

Heuristic Algorithms for Influence Maximization in Partially Observable Social Networks

Sebastian Stein¹, Soheil Eshghi², Setareh Maghsudi², Leandros Tassioulas²,
Rachel K. E. Bellamy³, and Nicholas R. Jennings⁴

¹ University of Southampton, Southampton, UK
ss2@ecs.soton.ac.uk

² Yale University, New Haven, USA

{soheil.eshghi, setareh.maghsudi, leandros.tassioulas}@yale.edu

³ IBM T.J. Watson Research, Yorktown Heights, NY, USA
rachel@us.ibm.com

⁴ Imperial College London, UK
n.jennings@imperial.ac.uk

Abstract. We consider the problem of selecting the most influential members within a social network, in order to disseminate a message as widely as possible. This problem, also referred to as *seed selection* for influence maximization, has been under intensive investigation since the emergence of social networks. Nonetheless, a large body of existing research is based on the assumption that the network is completely known, whereas little work considers partially observable networks. Yet, due to many issues including the extremely large size of current networks and privacy considerations, assuming full knowledge of the network is rather unrealistic. Despite this, an influencer often wishes to distribute its message far beyond the boundaries of the known network. In this preliminary study, we propose a set of novel heuristic algorithms that specifically target nodes at this boundary, in order to maximize influence across the whole network. We show that these algorithms outperform the state of the art by up to 38% in networks with partial observability.

1 Introduction

Today's vast online social networks connect people across geographical, cultural and organizational boundaries. This offers a tremendous opportunity for quickly disseminating a message to a global audience with only a limited budget. This could be exploited for product marketing, but also for quickly mobilizing large crowds to help during humanitarian crises [15], for collecting data through participatory sensing [23] or for supporting machine intelligence through human computation and crowdsourcing [20]. However, tapping into this vast distributed processing and information resource requires organizations to effectively disseminate a message (e.g., a call for action or participation) far beyond their usual domain of direct influence (e.g., their existing customers or social media followers). This can be achieved by asking key individuals to

Copyright ©2017 for the individual papers by the papers' authors. Copying permitted for private and academic purposes. This volume is published and copyrighted by its editors.

spread the message to contacts within their respective social networks, ideally creating far-reaching cascades of repeated messages.

How to choose such key individuals within a large network has been widely studied in the computer science literature and is typically referred to as *influence maximization* (IM) [7, 12, 10, 3]. This body of work provides models that allow the prediction of network spread, given a fully known network, enabling those who wish to have their messages reach the most people, e.g., advertisers, marketers or public health professionals, to choose the best few people to target with their initial messaging.

Although IM has attracted a great deal of attention in the past few years, the majority of existing work focuses on known networks, rather than the real-world situation where the full network is not known. In essence, little work considers partially observable networks, where only part of a network is visible to the decision maker. This is especially relevant, for instance, when an organization may be aware of its own members' network links, but has little information about the wider network.

In this paper, we take the first steps towards addressing this problem. Specifically, we survey the state of the art and propose a new model for IM in partially observable networks. Furthermore, we propose a number of heuristic algorithms that target nodes at the boundary of the known network and, through an empirical evaluation on a real-world network, we show that these can outperform the current state of the art by up to 38% in settings with very limited observability.

2 Related Work

In the following we first introduce the general influence maximization problem in fully observable settings and then we detail work that has looked at partially observable settings.

2.1 IM in Fully Observable Networks

In the influence maximization problem, a (fully observable) network is specified along with an *activation function* that describes the spread of influence. Initially, no individuals in this network are *activated*, i.e., influenced by the message. Given this, the decision maker picks a fixed number of individuals as seeds. These become activated and spread the message, i.e., activate their neighbors, according to the specified activation function. The aim of the decision maker is to maximize the total number of activated individuals [10].

While solving this problem is NP-hard in general, there are several computationally efficient approximate solutions with performance guarantees [10, 14, 6, 21, 3]. Some of these scale to realistic-sized networks with millions of nodes or more.

This work focuses on partially observable networks, and hence we avoid a detailed discussion of the state-of-the-art IM schemes for fully observable networks. Nonetheless, we describe the state-of-the-art IM algorithms we use in our numerical analysis.

