2023 17th International Conference on Ubiquitous Information Management and Communication (IMCOM) | 978-1-6654-5348-6/23/$31.00 ©2023 IEEE | DOI: 10.1109/IMCOM56909.2023.10035644

Multi-Objective Information Maximization in a Social Network

Kundan Kandhway
Department of Electrical Engineering and Computer Science
Indian Institute of Science Education and Research Bhopal
Madhya Pradesh, 462066, India
kundankandhway@iiserb.ac.in

***Abstract*—We address the problem of running a campaign on a social network. We formulate a bi-objective optimal control problem where the first objective is to minimize the fraction of uninformed population, and the second objective is to minimize the cost of running the campaign. Information diffusion is modeled using susceptible-infected (SI) process. Specifically, the degree based compartmental model for the SI process is employed. We have modified the standard model to include effects of campaigning. Our formulation is capable of handling heterogeneous social networks with arbitrary degree distribution. We have used the multi-objective genetic algorithm to find the Pareto front for the formulated bi-objective optimal control problem. This work finds application in viral marketing of products, running public awareness campaigns, political campaigns etc.**

***Index Terms*—Epidemics, influence maximization, information dissemination, multi-objective programming, optimal control, susceptible-infected process**

I. Introduction

Apart from the face-to-face social interactions; traditionally, mass/print media (radio, television, newspapers etc.) were the source of news and information for people. Since the beginning of this century social networks and media have taken over as a major channel for spread of information, news, innovation, fashion trends etc. The volume of social interactions has increased many fold because physical proximity of individuals is no longer a requirement. A piece of information often influences our decision. For example, it is not uncommon to have a discussion about an upcoming election, or a service/product, before making a voting or purchase decision. This is exploited by social influencers from political parties and companies, often turning cyberspace into a battleground for these entities. Thus, influence maximization has drawn attention of researchers from multiple fields such as applied mathematics, physics, computer science, sociology, psychology, etc. and is also the focus of current paper.

Researchers have modeled information and influence diffusion using a variety of methods. Some examples include independent cascade model [1], linear threshold model [2], rumor models [3], epidemic models [4] etc. These models have been employed in studying a variety of problems, for example, influence maximization [5], [6]; rumor containment [7], [8]; viral marketing [9]; etc. In this paper, we will use the epidemic process as a model for information dissemination. Our decision is not only guided by the fact that information and biological epidemics spread in a similar manner, but also due to the simplicity of the epidemic models. In particular, unlike linear threshold and independent cascade models, the epidemic models are deterministic in nature. They are computationally cheaper to use in optimization formulations where the information dissemination model needs to be solved multiple times to arrive at a solution.

In this work, the spread of information in a heterogeneous social network is modeled using (a modified) degree based compartmental model for the susceptible-infected (SI) epidemic process. We incorporate the effect of running the campaign on the standard SI process, thus modifying the standard model. To maximize the reach of information in the social network, we formulate a bi-objective optimal control problem. The aim is to calculate the optimum campaigning strategies while keeping the following objectives in mind: The first objective is to minimize the number of uninformed individuals in a social network by the time the campaign ends, and the second objective is to keep the campaigning costs to the minimum. As can be noticed, these are conflicting in nature—the number of uninformed individuals can only be reduced by running stronger campaigns which is going to increase financial costs.

Our formulation differentiates itself from the existing literature in two aspects. The literature on multi-objective influence maximization, although available, is limited (see, for example, [10]). Majority of the existing literature address the single objective case (for example [11], [12] and many others). In addition, we have obtained the optimum time varying campaigning strategies in a period $0 \leq t \leq T$, where T is the duration of campaign (multi-objective optimal control formulation [13]), instead of a single shot seed selection problems where decision is made only at $t = 0$. Although sizable literature exists for optimal control (mitigation) of biological epidemics (for example [14]), maximizing information epidemics have received less attention [4], specifically in multi-objective optimal control setting. We have solved the formulated multi-objective problem numerically using genetic algorithm [15]. This technique is effective in handling the large system of ODEs in our model, and generating the Pareto front as a solution to the formulated problem. Genetic algorithm has been used before in the single objective single shot influence maximization problem [16].

