Errata to "Scalable influence maximization for independent cascade model in large-scale social networks"

Chi Wang, Wei Chen, Yajun Wang

May 28, 2013

1 The issue with the Prefix excluding MIA model

In Section 3.3 of [1], we define the Prefix excluding MIA (PMIA) model. In particular, after several seeds are selected, we compute a maximum influence path from node u to node v avoiding all previously selected seeds. One can then define the influence spread function $\sigma_P(\cdot)$ based on the model, as in Equation (3.2) in [1]. We claimed that $\sigma_P(\cdot)$ is sequential submodular in Theorem 5 in [1]. However this claim is wrong. We give a counter example in Figure 1 to demonstrate that. The numbers on the edges are the propagation probabilities of the edges.

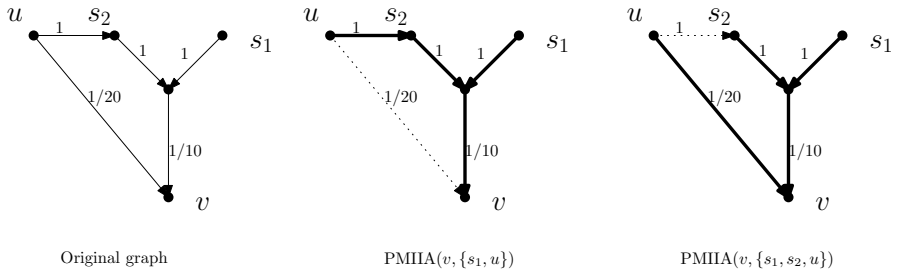

Figure 1: The counter example

Consider the leftmost graph in Figure 1. We are interested in studying the marginal influence spread of node u. Without loss of generality, we assume the weight of v is large, e.g., this can be simulated by a long chain of nodes from v such that the probabilities on the edges is 1. In particular, it is sufficient to study the activation probability of v. When only s_1 is selected, i.e., the middle case, the marginal probability that u activates v with MIP $u \to s_2 \to \cdots \to v$ is 0. However, when both s_1 and s_2 are selected (in the right case), the PMIA model will select another path with probability $1/20$ from u to v. In particular, the marginal probability that u activate v in this case is $\frac{1}{20} \cdot (1 - \frac{1}{10}) > 0$. Therefore, the function $\sigma_P(\cdot)$ is not sequential submodular.

2 The implications of the issue

First, Theorem 5 in [1] does not hold. The problem comes from the fact that in PMIA model, the underlying arborescences are changing. On the other hand, in MIA model, these arborescences are fixed. Therefore, the submodularity in MIA model hence Theorem 3 in [1] still hold.

Second, it is possible to recover the submodularity by revising the PMIA model significantly. More specifically, we have to compute maximum *marginal* influence paths instead. Furthermore, the local arborescences formed by the selected seeds can only increase monotonically so that the path search space always shrinks. However, such revision is mainly for satisfying sequential submodularity and does not reflect significant practical improvement. Therefore, we consider it as only of theoretical interest and do not provide further details here.

Finally, since both PMIA and MIA are heuristics for the influence maximization problem, we believe PMIA is still a valuable heuristic as demonstrated by our empirical evaluations, despite the lack of submodularity (and hence the approximation) we mistakenly claimed in [1], for which we sincerely apologize.

References

[1] Chi Wang, Wei Chen, and Yajun Wang. Scalable influence maximization for independent cascade model in large-scale social networks. *Data Mining and Knowledge Discovery*, 25(3):545–576, 2012.