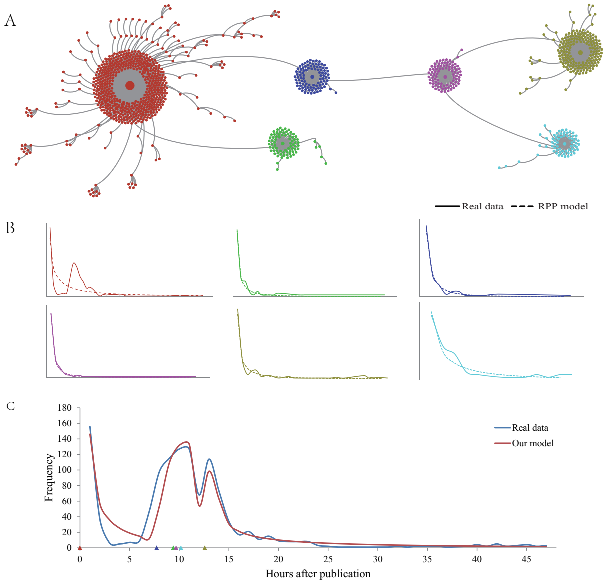

Figure 1: Retweeting dynamics of a message in Sina Weibo. (A) Diffusion tree of the retweeting process; (B) Subprocesses, each being initiated by a key node; (C) Retweeting dynamics of the message.

up to time t. In this paper, we use log-normal function as the relaxation function as in [2].

The probability of the i-th retweet arrives at t_i, given the $(i-1)$-th retweet arrives at t_{i-1}, can be written as

$$p_1(t_i|t_{i-1}) = x(t_i)\exp\left(-\int_{t_{i-1}}^{t_i} x(t)dt\right), \quad (2)$$

and the probability of no retweet between $[t_n, T]$ is

$$p_0(T|t_n) = \exp\left(-\int_{t_n}^{T} x(t)dt\right). \quad (3)$$

Therefore, the likelihood of the retweeting dynamics $\{t_i\}_{i=1}^n$ during $[0, T]$ is

$$\mathcal{L} = p_0(T|t_n)\prod_{i=1}^{n} p_1(t_i|t_{i-1}). \quad (4)$$

The parameters (i.e., λ_l, θ_l, τ_l) are estimated by maximizing the log likelihood, and the retweeting dynamics of message could be predicted as

$$c(t) = n * \exp\left(\int_T^t \sum_{l=1}^{k} \lambda_l f(s-\tau_l;\theta_l)ds\right). \quad (5)$$

To evaluate the effectiveness of the proposed model, we compare it with the RPP model [2] on a dataset crawled from Sina Weibo, the largest microblogging platform. We filter out the tweets with less than 100 retweets, and obtain the retweeting dynamics of 164 tweets. For each tweet, we use its retweeting dynamics in the first two hours for training, and predict its dynamics at the following hours. In the experiments, we set k=3 with the prior that few tweets has more than 3 key nodes in its retweeting dynamics in the first two hours. We use MAPE (mean average percentage error) to evaluate prediction accuracy. Results are shown in Fig. 2.

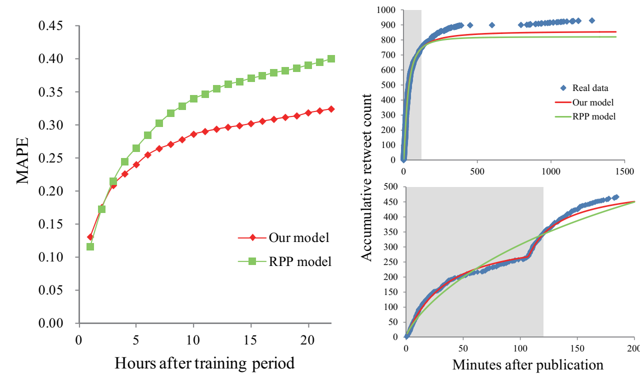

Figure 2: Comparison with the RPP model.

To offer some intuitions about the comparison, we also give two examples. We can see that our model not only fits well the retweeting dynamics in training period, but also has a more desirable prediction accuracy than competing model.

3. CONCLUSION

In this paper, we proposed a mixture process to model and predict retweeting dynamics in social media. The idea is motivated by the fact that the whole process of retweeting dynamics could be effectively captured by a handful of subprocesses initiated by retweeters with high number of followers. The proposed model is efficient to be trained using only the retweeting dynamics of individual tweet. Experimental results demonstrate the effectiveness of the proposed model, compared with the state-of-the-art model.

4. ACKNOWLEDGMENTS

This work was funded by the National Basic Research Program of China (the 973 program) with grant numbers 2014CB340401 and 2013CB329602, and the National Natural Science Foundation of China with grant numbers 61472400, 61232010, and 61572467.

5. REFERENCES

[1] G. Szabo and B. A. Huberman. Predicting the popularity of online content. *Communications of the ACM* 53(8):80–88, 2010.

[2] H. W. Shen, D. Wang, C. Song, and A.-L. Barabási. Modeling and predicting popularity dynamics via reinforced poisson processes. In *AAAI 2014*, pp. 291–297.

[3] S. Gao, J. Ma, and Z. Chen. Modeling and predicting retweeting dynamics on microblogging platforms. In *WSDM 2015*, pp. 107–116.

[4] R. Crane and D. Sornette. Robust dynamic classes revealed by measuring the response function of a social system. *PNAS*, 105(41):15649–15653, 2008.

[5] P. Bao, H. W. Shen, X. Jin, and X. Q. Cheng. Modeling and predicting popularity dynamics of microblogs using self-excited hawkes processes. In *WWW 2015*, pp. 9–10.

[6] Q. Zhao, M. A. Erdogdu, H. Y. He, A. Rajaraman, and J. Leskovec. SEISMIC: A Self-Exciting Point Process Model for Predicting Tweet Popularity. In *KDD 2015*, pp. 1513–1522.