The rest of this paper is organized as follows: Section II presents the standard and modified model for the SI process employed for information diffusion in this work. This section also formulates the bi-objective optimal control problem. Section III presents the method to compute a solution for the bi-objective optimal control problem formulated in Section II. The numerical results are discussed in section IV. Finally, section V concludes the paper.

II. Information Dissemination Model and Problem Formulation

We have a social network (graph) to capture the social interactions among the population. Individuals of the population are vertices of this graph. Two individuals share an edge, or are neighbors of each other, if they interact with one another. We have to run a campaign on this social network in the period $[0, T]$. Thus, the campaign starts at $t = 0$ and finishes at $t = T$. We assume that the network is undirected and remains fixed during the campaign duration. Also, we assume the social network to be free from self-edges and multi-edges, such graphs are called simple graphs.

In an undirected simple graph, the number of neighbors of an individual is called the degree of the individual. We collect all vertices of the social graph with degree k into degree class k. The set of all degree classes present in the social graph is given by $\{K_l, K_{l+1}, ..., K_h\}$, where K_l and K_h are minimum and maximum degrees in the graph. The social network is characterized by the probability mass function of the degrees, p_k, which is also called the degree distribution.

The vertices of the graph are in either of the two states– uninformed (susceptible) or informed (infected). Initially (at $t = 0$), all but a tiny set of vertices in the graph are in the uninformed state. The set of vertices that are in the informed state at $t = 0$ starts the information spreading process. As time progresses, the informed vertices transmit the information at a rate eta whenever it encounters an uninformed vertex. This spreading dynamics is called the SI process in the literature [17]. This model is able to capture the spreading process in a heterogeneously mixed population (the one organized in a graph) as opposed to a homogeneously mixed population where interaction graph is absent and interactions are completely random.

The fraction of uninformed and informed vertices in the degree class k in the graph at time t, $0 \leq t \leq T$, is given by $u_k(t)$ and $i_k(t)$ respectively. At any time t, any vertex is either in uninformed or informed state, hence, $i_k(t) = 1 - u_k(t)$. Given the spreading rate η, the fraction of uninformed and informed vertices in degree class k in the SI process is given by [17]:

$$\dot{u}_k(t) = -\eta u_k(t) k \sum_{m=K_l}^{K_h} (r_m i_m(t)), \; k \in \{K_l, ..., K_h\}.$$

$$\dot{i}_k(t) = \eta u_k(t) k \sum_{m=K_l}^{K_h} (r_m i_m(t)), \; k \in \{K_l, ..., K_h\}.$$

In the literature, the above model is referred to as the degree based compartmental model. Here, dot represents the differential with respect to time, $r_k = k \times p_k / \sum_{m=K_l}^{K_h} (m p_m)$, is the neighbor degree distribution. Neighbor degree distribution is the probability of finding a vertex with degree k by following an edge of a randomly chosen vertex of the graph. The initial conditions for the system of ordinary differential equations defined above are:

$$u_k(0) = u_0, \; k \in \{K_l, ..., K_h\},$$
$$i_k(0) = 1 - u_0, \; k \in \{K_l, ..., K_h\},$$

for a constant $u_0 \in [0, 1]$.

As stated before, $u_k(t) + i_k(t) = 1$. Thus, any one state variable is enough to characterize the system of ODEs. This reduces the size of the system of ODEs to $K_h - K_l + 1$ from $2 \times (K_h - K_l + 1)$. The complete system may now be represented as:

$$\dot{u}_k(t) = -\eta u_k(t) k \sum_{m=K_l}^{K_h} (r_m (1 - u_m(t))), \quad k \in \{K_l, ..., K_h\}. \tag{1}$$

The system has the same initial condition as before:

$$u_k(0) = u_0, \; k \in \{K_l, ..., K_h\},$$

for a constant $u_0 \in [0, 1]$.