Currently, the most successful IM algorithms with theoretical performance guarantees [1] are TIM and TIM+ [19], and IMM [18]. These algorithms are based on the concept of *reverse reachable sets* [3], defined below.

Definition 1 (Reverse Reachable Set). Let v be a node in some graph G , and g be a graph obtained by removing each edge e in G with probability $1 - p(e)$. The reverse reachable (RR) set for v in g is the set of nodes in g that can reach v . That is, for each node u in the RR set, there is a directed path from u to v in g .

Theoretically, it can be shown that if an RR set generated for v overlaps with a node set S with probability ρ , then when we use S as the seed set to run an influence propagation process on G , v will be activated with probability ρ .

These algorithms first determine how many RR sets need to be generated to provide statistical guarantees for the IM problem. Choosing too many RR sets increases the computational complexity, while choosing too few RR sets affects the algorithm's performance guarantees. TIM/TIM+ predetermine the required number of RR sets and then generate them, while IMM produces RR sets one after another until a certain stopping condition is satisfied. Once the RR sets are generated, a greedy heuristic for the max-cover problem derived from these RR sets is applied to select the seeds, which is guaranteed to at least be within a factor $(1 - \frac{1}{e})$ of the optimal case.

2.2 IM in Partially Observable Networks

Influence maximization under uncertainty corresponds to a general class of problems, where given limited network information and a fixed budget, the desire is to allocate the budget to a set of seeds, so that the spread of influence is maximized. In particular, the probability of influence among each pair of nodes and/or the network structure might be (partially) unknown. This problem has been investigated in the literature in two general settings:

One shot: Here, seed selection is performed only once given the available information. For example, robust influence maximization [5] addresses uncertainty in the edge influence probability, which is represented as an interval in which the true probability may lie. Its goal is to maximize the worst-case ratio between the influence spread of the chosen seeds and the optimal set of seeds.

Repeated: Here, seeds are selected incrementally, so that sampling/learning approaches can be used to acquire information about the *a priori* unknown propagation characteristics. These work can be classified into two sub-groups:

- *Bandit-based approaches:* In [16], a combinatorial bandit approach is used to model IM, where the agent selects a set of nodes as seeds at every round, and incurs a regret due to not selecting the optimal set. Existing regret-minimizing algorithms are used to solve the problem. In another bandit-based IM method [11], the seed selection scheme has two phases: (i) Exploration using a conventional influence maximization algorithm and (ii) Exploitation using a bandit algorithm. In influence maximization with semi-bandit feedback [22], upon activating the seeds, the learner only observes the influenced part of the network and uses an algorithm based on the well-known UCB strategy to select seeds. This latter algorithm can also accommodate a combinatorial bandit setting, since at every round a set of nodes can be selected.

- *Other approaches*: [9] learn the diffusion probability in an information-cascade model by likelihood maximization, and develop an expectation maximization seed selection algorithm. [17] investigates a scenario in which seed selection is performed under partial feedback, where at every round the result of past action(s) is not completely revealed. This creates a trade-off between delay (waiting time to see a desired partial outcome) and efficiency of seed selection in terms of spread of influence. This is somewhat similar to the trade-off in influence maximization in changing networks [24], which is between recency and accuracy of network information and efficiency of spread. Finally, [13] proposes a heuristic algorithm which at every round probes a set of nodes which have a high probability of large degrees. All probed nodes are ranked based on their degree, and nodes with the highest ranks are selected as seeds.

In this work, we explore influence maximization under a type of uncertainty which has not been investigated so far. In our setting, a part (or some parts) of a network is known (e.g., individuals that belong to the decision maker’s organization), while the rest is completely unobservable. We investigate the applicability of the state-of-the-art approaches, and propose new heuristic algorithms for seed selection.

