Influence maximisation in real-world social networks

Valerio Restocchi[1], Soheil Eshghi[2] , Setareh Maghsudi[3] , Leandros Tassiulas[2] , Sebastian Stein[1]

(1)University of Southampton, Electronics and Computer Science, Southampton, UK
(2)Yale Institute for Network Science, Yale University, New Haven, USA
(3)Electrical Engineering and Computer Science, Technische Universität Berlin , Berlin, Germany

Influence maximisation, i.e., the study of strategically influencing agents that aim at spreading opinions, behaviours, or rumors on social networks, has been extensively investigated in recent work, mostly via variants of the seminal independent cascade model [1]. However, algorithms that provide optimal and nearly-optimal solutions to the problem of triggering an influence cascade, by selecting the optimal seed node sets, often achieve this only if certain assumptions are met. Among these assumptions, the most stringent and common is the availability of complete information about a social network. Not surprisingly, this is usually not the case, especially when dealing with Big Data and offline social etworks. In this presentation, we discuss our most recent work on influence maximisation with partial information, and focus specifically on our results on partially observable networks. We analyse how the effectiveness of influence maximisation strategies varies depending on visibility and properties of the network topology, such as degree distribution, community size distribution, overlaps between communities, and distribution of link weights. We find that, in most cases, simple heuristics significantly outperform algorithms with theoretical performance guarantees (up to around 50%) at spreading influence when only a small portion of the network is visible.
Interestingly, we also find that, for most of the network topologies we consider, such state-of-the-art algorithms achieve only marginally better results than both the proposed heuristics and random selection of seeds, even on highly-visible networks.
From our results we draw two important conclusions: first, we argue that, when designing new influence maximisation algorithm, complete information is too stringent an assumption, as results obtained within this framework often do not transfer to real-world networks. Second, we consider the practicality of influence maximisation algorithms. In particular, our results suggest that the performance of algorithms is highly variable and depends on the network topology, and that algorithms with theoretical guarantees are usually far more computationally expensive than simple heuristics (the running time difference is typically in 2-3 order of magnitudes). Given these observations, we argue that, for real-world networks, and especially for large ones, it is better to design tailored heuristic algorithms that can outperform more generic ones, both in terms of the size of the influence cascade, and computational costs.

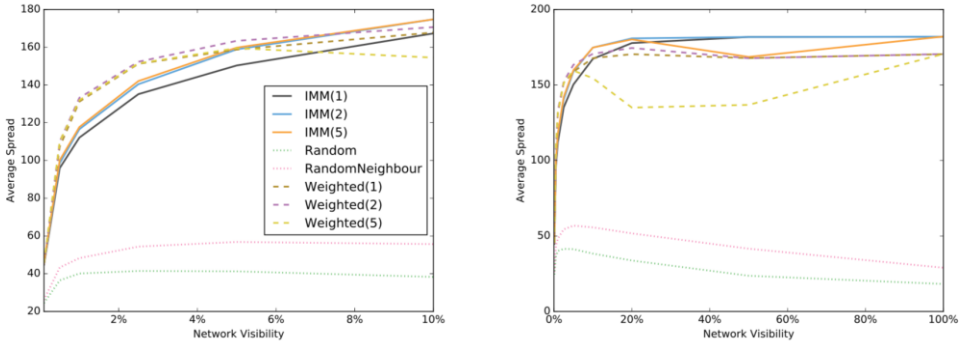

Figure 1: Average spread the NetHept collaboration network for 5 seeds. We compare a state-of-the-art algorithm with theoretical guarantees (IMM(1)) with several heuristics designed for partially observable networks and a random selection benchmark. These results show that the heuristics outperform IMM(1) (up to 29%) when network visibility is lower than 20%.

[1] D. Kempe et al., ”Maximizing the spread of influence through a social network.”, Proceedings of the 9th International Conference on Knowledge discovery and Data Mining (KDD), Washington, DC, USA (2003).