The SI process described in (1) is modified to include a control effort used to speed up information spread in the social graph. A campaigner may employ strategies—for example, placing advertisements in the social network timeline of vertices—to achieve this. Mathematically, it is akin to the campaigner applying a control signal $\lambda(t)$ at time t that transfers uninformed vertices to the informed state. This modifies the SI process as follows:

$$\dot{u}_k(t) = -\eta u_k(t) k \sum_{m=K_l}^{K_h} (r_m (1 - u_m(t))) + \lambda(t) u_k(t), \; k \in \{K_l, ..., K_h\}. \tag{2}$$

The initial condition for the above system of ODEs:

$$u_k(0) = u_0, \; k \in \{K_l, ..., K_h\},$$

for a constant $u_0 \in [0, 1]$.

For the controlled SI process in (2), we formulate a multi-objective optimal control problem with two objective functions. It is natural for a campaigner to try to increase the population of informed vertices by the time the campaign ends, that is at $t = T$. In other words, the campaigner would want

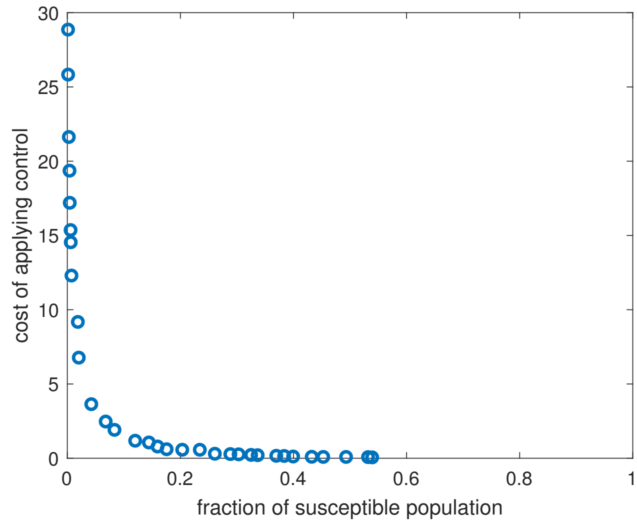

Fig. 1. Pareto front for the bi-objective problem formulated in this paper.

to minimize the fraction of uninformed vertices in the graph at $t = T$,

$$\min J_1 = \sum_{k=K_l}^{K_h} p_k u_k(T).$$

Apart from this, campaigning incurs costs, hence, the second objective for the campaigner is to minimize the aggregated cost of running the campaign. This is expressed mathematically as:

$$\min J_2 = \int_0^T \lambda^2(t)dt.$$

With the cost functions defined above and the controlled dynamical system defined in (2), the multi-objective optimal control problem is formulated below:

Problem (A)

$$\min: \ (J_1, J_2) = \Big(\sum_{k=K_l}^{K_h} p_k u_k(T), \quad \int_0^T \lambda^2(t)dt \Big)$$

$$\text{Subject to: } \dot{u}_k(t) = -\eta u_k(t) k \sum_{m=K_l}^{K_h} (r_m(1 - u_m(t))) + \lambda(t) u_k(t), \ k \in \{K_l, ..., K_h\},$$

$$u_k(0) = u_0, \ k \in \{K_l, ..., K_h\}.$$

III. Numerical Solution

The solution of a multi-objective optimal control problem in (A) is characterized by a Pareto front. A solution belongs to the Pareto front (or set) if no other candidate solution Pareto dominates it. A solution Pareto dominates another point if it strictly reduces (in our formulation) the value of any one objective, while not degrading the other objective value. In other words, a solution $\lambda_a(t)$ Pareto dominates $\lambda_b(t)$ if, either $J_1(\lambda_a(t)) < J_1(\lambda_b(t))$ and $J_2(\lambda_a(t)) \leq J_2(\lambda_b(t))$; or $J_1(\lambda_a(t)) \leq J_1(\lambda_b(t))$ and $J_2(\lambda_a(t)) < J_2(\lambda_b(t))$. A set of points that are not Pareto dominated by any other points in the feasible set is said to be in Pareto front (or set).