3 Model and Problem

We consider a dense network modeled by a directed graph $G(\mathcal{V}, \mathcal{E}, w)$ and called an influence network:

- $\mathcal{V} = \{1, 2, \dots, N\}$ indicates the set of nodes. Each node represents a network entity, for instance an individual. G represents a social network;
- $\mathcal{E} \subseteq \mathcal{V} \times \mathcal{V}$ is the set of directed edges;
- $w : \mathcal{E} \rightarrow [0, 1]$ is a weight function, where $w(i, j)$ is the likelihood (possibility) of j being influenced by i . Alternatively and with a slight change of notation, w can be defined as the weight matrix.

In an influence network, initially a seed set $\mathcal{S} \subset \mathcal{V}$ becomes activated. Every activated node causes its neighbors to become (and stay) activated according to the *independent cascade* model [8]. In this model, an activated node $i \in \mathcal{V}$ might activate each neighbor j with probability $w(i, j)$. In a *weighted cascade* model, we have $w(i, j) = \frac{1}{|\{(k, j) \in \mathcal{E}\}|}$, where $|\{(k, j) \in \mathcal{E}\}|$ is the in-degree of j .

It should be mentioned that in addition to the independent- and weighted cascade models, the spread of influence through a network can be described by some other models, one of them being the well-known *linear threshold* model. In this work, however, we confine our attention to the cascade influence propagation model.

Departing from existing models in the literature, we assume that the network is only partially observable. Specifically, we assume that the decision maker, who selects the seeds, observes only one part of the network, denoted by $G' = (\mathcal{V}', \mathcal{E}', w')$. Naturally we have $\mathcal{V}' \subseteq \mathcal{V}$, $\mathcal{E}' \subseteq \mathcal{E} (\subseteq \mathcal{V}' \times \mathcal{V}')$, and $w' : \mathcal{E}' \rightarrow [0, 1]$. Note that the range of function w' includes the weights of the edges belonging to \mathcal{E}' .

One specific case of partially observable networks is the *organization-partitioned* model of network. In such networks, a subset of nodes, $O \subseteq \mathcal{V}$, reveal their neighbors (and possibly the relevant edge-weights) to the decision makers, e.g., because they are members of the decision maker’s organization. We denote this subset of nodes as *fully observable* nodes, while the neighbors of nodes in O (that are not also in O) are denoted as *boundary* nodes. We assume only nodes in O can be selected as initial seeds. Figure 1 provides an illustrative example of an organization-partitioned network.

Concluding this section, we define the general problem of *influence maximization in partially observable networks* as follows: Given a specific activation function, a seed set size M and the observable part \mathcal{G}' of a network \mathcal{G} , it is desired to find a seed set $\mathcal{S} \subseteq \mathcal{V}'$ (with $|\mathcal{S}| \leq M$), so that in the *underlying* network G , the total number of activated nodes through seeding \mathcal{S} is maximized.

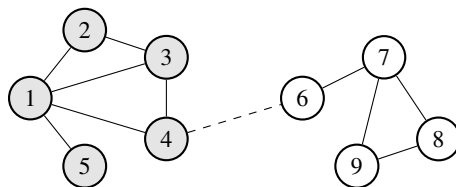


Fig. 1. Illustrative example of an organization-partitioned network, which is a special case of partially observable networks. The light nodes are observable. Note that there is only a weak tie between the observable and unobservable parts of the networks, which makes the influence maximization, and thereby the seed selection, particularly challenging.

In this paper, we confine our attention to heuristic-based IM (described in the following section), and leave the investigation of inference-based IM in partially observable network (where the unseen part of the network is reconstructed using generative models) for our future work.

3.1 Heuristic-based Influence Maximization

Here the structure of the problem is such that reliable inference is not possible. This might arise, for instance, due to restricted availability of samples or limited computational capacity. In such circumstances, instead of inference, we explore simple rules for selecting seed nodes:

- **Random Selection:** In this case, we select the seeds simply at random.
- **Random with Neighbor Activation:** Here, we first select the seeds at random, and then for each seed, we activate one of its neighbors. This approach is based on the *friendship paradox* [25], first discovered by S. L. Feld. It states that on an average basis, most people have fewer friends than their friends have. Thus, if we select some seed at random, it is beneficial to activate one of its neighbors, instead of the original node itself, due to possible larger number of connections.

