Microscopic Description and Prediction of Information Diffusion in Social Media: Quantifying the Impact of Topical Interests

Przemyslaw Grabowicz
MPI-SWS, Germany

Niloy Ganguly
IIT Kharagpur, India

Krishna Gummadi
MPI-SWS, Germany

ABSTRACT

A number of recent studies of information diffusion in social media, both empirical and theoretical, have been inspired by viral propagation models derived from epidemiology. These studies model propagation of memes, i.e., pieces of information, between users in a social network similarly to the way diseases spread in human society. Naturally, many of these studies emphasize *social exposure*, i.e., the number of friends or acquaintances of a user that have exposed a meme to her, as the primary metric for understanding, predicting, and controlling information diffusion.

Intuitively, one would expect a meme to spread in a social network selectively, i.e., amongst the people who are interested in the meme. However, the importance of the alignment between the topicality of a meme and the topical interests of the potential adopters and influencers in the network has been less explored in the literature. In this paper, we quantify the impact of the topical alignment between memes and users on their adoption. Our analysis, using empirical data about two different types of memes, i.e., hashtags and URLs spreading through the Twitter social media platform, finds that topical alignment between memes and users is as crucial as the social exposure in understanding and predicting meme adoptions. Our results emphasize the need to look beyond social network-based viral propagation models and develop microscopic models of information diffusion that account for interests of users and topicality of information.

1. INTRODUCTION

Numerous studies have shown that information and behaviors propagate over online social networks due to social exposure [3, 6]. As such, the phenomenon of information diffusion in social networks is predominantly modeled by building parallels with the spread of diseases, e.g., as in the independent cascade or linear threshold models [5]. Such approaches make intuitive sense, since we learn information or copy behaviors from our friends and acquaintances. However, is the diffusion of information similar to the diffusion of indistinguishable particles between indistinguishable nodes of a network?

The increasing availability of data characterizing users and content in online social networks as well as the development of reliable topic modeling methods allow us to explore information diffusion much beyond social network. For instance, some of the recently proposed models suggest that latent factors impact information diffusion [1] and measure topical social influence [2]. However, large-scale empirical studies analyzing the extent to which topical alignment impacts diffusion are missing. There are several important and interesting questions which can be asked. For example, how topical alignment compares against social exposure; is it more or less important? Can we measure the extent to which topical alignment helps in predicting information propagation? These are the questions that we explore here.

2. PRESENT WORK

To understand the influence of topical alignment, we require (i) a closed social system enabling information diffusion and (ii) a comprehensive annotation of topicality of users and diffused pieces of information. On the one hand, we analyze a dataset from Twitter containing almost all tweets that were messaged during the observation period of three months and a snapshot of the follower graph from the end of that period. We consider two distinct types of information that diffuse over the follower graph, namely hashtags and URLs, which for simplicity we jointly call *memes*. We say that a user *diffuses* or *adopts* a meme if she writes it in one of her tweets. On the other hand, we infer with an unsupervised method, namely with LDA, the topical distributions, i.e., the *topicality*, of active users and popular memes. To this end, we use either nouns and proper nouns extracted from tweets or an independent dataset of expertise tags [4]. We obtain qualitatively and quantitatively the same results with both methods. We leverage these datasets to understand the impact of topical alignment, defined as a cosine similarity between topicalities, on information diffusion.

2.1 Adoption probability

We measure adoption probability as a function of social exposure κ following a standard definition [6]. We extend this definition to include the topical alignment $S(\text{user}, \text{meme})$ (details omitted). In Figure 1, we show the probabilities of adoption $P_a(\kappa)$ and $P_a(S(\text{user}, \text{meme}))$. We find that the probability of adoption increases with social exposure, as well as with the topical alignment between user and meme,

WWW 2015 Companion, May 18–22, 2015, Florence, Italy.
ACM 978-1-4503-3473-0/15/05.
http://dx.doi.org/10.1145/2740908.2744106.

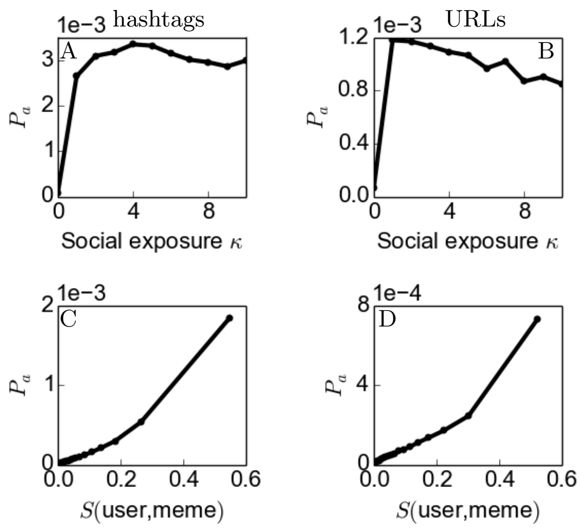

Figure 1: The adoption probability of hashtags and URLs as a function of: (A-B) social exposure κ, (C-D) topical alignment $S(\text{user}, \text{meme})$ for hashtags and URLs.