To compute the objective values in Problem (A) we discretize the continuous time system. The system of ODEs representing the state equations (2) is solved using Runge-Kutta method. The cost function J_1 is the final state of the system of ODEs, and J_2 is evaluated using the trapezoidal rule. The numerical solution is now obtained using the multi-objective genetic algorithm solver. Implementation of the multi-objective genetic algorithm is available in many languages such as Python and Matlab.

IV. Results

We present our results on a scale free network. A scale free network has a power law degree distribution, $p_k = Ck^{-\alpha}$, $k \in \{K_l, ..., K_h\}$, where α is the power low factor and C is an appropriate real constant that normalizes the degree distribution. We choose $\alpha = 3, K_l = 40$, and $K_h = 80$. The spreading rate for the SI process has been chosen to be $\eta = 0.04$, and the campaign deadline is normalized to $T = 1$ time unit. With these parameters we get the final fraction of the uninformed population at $t = T$ to be about 0.92. This is quite high, and we would like to reduce this fraction according to Problem (A), of course accounting for the costs of running the campaign as well.

For the set of parameters selected, the Pareto front for the bi-objective optimal control Problem (A) is presented in Figure (1). We can notice that the right extreme of the front has large fraction of uninformed (susceptible) population, and cost of application of control is nearly zero 0. Where as the other extreme has fraction of uninformed (susceptible) population close to 0 at a high cost of application of control. The evolution of the fraction of uninformed (susceptible) and informed (infected) individuals for one of the points on the Pareto front corresponding to fraction of uninformed population as 0.24

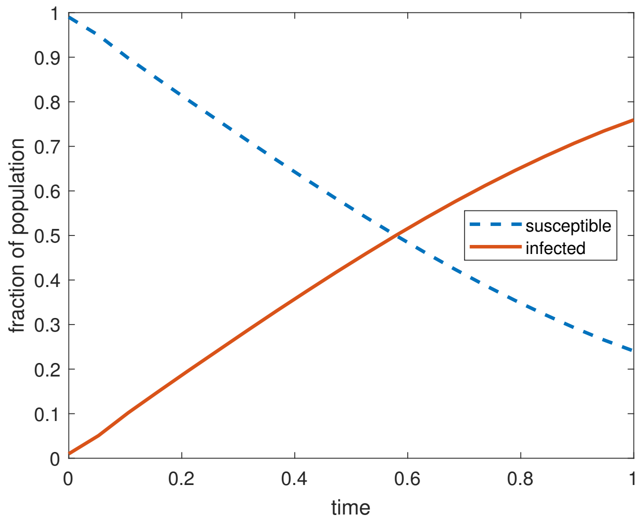

Fig. 2. Controlled SI information epidemic with optimally calculated control. More than 80% of the individuals are in the infected/informed state by the campaign deadline, much more than the uncontrolled case.

and cost of application of control as 0.31 is presented in Figure (2).

The scale free network chosen in this paper has $80-40+1 = 41$ degree classes, so is the number of state variables and number of ODEs in the dynamical system. This shows that the solution approach can easily handle large sized problems. We compute the fraction of uninfected population in the network as $\sum_{k=Kl}^{K_h} p_k u_k(t),\ t \in [0, T]$. This is nothing but a weighted linear combination of state variables, where weights are the fraction of nodes in a given degree class. Fraction of the informed population is simply $1-\sum_{k=Kl}^{K_h} p_k u_k(t),\ t \in [0, T]$.

