Modeling Influence Diffusion in Social Networks for Viral Marketing

Wenjun Wang and W. Nick Street

Department of Management Sciences, University of Iowa, Iowa City, IA 52242, USA
wenjun-wang@uiowa.edu
nick-street@uiowa.edu

1 Introduction

Modeling influence diffusion in social networks is an important challenge. There are a large number of models in the literature addressing influence diffusion and viral marketing [2], [4]. However, there exist some significant limitations in these models. First, they approach a social entity's adoption likelihood from a confined scope, considering only the direct influence from the activated neighbors of the entity. Second, most models overlook the fact that influence does not remain static or constant, but rather attenuates along diffusion paths and decays with time. Third, these models fail to capture the individual temporal diffusion dynamics. In this paper, we propose a novel *multiple-path asynchronous threshold* (MAT) model to address these issues, and evelop an effective and efficient heuristic to tackle the influence-maximization problem.

2 MAT Model

We categorize all the nodes in the network into two types: influencers and messengers. An *influencer* is an active node that has adopted the product and can originate and spread its influence in the network. A *messenger* is an inactive node that may acquire influence and pass the influence on to others. Once a messenger acquires influence that is greater than or equal to its threshold, it is activated and turns into an influencer who starts to spread out its own influence. Our model captures an important characteristic of word-of-mouth (WOM) communication in that: *anyone can pass along WOM messages and potentially influence the recipient*. In other words, a node can be activated by not only the *direct influence* from its active neighbors (influencers), but also the *indirect influence* passed along by its inactive neighbors (messengers). This is a distinguishing and more realistic feature built in our MAT model.

To differentiate the relationship strength on influence, we introduce a *weight normalization scheme* that measures the fraction of influence a node receives from a specific in-neighbor relative to the total influence it receives from all of its in-neighbors. To quantify the influence attenuation along a diffusion path, we define a depth-associated attenuation coefficient $\alpha = d^{-2}$, where d is the depth (number of hops) from an influencer to the node of interest along the diffusion path. It can be interpreted as a compounding factor that incorporates the trustworthiness decay, information corruption, and decreasing reaching probability. Specifically, we set the depth limit d_{max} to 3 for each influencer to capture the *three-degrees-of-influence* phenomenon [3]. On the other hand, we model the temporal influence decay as an exponential function of time, $I(t) = e^{-\lambda t}$, where λ is a user-specified tunable parameter of decay rate. It can be tuned to account for different products on various social networks. To capture the individual temporal diffusion dynamics, we model the heterogeneity of WOM messaging from a node to its neighbors as a *Poisson process* with a rate that is determined by the node's relative activness at both local and global level in terms of its out-link weights.

The diffusion process starts with an initial set of influencers (seed nodes) S_0 with $|S_0| = K$, and unfolds in discrete time steps. At each time step, the influence is propagated one hop from a node u to each out-neighbor v with a probability based on their contact frequency and node u's global activenss. Each inactive node v is assigned with an activation threshold selected uniformly at random in the range $[0,1]$. When the total influence that v receives is greater than or equal to it threshold, it is activated and turns into an influencer. Then it not only continues passing other influencers' influence as a messenger, but also starts to spread out its own influence as an influencer. The diffusion process stops when the number of hops of influence diffusion of each influencer reaches the depth limit (set to 3 by default) and no new activation is possible.

3 IV-Greedy Algorithm

The influence-maximization (IM) problem is to find a small set of K seed nodes (initial adopters) who can trigger the largest further adoptions in the network. The IM problem is NP-hard under the MAT model. Using the *influence vector* (IV) of each node, we develop a heuristic algorithm called *IV-Greedy*. The influence vector of a node captures where and how much influence it spreads out in its neighborhood based on a *static* version of the MAT model, in which we ignore the temporal diffusion decay, individual diffusion dynamics, and the activation of any nodes. We use it as a proxy for the *dynamic* influence diffusion so as to avoid the expensive Monte Carlo (MC) simulation. We sweep over the influence vector of each node to repeatedly pick the node with the *maximum marginal gain* and add it to the seed set until all K seeds are found.

4 Experiments

To evaluate our MAT model and the performance of IV-Greedy, we conduct experiments on three widely-used real-life network datasets, which include PGP [1], NetHEPT [1], and WikiVote [2]. We compare the performance of IV-Greedy against a set of baseline algorithms in terms of both influence spread and time efficiency. The simplest baseline is to select the seed nodes uniformly at random. The most frequently used is the *degree-centrality* heuristic, in which the seed nodes are chosen in descending order of their out-degrees. The next baseline is the Top-K algorithm, which selects the top K nodes with the largest *individual* influence spread based on MC simulation. As shown in Fig. 1 and Fig. 2, IV-Greedy achieves the largest influence spread, and it is up to three orders of magnitude faster than Top-K. IV-Greedy is the best performing algorithm overall.

[1] http://research.microsoft.com/en-us/people/weic/graphdata.zip

[2] https://snap.stanford.edu/data/wiki-Vote.html

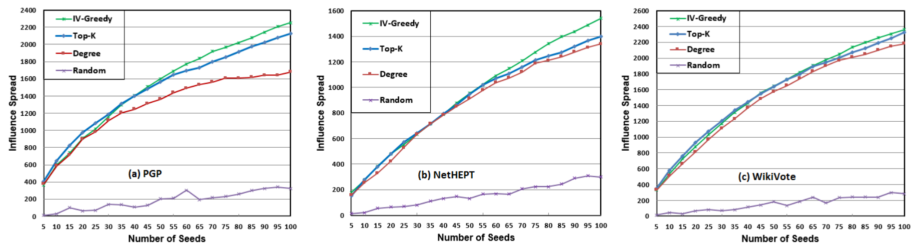

Fig. 1. Performance comparison on influence spread

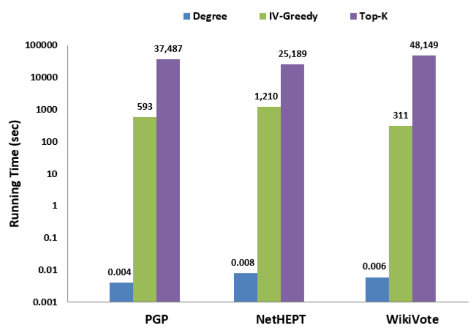

Fig. 2. Performance comparison on running time (CPU seconds)

5 Conclusion

In this paper, we propose a novel *multiple-path asynchronous threshold* (MAT) model for viral marketing in social networks. Our MAT model captures both direct and indirect influence, influence attenuation along diffusion paths, temporal influence decay, and individual diffusion dynamics. It is an important step toward a more realistic diffusion model. Further, we develop an effective and efficient heuristic, IV-Greedy, to tackle the influence-maximization problem. Our experiments on three real-life networks demonstrate its excellent performance in terms of both influence spread and time efficiency. Our work provides preliminary but important insights and implications for diffusion research and marketing practice.

References

1. Boguna, M., Pastor-Satorras, R., Diaz-Guilera, A., Arenas, A.: Models of social networks based on social distance attachment. Phys. Rev. E 70, 056122 (2004)
2. Chen, W., Lakshmanan, L., Castillo, C.: Information and Influence Propagation in Social Networks. Morgan & Claypool Publishers (2013)
3. Christakis, N.A., Fowler, J.H.: The spread of obesity in a large social network over 32 years. New England J of Medicine 357, 370–379 (2007)
4. Kempe, D., Kleinberg, J., Tardos, E.: Maximizing the spread of influence through a social network. Theory of Computing 11(4), 105–147 (2015)