

Finding Influential Nodes for Initiating Successful Campaigns in Social Networks

Mohammadreza Samadi¹, Alexander Nikolaev¹, Rakesh Nagi²

¹ Department of Industrial and Systems Engineering, University at Buffalo (SUNY), Buffalo, NY 14260

² Department of Industrial and Enterprise Systems Engineering, The University of Illinois at Urbana-Champaign, IL 61801
(msamadi@buffalo.edu, anikolaev@buffalo.edu, nag@illinois.edu)

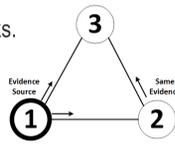
Motivation and Problem Statement

- Online social networks are growing fast.
- Knowledge transfer between individuals affects their purchasing/voting decisions [2].
- Calculated initial influence may lead to widespread campaign success.



Word-of-Mouth and Viral Marketing strategies direct the social effects in the campaign of choice.

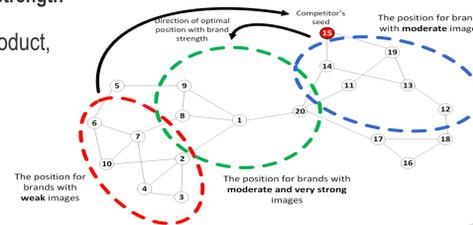
- The notion of subjective evidence explains belief reinforcement in social networks.
- The Influence Maximization problem finds a set of influential nodes (sponsored users) to start a social influence campaign [3].
- Influence maximization considers both the number of activated people and the time of activation.



Insights

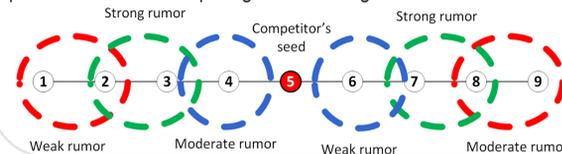
Case Study 1: The optimal strategic positions of the seeds are governed their relative strength

- Product distributors in a regional market, not yet exposed to a new emerging product, opt to carry the new product and influence each other through carrying it.
- Manufacturer F1 plans to offer the product at a discounted price to two local distributors to motivate other distributors.
- Node 15 is an agent for F2, that produces an alternative product.
- The relative strength of brands determines the strategic position of the seeds.



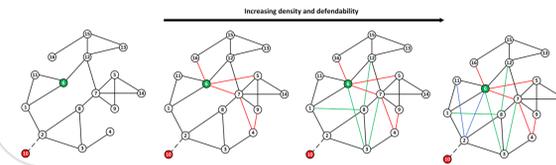
Case Study 2: Strategic positioning in an information war problem on a simple tree network

- A rumor initiated by the opponent (negative party) at node 5 and he "positive" party, plans to initiate a competing rumor through two nodes.



Case Study 3: The ability to penetrate a cluster depends on its cohesion and attack timing

- A social community has been exposed to a single political opinion for d time periods. After that a competing party creates a connection to the cluster to penetrate into it.
- Defendability increases with delay and cluster density.

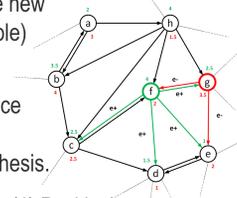


Challenges

- How can the diffusion of influence be realistically modeled in social networks?
- Can we use mathematical programming to model the spread of social influence?
- How can the mathematical model for Influence Maximization problem be solved for large-scale social networks?
- What managerial insights can be derived by solving different instances of the Influence maximization problem?

Parallel Cascade (PC) Diffusion Model

- The decision making is modeled as a hypothesis testing process based on the Bayesian Inference logic.
- Null hypothesis reflects the opinion supported by decision maker (e.g., the hypothesis that the new iPhone released in the market is reliable)
- Each node is considered as an intelligent agent that collects all evidence in its social neighborhood and decides to either accept or reject the null hypothesis.
- At each time period, a node can be: (1) Positively activated, (2) Negatively activated, (3) Neutral [1].



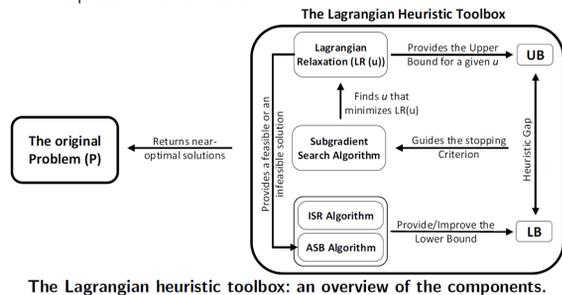
Optimization Model and Solution Methodology

- A mixed-Integer program is developed to maximize the number of influenced nodes in the shortest time possible.

$$(P) \quad \max Z = \sum_{i=1}^N \sum_{t=0}^T (X_{it} - Y_{it})$$

Summation over all nodes (X_{it} - Y_{it})
Summation over all time periods (i, t)

- The problem is NP-hard.



The Lagrangian heuristic toolbox: an overview of the components.

