Accepted Manuscript

Modeling of information diffusion on social networks with applications to WeChat

Liang Liu, Bo Qu, Bin Chen, Alan Hanjalic, Huijuan Wang

PII: S0378-4371(17)31278-5
DOI: https://doi.org/10.1016/j.physa.2017.12.026
Reference: PHYSA 18956

To appear in: *Physica A*

Received date : 20 September 2017
Revised date : 25 November 2017

Please cite this article as: L. Liu, B. Qu, B. Chen, A. Hanjalic, H. Wang, Modeling of information diffusion on social networks with applications to WeChat, *Physica A* (2017), https://doi.org/10.1016/j.physa.2017.12.026

Modeling of Information Diffusion on Social Networks with Applications to WeChat

Liang Liu[a,b], Bo Qu[b], Bin Chen[a], Alan Hanjalic[b], Huijuan Wang[b,*]

[a] *College of Information System and Management, National University of Defense Technology, Changsha, China, 410073*
[b] *Faculty of Electrical Engineering, Mathematics, and Computer Science, Delft University of Technology, Mekelweg 4, Delft, The Netherlands, 2628 CD*

Abstract

Traces of user activities recorded in online social networks open new possibilities to systematically understand the information diffusion process on social networks. From the online social network WeChat, we collected a large number of information cascade trees, each of which tells the spreading trajectory of a message/information such as which user creates the information and which users view or forward the information shared by which neighbors. In this work, we propose two heterogeneous non-linear models, one for the topologies of the information cascade trees and the other for the stochastic process of information diffusion on a social network. Both models are validated by the WeChat data in reproducing and explaining key features of cascade trees.

Specifically, we apply the Random Recursive Tree (RRT) to model the growth of cascade trees. The RRT model could capture key features, i.e. the average path length and degree variance of a cascade tree in relation to the number of nodes (size) of the tree. Its single identified parameter quantifies the relative depth or broadness of the cascade trees and indicates that information propagates via a star-like broadcasting or viral-like hop by hop spreading. The RRT model explains the appearance of hubs, thus a possibly smaller average path length as the cascade size increases, as observed in WeChat. We further propose the stochastic Susceptible View Forward Removed (SVFR) model to depict the dynamic user behavior including cre-

*Corresponding author
Email address: h.wang@tudelft.nl (Huijuan Wang)