

# Homophily and Controversy: On the Role of Public Opinion in Online Viral Diffusion

Jun Yang<sup>1,2</sup>, Nan Zhou<sup>1</sup>, Yifan Li<sup>1</sup>, Xueer Xue<sup>3</sup>, Yingfei Dong<sup>4</sup>, Yibo Xue<sup>2,\*</sup>, Jun Li<sup>2</sup>

<sup>1</sup>Department of Automation, Tsinghua University, Beijing, China.

<sup>2</sup>Research Institute of Information Technology, Tsinghua University, Beijing, China.

<sup>3</sup>Department of Computer Engineering, San Jose State University, San Jose, USA.

<sup>4</sup>Department of Electrical Engineering, University of Hawaii, Honolulu, USA.

{jun-yang15, zhoun14, liyf14}@mails.tsinghua.edu.cn, cindyxue1405@gmail.com, yingfei@hawaii.edu, {yiboxue, junl}@tsinghua.edu.cn

**Abstract**—It is critical in social network analysis to understand the underlying mechanisms of online information diffusion. Although there has been much progress on the influential factors that lead to online viral diffusion, little is known about the impact by public opinion. In this paper, we examine the relations between the public opinion among information propagators and the virality of online diffusion based on a large-scale real-world dataset. We propose a set of new metrics for public opinion in online diffusion to reveal their correlation with diffusion structural virality, and further apply our understanding to predict diffusion virality based on public opinion. The experimental results show the effectiveness of the proposed analysis in the prediction of viral diffusion events.

**Keywords**—Information diffusion; structural virality; public opinion; sentiment analysis; online social networks

## I. INTRODUCTION

Online social networks (OSNs) have enabled billions of users to produce, share, and comment on various information in their daily lives. The mechanism of online information diffusion has 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 decide whether to share certain received information with others. Such mechanisms usually spread information in tree-like cascades. Based on these cascade structures, Goel et al. [1] introduced a measure of diffusion structural virality to quantify the distinctions between viral and non-viral diffusions, when diffusions with higher branching factors and numbers of generations are considered more viral. Although influential factors that lead to viral diffusions have been studied [2], to the best of our knowledge, the potential influence by public opinion has not yet been considered. In certain social networks like Twitter and Weibo, users are allowed to share information with their own comments, which often express their sentiments and thus reflect the public opinion of certain diffusion events.

In this paper, we aim to examine the relations between public opinion and diffusion virality. We randomly collect 60,551 Weibo popular diffusion events from December 30, 2016 to May 24, 2017, and the explicit diffusion structures are reconstructed. The *individual opinion* of each propagator is labeled by the hybrid sentiment analysis approach in [3], and we then quantify the public opinion by a set of new metrics combining emotional and structural knowledge.

By observation and experiment on our dataset, we empirically revealed the significant influences of public opinion on diffusion virality. We also applied our analysis to predict viral diffusion based on the proposed public opinion features for early detection.

## II. PUBLIC OPINION METRICS

### A. Individual Opinion Definition

The *individual opinion* of each propagator is defined following Ekman's basic emotion theory [4], which categorizes *individual opinions* into six basic emotion states, including *happiness, surprise, fear, sadness, anger, and disgust*.

### B. Proposed Metrics

**Source opinion.** The opinion of the source publisher is a natural choice. Table I shows the viral probabilities for diffusion events with different *source opinion (SO)*, and we find that *angry* and *surprising* information is more likely to be viral compared to other information such as *happiness*.

**Public opinion composition and weighted public opinion composition.** We also considered the composition of *individual opinions* among all propagators. Furthermore, the social influences of different users are distinguished based on the numbers of their followers, indicating that the opinions of popular users have greater impacts on *weighted public opinion composition*. We notice that both metrics are greatly influenced by the *source opinion*, as shown in Fig. 1, which can be explained as a sympathy-driven phenomenon in online diffusions. The opinion with a maximum proportion in *public opinion composition (MPO)* and *weighted public opinion composition (MWPO)* is defined as the dominant opinion of a diffusion event. Similar to the *SO* mentioned above, the results in Table I reveal the significant impacts of dominant opinion on diffusion virality.

**Shannon entropy of opinion distribution.** The Shannon entropy of opinion distribution is used to evaluate the homophily of public opinion in a diffusion event. We find a significant monotonically increasing curve in Fig. 2.(a), which is not surprising that a contending discussion is more likely to become a viral diffusion, compared with a discussion with identical opinions.

