

A Cascade-Driven Approach Towards the Understanding of External Influence on Information Diffusion

Antigoni Maria Founta
Aristotle University of Thessaloniki
Department of Informatics
Thessaloniki, Greece 541 24
founanti@csd.auth.gr

Hakim Hacid
Zayed University
P.O. Box 19282
Dubai, United Arab Emirates
hakim.hacid@gmail.com

Athina Vakali
Aristotle University of Thessaloniki
Department of Informatics
Thessaloniki, Greece 541 24
avakali@csd.auth.gr

ABSTRACT

In this work we experiment with measuring the impact of external influence in the information diffusion process of social networks. Initially, we collect various data from a number of preselected topics from Twitter and make a two-step analysis on these data, in order to determine the factor of external influence. Additionally, we conduct a topic analysis to investigate how topics affect the process of diffusion, as it has been determined from the previous results. Our primitive results find some importance on the influence of the external factor, but do not show any significance yet. This early study, nevertheless, adds valuable insights to the immature research topic of external influence, while it also compares a number of potential factors of influence and discovers correlations between them.

KEYWORDS

Information Diffusion, External Influence, Social Networks

1 INTRODUCTION

With the rise of Web 2.0 and the dawn of Social Networks, tremendous amounts of social data have been produced. In fact, social networks, as their name indicates, are graphs of users that are partially connected according to their relationship status. In these graphs, information is being propagated through their edges and that can happen regardless of the connection between them. Thus, it is important to study alternative influence factors other than social relationship in matters of Information Diffusion. Some of these elements of influence seem apparent, but other equally potential have not been studied extensively. One such important factor that has only been superficially studied is the effect of transmitting information that originates outside the network, or else the External Influence. The main goal of this work is to study the effects of External Influence on the Information Diffusion process of the Twitter network, in order to understand and quantify the importance of this factor. Based on our extensive research on the current state-of-the-art [1] [2] [3], there is no previous work that achieves what we attempt to do.

2 METHODOLOGY

Our approach begins with the construction of diffusion cascades, originating from a dataset of tweets that was collected based on miscellaneous topics. For each of these cascades we calculate various features. The features are consisted of some properties of the tweets as well as some dynamics of the cascades, selected based on ground theory [4].

Afterwards follows our two-step approach to measure external influence and compare it with other major influence factors that we captured as features. The first step of the approach includes a statistical analysis of the cascades, using descriptive statistics and correlations, and the following step is a clustering analysis. For this part, we apply a simple k-means algorithm to the extracted features and investigate the results. Our main target is to recognise patterns of behaviours of external and internal cascades and understand external influence.

The final step of this work is a conducted topic analysis to investigate the impact of topics on our previous results and how it correlates with external influence. We separate the dataset into five trending topics, construct the cascades and calculate the features. Then we apply the same statistical analysis on each one of the topics and observe any differences from the results of our previous analysis on the entire dataset. That way we aspire to understand the effect topics have on the analysed diffusion process and their correlation with the external factor.

3 RESULTS

In general we observe some differences between the performance of the two information categories, namely internal and external. We find that external cascades seem to be more affected by user influence but not significantly. Temporal influence also looks like having an impact on them, but only when the user influence is low. On the other hand, we notice that the basic dissemination actions of Twitter seem to be much more connected to internal cascades, as in most cases the amounts of Favourites and Retweets are almost double. Unfortunately, there was no extracted pattern worth mentioning regarding the cascade dynamics. Finally, the results of the topic analysis show that the behaviour of topics is closely related to the case of the entire dataset. Yet, there are many variations on their statistics, working as an indicator that topics do in fact have an impact on the diffusion of information and might also be affected by the external influence factor.

REFERENCES

- [1] Jiang Li, Jiagui Xiong, and Xiaojie Wang. 2015. Measuring the External Influence in Information Diffusion. In *Mobile Data Management (MDM), 2015 16th IEEE International Conference on*, Vol. 2. IEEE, 92–97.
- [2] Chuang Liu, Xiu-Xiu Zhan, Zi-Ke Zhang, Gui-Quan Sun, and Pak Ming Hui. 2015. How events determine spreading patterns: information transmission via internal and external influences on social networks. *New Journal of Physics* 17, 11 (2015), 113045.
- [3] Seth A Myers, Chenguang Zhu, and Jure Leskovec. 2012. Information diffusion and external influence in networks. In *Proceedings of the 18th ACM SIGKDD international conference on Knowledge discovery and data mining*. ACM, 33–41.
- [4] Jiang Yang and Scott Counts. 2010. Predicting the Speed, Scale, and Range of Information Diffusion in Twitter. *ICWSM 10* (2010), 355–358.