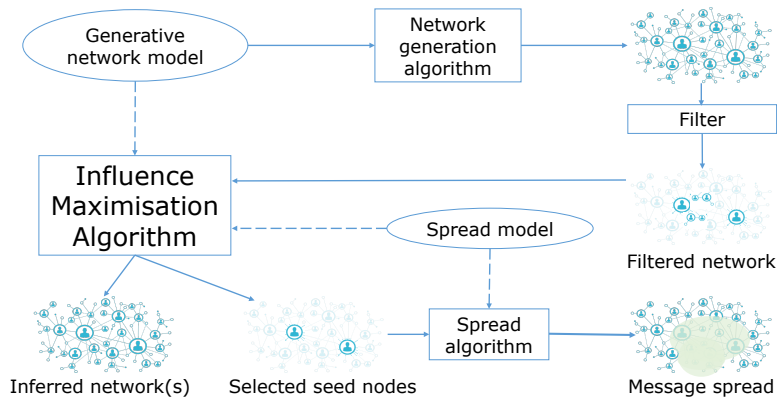


Fig. 2. Benchmarking framework

- **Selection based on (Weighted) Degree:** In this heuristic, we first rank the nodes in the known part of the network based on their degree, so that nodes with larger number of neighbors have higher ranks. Then we select node with highest rank as seeds.

Here we use the intuition that nodes with larger number of connections are very likely to be highly influential. In the settings we considered in this paper, neighbours typically had non-negligible influence on each other. However, in some real-world social networks, users may have large numbers of followers but actually exert little influence on them [4]. Here, links could be pruned based on historical influence interactions, but we leave this to future work.

In the weighted version of this approach, we still rank the nodes based on their degree; however, in order to improve the influence probability in the unknown part of the network, we attach a higher weight for neighbours that are in the boundary set. For example, when $w = 3$, then node 6 in Figure 1 would have a weighted degree of 4, while node 7 would only have a weighted degree of 3 (since all its neighbours are fully observable).

- **State of the Art with Limited Visibility:** Here we simply use the IM algorithms which are originally developed for known networks, for instance *TIM* and *IMM* described in Section 2.1. In doing so, we intend to investigate the extent to which the state of the art is applicable to partially known networks. Similar to the activation based on degree, *TIM* and *IMM* can be used in the weighted version. In essence, we run the *TIM* and *IMM* in the known part of the network as conventional; however, for selecting the RR sets, we weight the edges with connections to outside the boundary, although such connections are not completely visible to us.

4 Experimental Evaluation

In order to evaluate influence maximization algorithms in partially observable networks, we implemented a comprehensive benchmarking framework, as shown in Figure 2 (here, dashed lines indicate the use of a probabilistic model, and solid lines indicate data flow). Briefly, the figure shows how a full network is generated, then filtered to a partially observable network. The IM algorithm then selects seed nodes (potentially using inference) and the outcome is simulated. We implemented a range of algorithms for generating synthetic networks and filtering these. However, due to the limited space, we only report on a subset of our results. In the following, we first describe our experimental setup and then outline the results.

4.1 Experimental Setup

We use the NetHept⁵ dataset, a network of 15k nodes and 31k edges (representing citations within the high energy physics theory community). We choose this because it constitutes a real-world dataset of reasonable size and because it has been widely used to benchmark influence maximization algorithms [18].

Throughout our experiments, we will generate partially observable networks with varying numbers of *fully observable* nodes (i.e., nodes that are observable and whose neighbours are also observable — see Section 3). We do this by initialising the set of fully observable nodes (O) to contain a number of seed nodes, chosen uniformly at random. In our experiment, we initially choose 4 seed nodes, but our results are similar for other settings. Then, we iteratively add one additional node to O at a time, either by picking a node from O and then adding one of its neighbours that is currently not in O to O (both uniformly at random), or, when no such neighbours exist, by adding a random node to O . This continues until O contains a target number of nodes.

Again, due to space reasons, we focus on the weighted cascade model (see Section 3), although we have observed broadly similar trends more generally in the independent cascade model (for various edge weight distributions).

