aspect-level sentiment analysis[J]. IEEE Transactions on Knowledge and Data Engineering, 2016, 28(3): 813-830.

[11] Ekman P. An argument for basic emotions[J]. Cognition & emotion, 1992, 6(3-4): 169-200.

[12] Yu X L L H P, et al. Constructing the Affective Lexicon Ontology [J][J]. Journal of the China Society for Scientific and Technical Information, 2008, 2: 006.

[13] Zhao J, et al. Moodlens: an emoticon-based sentiment analysis system for chinese tweets[C]//Proceedings of SIGKDD'12, 2012: 1528-1531.

[14] Boyd D, et al. Tweet, Tweet, Retweet: Conversational Aspects of Retweeting on Twitter[C]// HICSS'10, 2010:1-10.

[15] Suh B, et al. Want to be retweeted? large scale analytics on factors impacting retweet in twitter network[C]// SocialCom'10, 2010: 177-184.

[16] Bakshy E, et al. Everyone's an influencer: quantifying influence on twitter[C]//Proceedings of WSDM'11, 2011: 65-74.

[17] Petrovic S, et al. RT to Win! Predicting Message Propagation in Twitter[J]. ICWSM, 2011, 11: 586-589.

[18] Yang Z, et al. Understanding retweeting behaviors in social networks[C]//Proceedings of CIKM'10, 2010: 1633-1636.

[19] Goldenberg J, et al. Talk of the network: A complex systems look at the underlying process of word-of-mouth[J]. Marketing letters, 2001, 12(3): 211-223.

[20] Granovetter M. Threshold models of collective behavior[J]. American journal of sociology, 1978, 83(6): 1420-1443.

[21] Leskovec J, et al. Patterns of cascading behavior in large blog graphs[C]//Proceedings of SIAM '07, 2007: 551-556.

[22] Cui P, et al. Cascading outbreak prediction in networks: a data-driven approach[C]//Proceedings of SIGKDD'13, 2013: 901-909.

[23] Subbian K, et al. Detecting Large Reshare Cascades in Social Networks[C]//Proceedings of WWW'17, 2017: 597-605.

[24] Yang J, et al. Patterns of temporal variation in online media[C]//Proceedings of WSDM'11, 2011: 177-186.

[25] Matsubara Y, et al. Rise and fall patterns of information diffusion: model and implications[C]//Proceedings of SIGKDD'12, 2012: 6-14.

[26] Liben-Nowell D, et al. Tracing information flow on a global scale using Internet chain-letter data[J]. Proceedings of the National Academy of Sciences, 2008, 105(12): 4633-4638.

[27] Romero D M, et al. Differences in the mechanics of information diffusion across topics: idioms, political hashtags, and complex contagion on twitter[C]//Proceedings of WWW'11, 2011: 695-704.

[28] Zang C, et al. Quantifying Structural Patterns of Information Cascades[C]//Proceedings of WWW'17, 2017: 867-868.

[29] Rodrigues T, et al. On word-of-mouth based discovery of the web[C]//Proceedings of SIGCOMM'11, 2011: 381-396.

[30] Gomez Rodriguez M, et al. Structure and dynamics of information pathways in online media[C]//Proceedings of WSDM'13, 2013: 23-32.

[31] Han J, et al. Predicting Popular and Viral Image Cascades in Pinterest[C]// Proceedings of ICWSM'17, 2017: 82-91

[32] Yi C, et al. A novel method for information propagation model perceiving[C]//ICNC'14, 2014: 6-10.

[33] Wiener H. Structural determination of paraffin boiling points[J]. Journal of the American Chemical Society, 1947, 69(1): 17-20.

[34] Blondel V D, et al. Fast unfolding of communities in large networks[J]. Journal of statistical mechanics: theory and experiment, 2008, 2008(10): P10008.

[35] Dobrynin A A, et al. Wiener index of trees: theory and applications[J]. Acta Applicandae Mathematicae, 2001, 66(3): 211-249.

[36] Spearman C. The proof and measurement of association between two things[J]. The American journal of psychology, 1904, 15(1): 72-101.

[37] Yang J, et al. Propagator or Influencer? A Data-driven Approach for Evaluating Emotional Effect in Online Information Diffusion[C]// Proceedings of ASONAM '17, 2017: 836-843.

[38] Guillory J, et al. Upset now?: emotion contagion in distributed groups[C]//Proceedings of SIGCHI'11, 2011: 745-748.

[39] Fisher R A. On the probable error of a coefficient of correlation deduced from a small sample[J]. Metron, 1921, 1: 3-32.

[40] Yang J, et al. Homophily and Controversy: On the Role of Public Opinion in Online Viral Diffusion[C]//Computer Communications Workshops (INFOCOM WKSHPS), 2018 IEEE Conference on. IEEE, 2018.