# How Effectively Can We Form Opinions?*

AmirMahdi Ahmadinejad
Sharif University of Tech.

Sina Dehghani
University of Maryland

MohammadTaghi Hajiaghayi
University of Maryland

Hamid Mahini
University of Maryland

Saeed Seddighin
University of Maryland

Sadra Yazdanbod
Georgia Institute of Tech.

## ABSTRACT

People make decisions and express their opinions according to their communities. An appropriate idea for controlling the diffusion of an opinion is to find influential people, and employ them to spread the desired opinion. We investigate an influencing problem when individuals' opinions are affected by their friends due to the model of Friedkin and Johnsen [4]. Our goal is to design efficient algorithms for finding opinion leaders such that changing their opinions has great impact on the overall opinion of the society.

We define a set of problems like maximizing the sum of individual opinions or maximizing the number of individuals whose opinions are above a threshold. We discuss the complexity of the defined problems and design optimum algorithms for the non NP-hard variants of the problems. Furthermore, we run simulations on real-world social network data and show our proposed algorithm outperforms the classical algorithms such as degree-based, closeness-based, and pagerank-based algorithms.

## Categories and Subject Descriptors

F.2.2 [**Analysis of Algorithms and Problem Complexity**]: Nonnumerical Algorithms and Problems

## Keywords

social networks; opinion formation; influence maximization

## 1. INTRODUCTION

Opinions of individuals are often influenced by their friends. In many situations social pressures and influences form the opinion of an individual. For example, if the majority of an individual's friends have the same political view, she would have tendency toward choosing the same view; if an individual finds a certain number of her friends with a particular life-style, she would be interested to live based on that life-style as well; and if an appropriate fraction of an individual's colleagues are motivated to publish their work in a specific venue, it would be likely to influence her opinion on where to publish.

A well-known framework for modeling opinion formation dynamics is proposed by DeGroot [3], in which each person has an opinion and updates her opinion based on a weighted averaging procedure. This model is suitable for modeling a situation where the dynamics converges to *consensus* and all individuals reach the same opinion in a stable state. Since in many real-world applications there is no consensus and the opinions are usually fragmented into several parts, we use a variant of the Degroot model, which is introduced by Friedkin and Johnsen [4], in which consensus is not necessarily reached. This model recently has become popular in the computer science literature (see e.g., [1, 2, 5]). In this model, each node has an *inherent internal opinion* $s_i$ which remains unchanged during the process and an *expressed overall opinion* $z_i$ which is dynamically updated through weighted averaging of the node's internal opinion and its neighbors' expressed opinions. More precisely, during each time step node $i$ updates its *expressed opinion* to be:

$$z_i = \frac{s_i + \sum_j w_{j,i} z_j}{1 + \sum_j w_{j,i}} \quad (1)$$

The internal opinion $s_i$ is modeling inherent beliefs of a person, which may be related to her political orientation, religious thinking, background and/or education. However, expressed opinion $z_i$ which is the external appearance of a person's opinion and is perceivable by the others in a network is subject to change and can be adapted dynamically.

## 2. INFLUENCING STRATEGY

A key question, which is highlighted by growth of social activities and huge amount of network externalities, is how a planner would design a strategy to control the diffusion of opinions. The idea of designing an appropriate strategy usually takes place around finding and influencing the authoritative people, called *opinion leader*. Here, we focus on answering this question: If we could change the internal opinions, which set should we select as the target set to maximize the popularity of a desired opinion at the equilibrium?

We are given graph $G = (V, E)$ with $n$ nodes and vector $S = \langle s_1, s_2, \ldots, s_n \rangle$ of internal opinions. Each internal opinion is in $[0, 1]$ which can be interpreted as a selection between two choices or as an amount of agreement with an idea.It can

*This work was supported in part by NSF CAREER award 1053605, NSF grant CCF-1161626, ONR YIP award N000141110662, DARPA/AFOSR grant FA9550-12-1-0423.


*WWW'14 Companion,* April 7–11, 2014, Seoul, Korea.
ACM 978-1-4503-2745-9/14/04.
http://dx.doi.org/10.1145/2567948.2577201 .

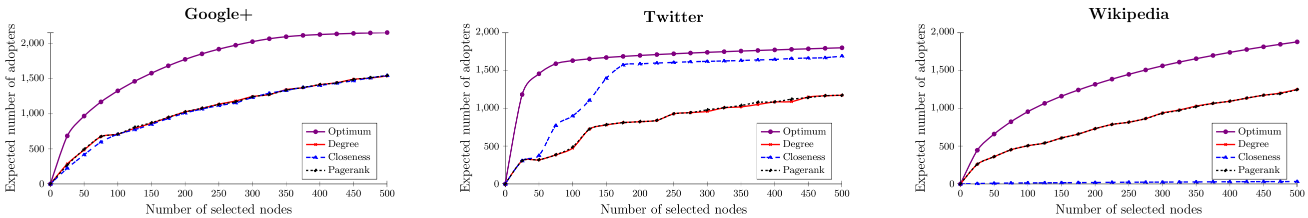


Figure 1: Performance of different algorithms on three real-world social networks Google+ with 2204 vertices and 411585 edges, Twitter with 2024 vertices and 111641 edges, and Wikipedia with 2215 vertices and 70002 edges. The data is available at http://snap.stanford.edu. Since the input social networks are unweighted, we set all weights equal to 1 in Equation (1). The $x$ axis is the number of selected nodes and the $y$ axis is the expected number of adopters.