We evaluate most of the heuristics described in Section 3.1. Here, we use **Random** and **Random Neighbour** to denote Random Selection (without or with Neighbour Activation). **Weighted(1)** represents selection based on degree, while **Weighted(w)** is the weighted version that attaches a relative weight w to boundary nodes. Finally, **IMM(1)** is the state-of-the-art IMM algorithm described in Section 2, and **IMM(w)** similarly represents the weighted version. We choose IMM here, due to its consistently high performance compared to other state-of-the-art algorithms [1]. We leave the evaluation of centrality-based heuristics and other state-of-the-art approaches in partially observable settings to future work.

To evaluate each influence maximization algorithm, we generate a partially observable network, run the algorithm on it and then simulate the influence spread on the chosen set of seeds 1,000 times to obtain an average spread. We repeat this process 1,000 times (to evaluate a wide range of partially observable networks) and present

⁵ This can be downloaded at <https://www.microsoft.com/en-us/research/wp-content/uploads/2016/02/weic-graphdata.zip>.

the overall average spread. The 95% confidence intervals are typically smaller than an absolute deviation of 2 from the average value, so are not plotted on the graphs.

4.2 Results

Figures 3 and 4 show the average spread of all algorithms in a setting with five seed nodes, as we vary the network visibility, i.e., the proportion of fully observable nodes ($|O| / |\mathcal{V}|$). Figure 3 focuses on networks with particularly low observability (up to 10%), while Figure 4 shows the full range of proportions.

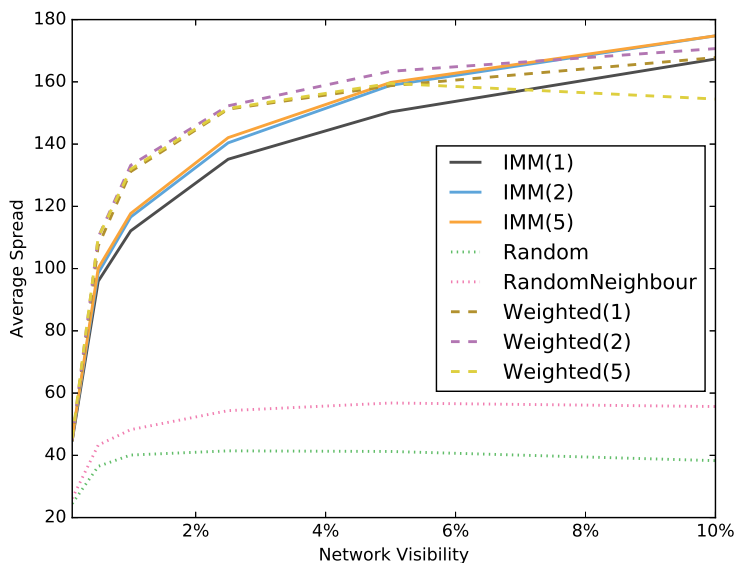


Fig. 3. Average spread in partially observable networks for 5 seeds (up to 10% visibility).

Here, several interesting trends emerge. As expected, the state-of-the-art IMM algorithm performs well throughout all settings. However, looking more closely at cases with low observability (Figure 3), some of the heuristic approaches outperform it. Specifically, the degree-based heuristics perform consistently well, sometimes achieving an up to 19% higher average spread than IMM. For example, when 1% of nodes are fully observable, IMM achieves an average spread of 112.12 ± 1.78 (with 95% confidence interval), while Weighted(2) achieves 133.19 ± 1.95 .⁶ This is likely because IMM focuses on maximising its influence only within the known part of the network. In settings with low visibility, this is a highly inaccurate view and the degree-based heuristics offer a better estimate of what nodes are influential across the full network.

⁶ A t-test confirms there is a statistically significant difference at the $p < 0.0001$ level.

