

Seeding Influential Nodes in Non-Submodular Models of Information Diffusion

(Extended Abstract)

Elliot Anshelevich
CS Department, RPI.
Troy, NY
eanshel@cs.rpi.edu

Ameya Hate
CS Department, RPI.
Troy, NY
hatea2@rpi.edu

Malik Magdon-Ismael
CS Department, RPI.
Troy, NY
magdon@cs.rpi.edu

ABSTRACT

We consider the model of information diffusion in social networks from [2] which incorporates trust (weighted links) between actors, and allows actors to actively participate in the spreading process, specifically through the ability to query friends for additional information. This model captures how social agents transmit and act upon information more realistically as compared to the simpler threshold and cascade models. However, it is more difficult to analyze, in particular with respect to seeding strategies. We present efficient, scalable algorithms for determining good seed sets – initial nodes to inject with the information. Our general approach is to reduce our model to a class of simpler models for which provably good sets can be constructed. By tuning this class of simpler models, we obtain a good seed set for the original more complex model. We call this the *projected greedy approach* because you ‘project’ your model onto a class of simpler models where a greedy seed set selection is near-optimal. We demonstrate the effectiveness of our seeding strategy on synthetic graphs as well as a realistic San Diego evacuation network constructed during the 2007 fires.

Categories and Subject Descriptors

F.2.2 [Algorithms]: Computations on discrete structure;
G.2.2 [Mathematics of Computing]: Graph algorithms;
I.6 [Computing Methodologies]: Simulation and modeling

Keywords

Diffusion, Multi-agent systems

1. INTRODUCTION

Networks (social, computer, and physical) are replete with the flow of information, ideas, innovations, etc., and it is these flows which affect the way people think, act, and bind together in a society. Ideally, important messages should disseminate quickly and reach the people who need to take action, and the diffusion of malicious gossip should, if possible, be terminated. Since diffusion of both useful and malicious

items is at the core of our society, it is vital to understand the mechanisms of diffusion through dynamic networks. A network can change as a result of the diffusion. For example, an evacuation message may not reach its intended audience because certain important people (critical conduits of information) left the community before the diffusion completed. In this paper, we study how to optimize a diffusion in realistic, large scale (multi-million node) complex networks; in particular, how to select those actors to be initially seeded with information so as to maximize the ultimate number of actors receiving the information *and* acting upon that information. Of particular interest to us is the diffusion of high-value actionable information – information which is asking the user to take some action – in particular diffusion of an evacuation warning. This will be the context of our discussion, although our methods are general.

Emergency warnings are an important and interesting context for diffusion. There is no universally trusted news source, and it is practically infeasible for such a news source to reach everyone. Thus, it is essential to make use of the social communication network to diffuse a warning through the network in such a way that people act. A person hearing the same information from multiple independent sources is more likely to act on it. The notion of trust, which measures the likelihood that a message from one person to another will be believed, plays an important role in such diffusions [1]. We will use the model developed in [2] to approximate the process of social information diffusion. In essence, this model captures:

- (i) the notion of trust and community structure;
- (ii) an actor’s ability to query friends for more information to seek confirmation and/or additional information, [3];
- (iii) multiple sources of information, each source trusted to a different level;
- (iv) network dynamics as a result of the diffusion (for example nodes evacuating the network);
- (v) tunable parameters that can model different types of diffusion, from gossip to evacuation warnings.

2. OUR CONTRIBUTION

We give efficient heuristics to select a subset of the actors (the seed set) to initialize with the information so that the final set of actors who believe and *act upon* that information is maximized. We call this a targeting algorithm.

Appears in: *Proceedings of the 12th International Conference on Autonomous Agents and Multiagent Systems (AA-MAS 2013)*, Ito, Jonker, Gini, and Shehory (eds.), May, 6–10, 2013, Saint Paul, Minnesota, USA.

Copyright © 2013, International Foundation for Autonomous Agents and Multiagent Systems (www.ifaamas.org). All rights reserved.