A Guaranteed-Performance Lagrangian Relaxation Heuristic

- Relaxes one of the constraint sets and attaches them to the objective function

$$(LR_u) \quad \max Z^{LR_u} = \sum_{i=1}^N \sum_{t=0}^T ((X_{it} - Y_{it}) + u(\sum_{i=1}^N X_{i0} - S^+))$$
- Solves the Lagrangian Dual problem to find the optimal Lagrangian multipliers

$$(LD_u) \quad Z^{LD_u}(u) = \text{Min}_u Z^{LR_u}(u)$$
- Uses Subgradient search algorithm for solving the Lagrangian Dual problem
- Two heuristics are developed for finding lower bound for the optimal solution and stopping the search procedure.

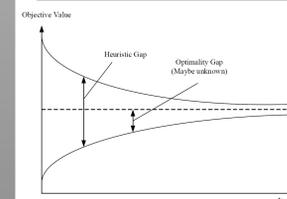
Heuristic Algorithms for the Lower Bound to (P)

Iterative Seed Removal (ISR) Algorithm

- Finding a dummy problem with more positive seeds
- The solution time for dummy problem is significantly lower than (P).
- Solution of the dummy problem is expected to include the original problem's solution.
- ISR iteratively removes the seeds in dummy solution.
- ISR provides a valid lower bound for (P).

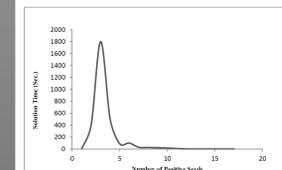
Adaptive Subgradient-Based (ASB) Algorithm

- ASB utilizes the information in subgradient algorithm.
- In each iteration, the Lagrangian Relaxation problem returns a solution with more positive seeds.
- ASB selects the first k_1 positive seeds from the relaxed problem solution.



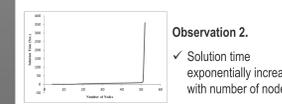
- The lower bound is obtained by two heuristic methods.
- Upper bound is obtained by Lagrangian Relaxation.
- The Lagrangian Relaxation heuristic guarantees the quality of the solution.

Optimization Model and Solution Methodology



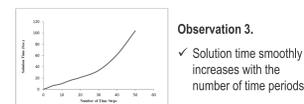
Observation 1.

- Solution time first increases with the number of positive seeds and then decreases.
- This diagram shows why the solution time for dummy problem is significantly lower than (P).
- Serves as the basic idea of ISR algorithm



Observation 2.

- Solution time exponentially increases with number of nodes



Observation 3.

- Solution time smoothly increases with the number of time periods.

Computational Results

- The algorithms are tested over Facebook datasets.

Dataset	Nodes	Edges	Directed	Density
F1	150	1693	No	0.151
F2	747	30025	No	0.108
F3	534	4813	No	0.034
F4	1034	26749	No	0.05



- All small and medium-sized problems can be solved to optimality using CPLEX.
- For the small problems, CPLEX outperforms the Lagrangian Relaxation heuristic in terms of solution time.
- When the problem size increases, the solution time of CPLEX increases rapidly but the Lagrangian Relaxation heuristic remains fast.

Dataset	Nodes	Periods	Pos. Seeds	LR Time (sec.)	LR LB	LR UB	Cplex Time (sec.)	Opt. Sol.	Opt. Gap (%)	Heu. Gap (%)	Iter. #
F1	30	40	6	11.53	1167	1186	0.60	1167	0	1.6	20
F1	40	100	9	72.01	4020	4021	7.89	4020	0	0.02	20
F1	60	50	7	81.32	2849	2931	74.32	2853	0.1	2.7	20
F2	45	64	7	26.59	2795	2870	7.71	2795	0	2.6	20
F2	60	50	9	32.96	2887	2929	45.84	2887	0	1.4	20
F2	85	75	14	49.66	6233	6289	3425.23	6233	0	0.8	30

- For large Facebook networks, CPLEX cannot even create a feasible solution in the computer memory.
- The Lagrangian Relaxation heuristic runs in a reasonable computational time and provides an acceptable heuristic gap.
- The runtime for the Lagrangian Relaxation heuristic smoothly increases with the dimensions of the problem instances.

Conclusion

- Our study presents a new diffusion model in social networks.
- This work sheds light on the phenomena of belief reinforcement and viral spread of innovations.
- We propose a mathematical model for solving Influence Maximization problem.
- We quantify how the social connection, campaign timing and opponent proximity impact the success of social campaigns.
- A managerial insight derived from this paper is that without creating a strong prior image of the product or opinion, spending money on triggering a campaign within a dense cluster, which has been exposed to an opposite opinion for a long time, is not profitable.

Future Studies

- We open a door to using location theory models for spread of influence investigations in social networks.
- Future studies can apply the proposed optimization scheme for modeling the spread of evidence in the growing social networks.
- Further research is required to employ network-level metrics, e.g., clustering coefficient, for reducing the size of large influence maximization problems to make them manageable.

References

1- Samadi, Mohammadreza, Alexander Nikolaev, Rakesh Nagi. 2014. Scalable methods for finding influential nodes in large social networks. Working paper, University at Buffalo SUNY, Buffalo, NY.
2- Aral, Sinan. 2011. Commentary-identifying social influence: A comment on opinion leadership and social contagion in new product diffusion. Marketing Science 30(2) 217-223.
3- Kempe, David, Jon Kleinberg, Eva Tardos. 2003. Maximizing the spread of influence through a social network. Proceedings of the ninth ACM SIGKDD international conference on Knowledge discovery and data mining. ACM, 137-146.