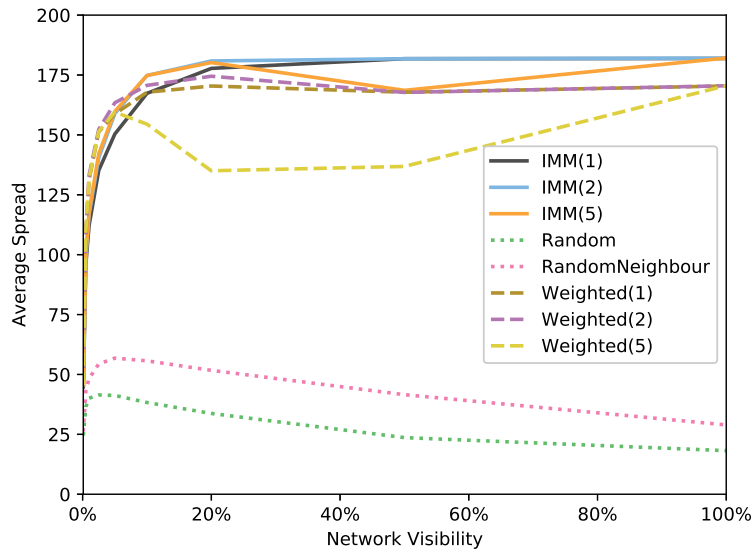


Fig. 4. Average spread in partially observable networks for 5 seeds (up to 100% visibility).

As visibility rises (also continuing in Figure 4), this difference becomes less pronounced, and by 10% visibility, they achieve a similar performance. As visibility rises further, IMM(1) eventually achieves the highest performance (from 50% visibility onwards). This is not surprising, since it is known to perform well in fully observable networks.

Looking specifically at the effect of adding higher weights for boundary nodes to the heuristics, this can lead to a significant increase in the average spread. However, the performance is sensitive to the exact parameter value, and for networks with higher visibility, a high weight can indeed lead to a decrease in performance, and is most pronounced for Weighted(5) in settings with 20–50% visibility. This is because the boundary nodes actually decrease in importance in the network as more of it is known to the algorithm.

Note that the relatively good performance of degree heuristics is surprising, as it goes against the distinction between influential and high-degree nodes that has been observed in some settings [4]. This may be due to the fact that NetHept, as a citation network, has high-degree assortativity [2, Chapter 7], meaning that higher degree nodes are likely to be connected to other high-degree nodes and vice versa. This would mean that targeting a high-degree node can lead to influence spread via its high-degree neighbors (including within the unseen part of the network).

Promisingly, adding the weight to IMM significantly increases the average spread of IMM — up to 6% in some cases, and IMM(2) consistently performs better or as well as IMM across all settings. However, as for the degree-based heuristic, choosing a very

large weight can lead to a decrease in performance for the same reasons as mentioned above.

Both random heuristics achieve a low level of performance, but using neighbour activation achieves a significantly higher spread than not using it (up to 100% more in some cases). This is potentially promising in settings where very little is known about the network, but where neighbours of known nodes can be activated. Note the apparent decrease in performance for larger networks is due to the way we filter nodes based on their neighbours — more connected nodes are likely to be included in our fully observable set earlier. Thus, when sampling a random node in a network with smaller visibility, it is more likely to be well connected, than when sampling from a network with full visibility.

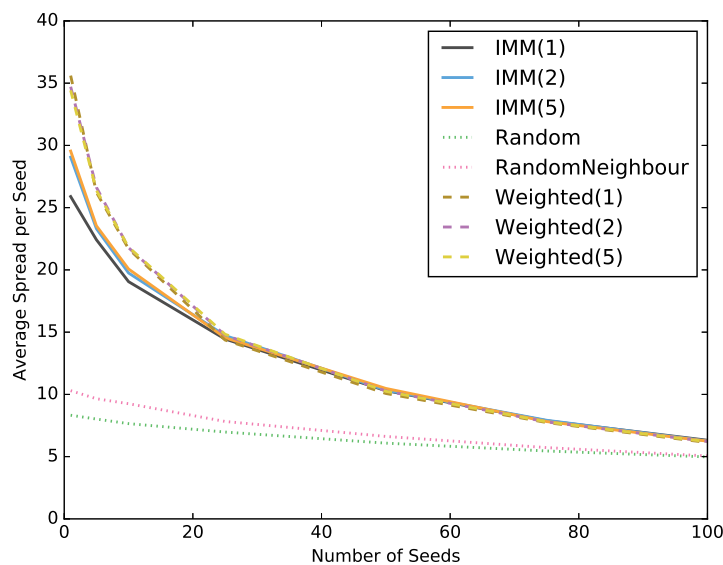


Fig. 5. Average spread for varying numbers of seeds for 1% visibility.

