Modification of ICM for social media information propagation study

This work is devoted to a modification of the independent cascades model (ICM, which is widely used for network contagion modeling), and to application of the modified model to real data of information propagation in the Vkontakte social network.

An information propagation process in social media has some issues that constrain a pure independent cascade model usage. Firstly, some nodes and links are not observable in the real network (hidden profiles, hidden friends). Secondly, the probability to pass a piece of information can differ for different nodes (popular publics and users vs. unpopular fake accounts with cheat subscribers). This difference is too significant to ignore by assuming that the probability to pass the information from one node to another is the same for all connected pairs of nodes. On the other hand, the assumption that all these probabilities are different (classic ICM) is unacceptable because of a relatively small number of observed cascades by comparison with the number of model parameters.

We propose to take into account these issues by adding the following modifications. The first is to assign to each publication two parameters of spreading: one is the probability of propagation through observable tie (just like in ICM where all probabilities are equal) and the other is the probability of propagation through unobservable tie (the same, but in assumption that network is fully connected). The second is to assign to each node a contagiousness modification parameter. This value is a multiplier for the probability of propagation through a tie going out of the node. We also assume that the most nodes in the network are not “unusual” in terms of information spreading (common users). Therefore we propose a mechanism to select nodes that should be treated like “uncommon”, so we should add into a model their contagiousness modification parameters. It significantly reduces the number of parameters. The byproduct of the approach is that we got a new measure of node ranking regarding their actual impact in contagion (information spreading).

Modified model is implemented as an open-source python/C++ library, allowing it to work with big networks (several million nodes). To test the model we use a dataset of more than 30000 cascades of charity-related publications, observed on Vkontakte social network. We compare the model with the classic contagion models. Also we study the interaction between two parameters of contagion through observable and unobservable ties.