V. Conclusion

In this paper we have formulated a bi-objective optimal control problem which aims to (i) minimize the fraction of uninformed population, and (ii) minimize the cost of running the campaign. Information diffusion is modeled using the degree based compartmental model for the susceptible-infected (SI) epidemic process. However, we have modified the standard model to include effects of campaigning. Our formulation is capable of handling heterogeneous social networks with arbitrary degree distribution. We have used the genetic algorithm to find the Pareto front for the formulated bi-objective problem. Our work finds application in viral marketing of products, running public awareness campaigns, political campaigns etc.

References

[1] Trivedi, N., & Singh, A. (2020). Efficient influence maximization in social-networks under independent cascade model. Procedia Computer Science, 173, 315-324.

[2] Pazura, P., Jankowski, J., & Bortko, K. (2020). Modeling the Performance of Limiting Information Spreading Processes under Competing Linear Threshold Model. Procedia Computer Science, 176, 3751-3760.

[3] Kandhway, K., & Kuri, J. (2014). Optimal control of information epidemics modeled as Maki Thompson rumors. Communications in Nonlinear Science and Numerical Simulation, 19(12), 4135-4147.

[4] Kandhway, K., & Kuri, J. (2014). Campaigning in heterogeneous social networks: Optimal control of SI information epidemics. IEEE/ACM Transactions on Networking, 24(1), 383-396.

[5] Kianian, S., & Rostamnia, M. (2021). An efficient path-based approach for influence maximization in social networks. Expert Systems with Applications, 167, 114168.

[6] Singh, A. K., & Kailasam, L. (2021). Link prediction-based influence maximization in online social networks. Neurocomputing, 453, 151-163.

[7] Tripathi, R., & Rao, S. (2022). Rumor containment in peer-to-peer message sharing online social networks. International Journal of Data Science and Analytics, 13(3), 185-198.

[8] Srinivasan, S., & LD, D. B. (2021). A social immunity based approach to suppress rumors in online social networks. International Journal of Machine Learning and Cybernetics, 12(5), 1281-1296.

[9] Tavasoli, A., Shakeri, H., Ardjmand, E., & Young II, W. A. (2021). Incentive rate determination in viral marketing. European Journal of Operational Research, 289(3), 1169-1187.

[10] Olivares, R., Munoz, F., & Riquelme, F. (2021). A multi-objective linear threshold influence spread model solved by swarm intelligence-based methods. Knowledge-Based Systems, 212, 106623.

[11] Kumar, S., Singhla, L., Jindal, K., Grover, K., & Panda, B. S. (2021). IM-ELPR: Influence maximization in social networks using label propagation based community structure. Applied Intelligence, 51(11), 7647-7665.

[12] Ghayour-Baghbani, F., Asadpour, M., & Faili, H. (2021). MLPR: Efficient influence maximization in linear threshold propagation model using linear programming. Social Network Analysis and Mining, 11(1), 1-10.

[13] Gambier, A. (2020, November). Multiobjective optimal control: Algorithms, approaches and advice for the application. In 2020 International Automatic Control Conference (CACS) (pp. 1-7). IEEE.

[14] Zada, I., Naeem Jan, M., Ali, N., Alrowail, D., Sooppy Nisar, K., & Zaman, G. (2021). Mathematical analysis of hepatitis B epidemic model with optimal control. Advances in Difference Equations, 2021(1), 1-29.

[15] Michalewicz, Z., Janikow, C. Z., & Krawczyk, J. B. (1992). A modified genetic algorithm for optimal control problems. Computers & Mathematics with Applications, 23(12), 83-94.

[16] Lotf, J. J., Azgomi, M. A., & Dishabi, M. R. E. (2022). An improved influence maximization method for social networks based on genetic algorithm. Physica A: Statistical Mechanics and its Applications, 586, 126480.

[17] Pastor-Satorras, R., Castellano, C., Van Mieghem, P., & Vespignani, A. (2015). Epidemic processes in complex networks. Reviews of modern physics, 87(3), 925.