Finally, Figure 5 shows the average spread per initial seed chosen as the number of initial seeds is increased (in a setting with 1% visibility). This highlights that our heuristic techniques achieve the highest performance gains over the state of the art in settings with fewer initial seeds (a gain of up to 38% when there is just a single initial seed).

Overall, these are promising results, showing that in settings where large parts of the network are not observable and where only few seeds can be chosen, the state-of-the-art algorithm does not necessarily perform best. Instead simple heuristics perform well, and both those heuristics and the current state of the art benefit from explicitly favouring nodes at the boundary of the known network. It should also be noted that the

heuristics are several order of magnitude faster than IMM — a typical run of IMM took about 0.1–0.2 seconds, while the degree-based heuristics typically completed within 0.2–0.3ms.

5 Conclusions and Future Work

In this paper, we presented initial results on influence maximization in partially observable networks. We showed that the current state of the art does not necessarily produce the best results, especially when only small parts of the network are visible. Indeed, we examined some settings where a degree-based heuristic leads to a 38% improvement over the state of the art.

In future work, we will examine these settings in more detail, to derive more principled inference-based algorithms that work well in a range of settings and that offer some performance guarantees.

We will look at influence maximization over different types of real-world networks to understand whether the relative performance of our simple heuristic depends on any parameters of the underlying network (e.g., degree assortativity).

We will then consider the case where the decision maker has access to a generative network model that can be used for inference about the structure of the unseen part of the network (conditional on the observed part) before application of IM algorithms. Two specific examples of such generative models include:

- **Sampling:** An inference model is used to sample multiple snapshots of the unseen network. Across all snapshots, the average marginal influence of all nodes in \mathcal{V}' is calculated and a greedy IM approach is applied.
- **Maximum Likelihood:** The model is used to calculate a single maximum likelihood network and an existing IM approach is applied to this.

Finally, we will investigate repeated settings, as described in Section 2.2, where seeds are selected incrementally and where information about the unknown part of the network is gradually revealed.

Acknowledgments

This research was sponsored by the U.S. Army Research Laboratory and the U.K. Ministry of Defence under Agreement Number W911NF-16-3-0001. The views and conclusions contained in this document are those of the authors and should not be interpreted as representing the official policies, either expressed or implied, of the U.S. Army Research Laboratory, the U.S. Government, the U.K. Ministry of Defence or the U.K. Government. The U.S. and U.K. Governments are authorized to reproduce and distribute reprints for Government purposes notwithstanding any copy-right notation hereon. The work of S. M. was supported by a post-doctoral fellowship from the German Research Foundation (DFG) under Grant Number MA 7111/1-1.