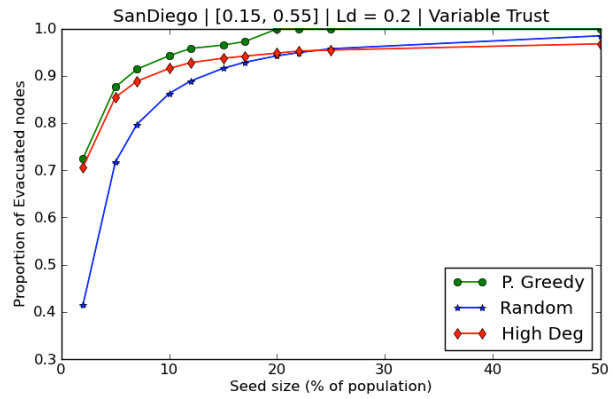
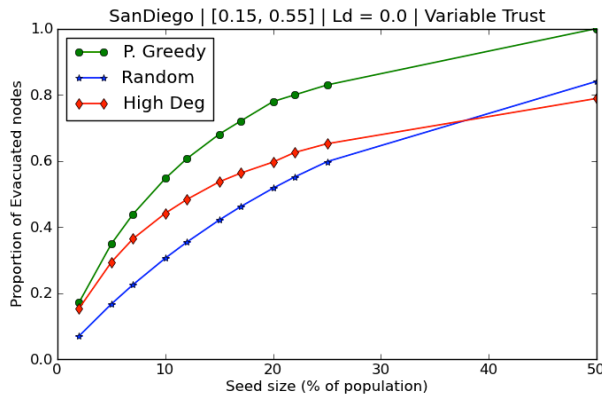


Figure 1: San Diego network, with group variable trust (high trust edges within a group).

We introduce a new paradigm for developing such heuristics in realistic diffusion models. Our approach works as follows. First we develop an appropriate simplification of the general model (with certain tunable knobs) for which we can develop *provably efficient and near-optimal targeting algorithms*. In contrast to the prior work, we develop the targeting algorithms for a *different, more efficient model*, not the original general model. The key aspect of our approach is that the simpler model has tunable parameters; these tunable parameters can be adjusted so as to optimize the ultimate diffusion for the true general model.

Projected Greedy Heuristic. For the particular diffusion model used in this paper, the heuristic we developed works as follows. Given an instance of the general model, we first form an instance of a simpler, submodular model with trust values chosen so that this new instance closely approximates the instance of the general model. Though the original general instance typically results in a non-submodular optimization problem, this new model is always submodular. This means we can find a greedy solution quickly, which is guaranteed to closely approximate the optimal seed set for the submodular model. If this submodular model is a good approximation to the original instance, then this seed set will be good for the original model. We can test the performance of the seed set on the general model explicitly (by forward simulation).

The particular simpler submodular model we use depends on the parameters of the model (trust values). By changing these parameters, we obtain different seed sets which are optimal for different submodular models, all of which approximate the general model to different extents. We test each seed set’s performance on the general instance and pick the best among these seed sets as our final output seed set for the instance of the general non-submodular model. Details can be found in the full version of the paper.

Experimental Testing. We demonstrate the effectiveness of our approach on several random graph models, as well as a model of the San Diego network during the 2007 fires that was created in [2]. In all cases, and for various types of diffusions (settings of parameters within the diffusion model) our approach gives the best seeds when compared with two benchmark algorithms: random targeting and high-degree targeting.

Figure 1 shows results of simulating the diffusion process

using seed sets selected by these targeting techniques on graph instances modeled after the San Diego network. The X-axis shows the percentage of nodes in a graph that were selected for seeding while the Y-axis shows the proportion of nodes eventually evacuated due to successful diffusion of warning messages. The two graphs pertain to different modes of information diffusion between nodes. Specifically, in the figure on the left nodes consider only the maximum level of information that their neighbors report, while in the figure on the right they also consider the sum of the information reported by their neighbors. For both scenarios shown in the figure, the Projected Greedy heuristic performs significantly better than both Random and High Degree seeding strategies.

There is a tradeoff, however. Random targeting, though not very effective, is simple and requires little information regarding the network. High degree targeting requires more information, namely a measure of prominence of the nodes as well, and often, though not always, does better than random seeding. Our algorithms require the most information, namely some model for the connectivity in the network (who links to whom). Though this may not be known in practice, it can be estimated as was done in [2] for the San Diego network (based on population densities and a geometric community based random graph model).

3. ACKNOWLEDGEMENTS

Anshelevich was partially supported by NSF grants CCF-0914782, CNS-1017932, and CCF-1101495 and Magdon-Ismael was partially supported by the Army Research Laboratory under Cooperative Agreement Number W911NF-09-2-0053.

4. REFERENCES

- [1] K. Chopra and W. A. Wallace. Modeling relationships among multiple graphical structures. *Comp. and Math. Org. Theory*, 6(4):361–380, 2000.
- [2] C. Hui, M. Goldberg, M. Magdon-Ismael, and W. Wallace. Simulating the diffusion of information: An agent-based modeling approach. *Int. J. Agent Tech. and Sys.*, 2(3):31–46, 2010.
- [3] J. Sorensen and D. Mileti. Decision-making uncertainties in emergency warning system organizations. *Int. J. of Mass Emerg. and Disas.*, 5(1):33–61, 1987.