be proved that if all nodes iteratively update their expressed opinions based on Equation (1), the expressed opinions will converge to a unique equilibrium. An *influencing strategy* is defined by vector $D = \langle d_1, d_2, \ldots, d_n \rangle$ which represents the changes in internal opinions, i.e., $s'_i = s_i + d_i$ denotes the new internal opinion of node $i$. Now we are ready to define our set of problems, each of them can be defined with a *fractional cost* function or an *integral cost* function. A fractional cost function is defined as $C(D) = \sum_i d_i$, and an integral cost function $C(D)$ is the number of nonzero elements of $D$.

Overall Influencing: In this problem, we maximize the sum of all expressed opinions within a given budget.

Targeted Influencing: In the Targeted problem, we want the expressed opinions of a set of target nodes become greater than threshold $t$ such that $C(D)$ is minimized.

Budgeted Influencing: In the Budgeted problem, we want to maximize the number of nodes whose final opinions are greater than threshold $t$.

Budgeted Stochastic Influencing: We further consider a stochastic version of the Budgeted problem, in which we do not know the threshold of each node and assume that these thresholds are drawn independently from given probability distributions. In the stochastic model, we want to find an algorithm which maximizes the expected number of individuals whose opinions are greater than their thresholds.

## 3. RESULTS

Trying to model the diffusion of opinions, one can write final expressed opinions $Z^*$ as $S \times A^*$, where $S$ is the vector of initial opinions and $A^*$ is a matrix and only depends on the structure of the graph. Gionis et al. [5] showed matrix $A^*$ has a nice property for any undirected graph, i.e., the sum of elements of each row is 1. This property can be used for designing efficient algorithm for our problems. The story is different for directed graphs which is the main concern of this study. We propose optimal algorithms and prove inapproximability results for aforementioned problems. In particular, we first propose a greedy algorithm for the Overall problem and the Budgeted Stochastic problem. We also present a polynomial-time algorithm for the Targeted problem with a fractional cost function. Furthermore, we show that the Targeted problem with an integral cost function is hard to approximate within a factor of $o(\log n)$ unless $P = NP$. At last, we show that the Budgeted problem is as hard as the notorious D$k$S problem. In D$k$S, given a graph $G$ and an integer $k$, we are interested in finding the densest subgraph of size $k$ in $G$.

| | **Fractional** | **Integral** |
|---|---|---|
| Overall | Polynomial | Polynomial |
| Targeted | Linear Program | Set Cover-Hard |
| Budgeted | D$k$S-Hard | D$k$S-Hard |
| Stochastic | Polynomial | Polynomial |

Table 1: Summary of our results.

## 4. EXPERIMENTS

The problem of determining influential nodes in a social network is a well-known problem in the literature and there are several algorithms for identifying such nodes. Here, we compare our algorithm with a set of popular heuristic algorithms such as the degree-based algorithm, the closeness-based algorithm, and the pagerank-based algorithm. Interestingly, we show the optimum set of influential nodes differs meaningfully from the results of these algorithms. In particular, we consider the Budgeted Stochastic problem and assume all thresholds are uniformly distributed in $[0, 1]$, and all internal opinions are 0. Our aim is to select a set of exactly $b$ nodes and set their internal opinions to 1, such that the expected value of activated nodes is maximized.

Figure 1 represents the results of different algorithms on real-world social networks such as Google+, Twitter, and Wikipedia. We also compute the pointwise performance ratio which is the ratio of expected number of activated nodes for the optimum algorithm and the best classical algorithm. We get mean pointwise performance ratio of 1.69, 1.47, and 1.72 for Google+, Twitter, and Wikipedia respectively. The experimental results show a significant gap between the optimum solution and the previous heuristic algorithms for determining influential nodes. This highlights the importance of our approach for finding opinion leaders during the opinion formation dynamics.

## 5. REFERENCES


[1] K. Bhawalkar, S. Gollapudi, and K. Munagala. Coevolutionary opinion formation games. In *STOC*, 2013.

[2] D. Bindel, J. Kleinberg, and S. Oren. How bad is forming your own opinion? In *FOCS*, 2011.

[3] M. H. DeGroot. Reaching a consensus. *J. Am. Statist. Assoc.*, 69(345):118–121, 1974.

[4] N. E. Friedkin and E. C. Johnsen. Social influence and opinions. *J. Math. Sociol.*, 15(3-4):193–206, 1990.

[5] A. Gionis, E. Terzi, and P. Tsaparas. Opinion maximization in social networks. In *SDM*, 2013.