References

1. Arora, A., Galhotra, S., Ranu, S.: Debunking the myths of influence maximization: An in-depth benchmarking study. In: Proceedings of the 2017 ACM International Conference on Management of Data. pp. 651–666. ACM (2017)
2. Barabási, A.L.: Network science. Cambridge University Press (2016)
3. Borgs, C., Brautbar, M., Chayes, J., Lucier, B.: Maximizing social influence in nearly optimal time. In: Proceedings of the Twenty-Fifth Annual ACM-SIAM Symposium on Discrete Algorithms. pp. 946–957. SIAM (2014)
4. Cha, M., Haddadi, H., Benevenuto, F., Gummadi, P.K.: Measuring user influence in twitter: The million follower fallacy. (2010)
5. Chen, W., Lin, T., Tan, Z., Zhao, M., Zhou, X.: Robust influence maximization. In: Proceedings of the 22nd ACM SIGKDD International Conference on Knowledge Discovery and Data Mining. pp. 795–804. KDD '16, ACM, New York, NY, USA (2016)
6. Chen, W., Yuan, Y., Zhang, L.: Scalable influence maximization in social networks under the linear threshold model. In: Data Mining (ICDM), 2010 IEEE 10th International Conference on. pp. 88–97. IEEE (2010)
7. Domingos, P., Richardson, M.: Mining the network value of customers. In: Proceedings of the seventh ACM SIGKDD international conference on Knowledge discovery and data mining. pp. 57–66 (2001)
8. Goldenberg, J., Libai, B., Muller, E.: Talk of the network: A complex systems look at the underlying process of word-of-mouth. *Marketing letters* 12(3), 211–223 (2001)
9. K. Satio, R.N., Kimora, M.: Prediction of information diffusion probabilities for independent cascade model. In: Knowledge-Based Intelligent Information and Engineering Systems. pp. 67–75. Springer (2008)
10. Kempe, D., Kleinberg, J.M., Tardos, E.: Maximizing the spread of influence through a social network. In: KDD. pp. 137–146 (2003)
11. Lei, S., Maniu, S., Mo, L., Cheng, R., Senellart, P.: Online influence maximization. In: Proceedings of the 21th ACM SIGKDD International Conference on Knowledge Discovery and Data Mining. pp. 645–654. KDD '15, ACM (2015)
12. M. Richardson, P.D.: Mining knowledge-sharing sites for viral marketing. In: Proceedings of the Eighth Intl. Conf. on Knowledge Discovery and Data Mining. pp. 61–70 (2002)
13. Mihara, S., Tsugawa, S., Ohsaki, H.: Influence maximization problem for unknown social networks. In: Proceedings of the 2015 IEEE/ACM International Conference on Advances in Social Networks Analysis and Mining. pp. 1539–1546. ASONAM '15, ACM (2015)
14. Mossel, E., Roch, S.: On the submodularity of influence in social networks. In: Proceedings of the thirty-ninth annual ACM symposium on Theory of computing. pp. 128–134. ACM (2007)
15. Okolloh, O.: Ushahidi, or 'testimony': Web 2.0 tools for crowdsourcing crisis information. *Participatory learning and action* 59(1), 65–70 (2009)
16. S. Vaswani, L.L., Schmidt, M.: Influence maximization with bandits. In: Proceedings of the 2015 International Conference on Neural Information Processing Systems. pp. 1–6 (2015)
17. Tang, S.: No time to observe: Adaptive influence maximization with partial feedback. *CoRR abs/1609.00427* (2016)
18. Tang, Y., Shi, Y., Xiao, X.: Influence maximization in near-linear time: A martingale approach. In: Proceedings of the 2015 ACM SIGMOD International Conference on Management of Data. pp. 1539–1554. ACM (2015)
19. Tang, Y., Xiao, X., Shi, Y.: Influence maximization: Near-optimal time complexity meets practical efficiency. In: Proceedings of the 2014 ACM SIGMOD international conference on Management of data. pp. 75–86. ACM (2014)

20. Von Ahn, L.: Human computation. In: Data Engineering, 2008. ICDE 2008. IEEE 24th International Conference on. pp. 1–2. IEEE (2008)
21. Wang, C., Chen, W., Wang, Y.: Scalable influence maximization for independent cascade model in large-scale social networks. *Data Mining and Knowledge Discovery* 25(3), 545 (2012)
22. Wen, Z., Kveton, B., Valko, M.: Influence maximization with semi-bandit feedback. *CoRR* abs/1605.06593 (2016)
23. Zenonos, A., Stein, S., Jennings, N.R.: Coordinating measurements for air pollution monitoring in participatory sensing settings. In: *Proceedings of the 2015 International Conference on Autonomous Agents and Multiagent Systems*. pp. 493–501 (2015)
24. Zhuang, H., Sun, Y., Tang, J., Zhang, J., Sun, X.: Influence maximization in dynamic social networks. In: *Data Mining (ICDM), 2013 IEEE 13th International Conference on*. pp. 1313–1318. IEEE (2013)
25. Zuckerman, E.W., Jost, J.T.: What makes you think you’re so popular? Self-evaluation maintenance and the subjective side of the “friendship paradox”. *Social Psychology Quarterly* pp. 207–223 (2001)