both for hashtags and URLs. The maximum value of the adoption probability is slightly larger for the social exposures than for the topical alignments, however, the adoption probability grows much more steadily with the topical similarity, while it rapidly reaches a plateau with just a couple of social exposures. We interpret this results as a strong signal that the topical similarity $S(\text{user}, \text{meme})$ can predict adoptions possibly even better than social exposure κ.

Furthermore, we deduce that neighbors in the Twitter follower graph are topically alike (details omitted). However, there are other aspects of topical alignments that impact information diffusion beyond the topical homophily of the graph (details omitted).

2.2 Predictions

We have shown that topical alignment and social exposure increase the probability of adoption. In this section we go one step ahead and check the predictive power of {social exposure, topical alignment, user, meme}-based features in predicting the adoption of a given meme by a particular user. To this end, we consider all users and memes that are in our preprocessed dataset, namely a few thousands of memes and half million of users. We split the user-meme pairs into balanced training and test sets. Namely, we use past user-meme pairs for training and future pairs for testing.

We present results from two classifiers, i.e., Random Forest (RF) and SVM with linear kernel (SVM). Both of the classifiers achieve very good results (Table 1). The RF classifier obtains the F1 score, which is a harmonic average of precision and recall, of 0.93 for hashtags and 0.90 for URLs. Next, we train the classifiers for the groups of features of our interest (Table 1). Most importantly, we find that topical alignments[1] are more predictive than social exposure for both meme types, independently of the classifier. Furthermore, we notice that it is harder to make predictions for URLs, because social exposure and user and meme features have less predictive power for URLs than for hashtags. Interestingly, the topical alignments remain the same highly predictive both for hashtags and URLs.

[1]Three different topical alignments are considered: $S(\text{user}, \text{meme})$, $S(\text{user}, \text{exposer})$, $S(\text{exposer}, \text{meme})$. We omit their detailed description.

	Classifier			
	For hashtags		For URLs	
Features	RF	SVM	RF	SVM
All	0.93	0.91	0.90	0.86
Social	0.73	0.74	0.66	0.67
Topical	0.79	0.80	0.80	0.82
User	0.82	0.82	0.74	0.67
Meme	0.72	0.74	0.59	0.63

Table 1: F1 scores of the prediction of who will adopt what for four groups of features. We focus on the comparison between social exposure and topical alignments.

3. DISCUSSION

This study reveals the importance of topical alignment for the diffusion of memes in Twitter. We find that topical alignment is a mechanism of equal importance for the diffusion of memes as viral spreading. In this context, present study suggests a novel perspective on information diffusion by providing empirical evidence on the matter. Future works, using similar methodology, could address similar questions in other social media platforms or could explore other types of social behaviors that happen in such systems.

4. REFERENCES

[1] N. Barbieri, F. Bonchi, and G. Manco. Topic-Aware Social Influence Propagation Models. *2012 IEEE 12th Int. Conf. Data Min.*, pages 81–90, Dec. 2012.

[2] B. Bi, Y. Tian, Y. Sismanis, A. Balmin, and J. Cho. Scalable topic-specific influence analysis on microblogs. *Proc. 7th ACM Int. Conf. Web search data Min. - WSDM '14*, pages 513–522, 2014.

[3] D. Crandall, D. Cosley, D. Huttenlocher, J. Kleinberg, and S. Suri. Feedback effects between similarity and social influence in online communities. *Proceeding 14th ACM SIGKDD Int. Conf. Knowl. Discov. data Min. - KDD 08*, page 160, 2008.

[4] S. Ghosh, N. Sharma, F. Benevenuto, N. Ganguly, and K. Gummadi. Cognos: Crowdsourcing Search for Topic Experts in Microblogs. In *Proc. 35th Int. ACM SIGIR Conf. Res. Dev. Inf. Retr. - SIGIR '12*, number 1, page 575, New York, New York, USA, 2012. ACM Press.

[5] D. Kempe, J. Kleinberg, and E. Tardos. Maximizing the spread of influence through a social network. In *Proc. ninth ACM SIGKDD Int. Conf. Knowl. Discov. data Min. - KDD '03*, page 137, New York, New York, USA, 2003. ACM Press.

[6] D. M. Romero, B. Meeder, and J. Kleinberg. Differences in the mechanics of information diffusion across topics. In *Proc. 20th Int. Conf. World wide web - WWW '11*, page 695, New York, New York, USA, 2011. ACM Press.