TABLE I. VIRAL PROBABILITY OF DIFFERENT OPINIONS

	Ang.	Dis.	Fea.	Hap.	Sad.	Sur.
SO	0.5839	0.4521	0.4904	0.4292	0.4622	0.5725
MPO	0.5501	0.4364	0.4562	0.4854	0.4364	0.5691
MWPO	0.5456	0.4615	0.4488	0.4823	0.4501	0.6002

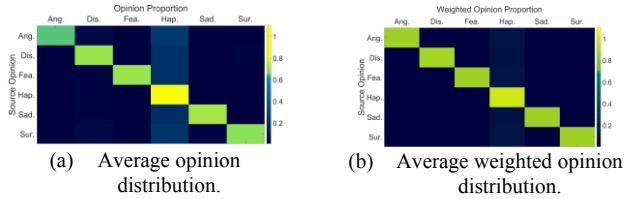


Fig. 1. Opinion correlation with source publisher. The bright color blocks in diagonals show that most users in a diffusion event usually express the same opinion as the root node’s.

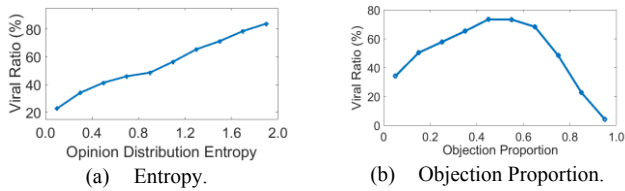


Fig. 2. Viral ratio for varying values of public opinion metrics.

**Objection proportion.** This metric is defined as the ratio that a user expresses an opinion different with its parent node’s in a diffusion tree, which can evaluate the homophily and controversy of individual opinions during the diffusion. We notice in Fig. 2.(b) that it would be more likely to achieve a viral diffusion when the *objection proportion* is around 0.5. In such cases, about a half of users follow the opinions of their parent nodes, while the other half have different opinions. Such a phenomenon shows that both homophily and controversy of public opinion play important roles in online diffusion, and the viral diffusion is driven by both of them collaboratively.

**Within-group opinion distance.** This metric represents the average distance among nodes with the same opinion. Such definition is similar to the structural virality in [1] but combines structural knowledge with public opinion. An opinion with a low average distance may prefer strong ties and homophily, while a longer average distance indicates stronger contagiousness through long travel paths, or a higher likelihood to occur spontaneously. Fig. 3 shows two statistical charts of *within-group opinion distances (WOD)* in various events, in which *happiness*, *surprise*, and *disgust* seem to be more contagious universally, according to our metrics as the higher values of medians and overall distributions.

### III. VIRAL DIFFUSION PREDICTION

To examine effectiveness of public opinion features in the viral diffusion prediction, we use the feature sets including both our proposed metrics and five baseline feature sets from the previous methods [2]. Similar with [2], the prediction tasks are defined as classification problems. By an early observation on a small initial portion (5%) of a diffusion event, we try to predict whether it will eventually be viral.

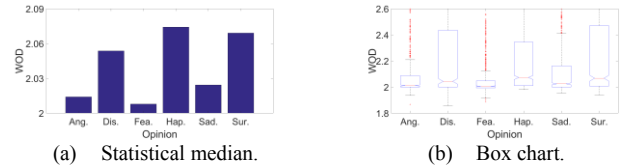


Fig. 3. Within-group opinion distance distribution.

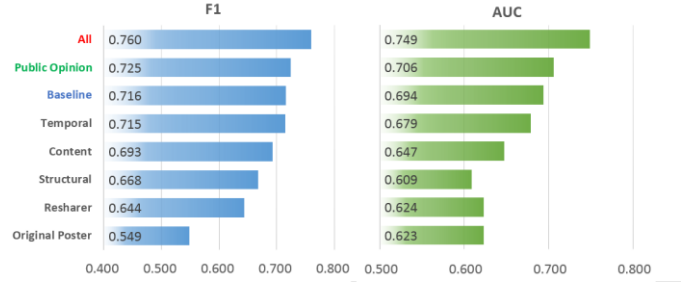


Fig. 4. Viral diffusion prediction with different features.

Performances in F1-score and AUC of logistic regression are reported in Fig. 4. We notice that our proposed features outperform all other individual feature sets, and perform slightly better than that of the baseline method using all other five feature sets. The baseline performance is significantly improved by adding our public opinion features (from 0.716 to 0.760 in F1-score, and from 0.694 to 0.749 in AUC).

### IV. CONCLUSION

In this paper, we have examined the influence of public opinion in online viral diffusion. We have revealed the significant correlation between our public opinion metrics and diffusion virality by statistical analysis, and applied our proposed opinion-based method to improve the performance of viral diffusion prediction. Currently, we are further investigating how the underlying mechanisms drive online diffusion with different opinions, and conducting model-based simulation to verify if the observed correlation is replicated.

### ACKNOWLEDGMENT

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

### REFERENCES

- [1] Goel S, et al. The structural virality of online diffusion[J]. Management Science, 2015, 62(1): 180-196.
- [2] Cheng J, et al. Can cascades be predicted?[C]//Proceedings of WWW'14, 2014: 925-936.
- [3] 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.
- [4] Ekman P. An argument for basic emotions[J]. Cognition & emotion, 1992, 6(3-4): 169-200.