1 Introduction

With the proliferation of social media networks in recent years, these communication platforms have become an attractive space for promoting products and attracting customers for both small and large brands. Many individuals who have a substantial following and exert influence in their respective communities (commonly referred to as influencers) offer advertising services based on the rates they define themselves.

Due to the absence of a common pricing policy among influencers, there is significant variability among individuals with similar characteristics. Moreover, these individuals primarily base their pricing on the number of their followers. Consequently, the expected impact of their advertising may not align with reality. In light of these descriptions, the issue at hand can be succinctly defined: how to allocate a limited budget among different influencers to achieve maximum advertising effectiveness. For instance, individuals with fewer followers tend to engage more with their audience, while those with a larger following possess greater message penetration. It is also anticipated that an increase in the number of followers will lead to an increase in advertising costs. These factors, when considered together, raise the question of what the optimal approach should be for advertising a brand.

2 Problem Statement

The question we are seeking to answer here is, in simplified terms, the optimization of a multi-variable function. In essence, the real-world problem and the literature under discussion revolve around how to allocate a certain amount of money among different users/pages on social media networks to spend on advertising in order to maximize the impact.

Thus, the problem can be divided into three smaller subproblems, including:

1. Estimating the cost that each of the users incurs for advertising.

2. Providing a criterion for evaluating the effectiveness of the selected set of users.

3. Determining the relationship between these two functions.

In other words, this complex problem involves estimating advertising costs, evaluating the effectiveness of user/page selections, and finding the optimal way to distribute the budget across these variables for maximum impact.

A Cost Estimation

Advertising costs indeed fluctuate as a function of the number of followers, engagement rate, and community of influencers ($followers(a_i)$). The pricing function can be represented as follows:

$$cost(a_i) = CEF(followers(a_i), engagementRate(a_i), Community(a_i)) + n_i$$

Where:

- " $cost(a_i)$" represents the advertising cost for influencer/user "a_i"

- "CEF" is a function that captures the relationship between the number of followers and advertising cost. However, the specific form of this function needs further investigation and refinement.

- "$followers(a_i)$" represents the number of followers for influencer/user "a_i."

- $engagementRate(a_i)$ represents the impression and engagement of user, may be assumed as a function of the number of likes and replies. (Something like what is explained in next section)

- $Community(a_i)$ represents the label of community of inlfuencer.

- "n_i" is an additional component, which may account for fixed costs or other factors that contribute to the advertising cost for influencer/user "a_i."

As mentioned, determining the precise nature of the "CEF" function will be a crucial step in estimating this pricing function accurately. This function characterizes how advertising costs vary with the number of followers, and further research and analysis will help in better defining its behavior.

B Effectiveness Assessment

It is possible to define a function that represents the effectiveness of each user on other users based on the average exposure each user's posts (or activities in general) receive. On platforms like Twitter, this can rely on metrics such as impressions or views per post, while on other social networks, it may involve a combination of metrics like likes, replies, reposts, and the number of followers. For instance, we can define the following function:

$$imp(a_i)$$

An important point to note is that we must be careful about not double-counting impressions for shared followers when assessing the overall impact of a selected set of users. Therefore, we can define a function "IMP(A)" to evaluate the collective impact of these users while removing the overlapping effects.

In other words, "$imp(a_i)$" indicates how much, on average, each post by user "a_i" engages other users, while "IMP(A)" examines the impact of the set of these users on the network. For simplicity, we can assume that:

$$imp(A) \neq \sum_{i=1}^{n} imp(a_i)$$

However, it's essential to recognize that in reality, it doesn't work exactly this way, and we should not consider the shared views between two users twice (i.e., if one user sees the same ad/post twice). As an initial (albeit likely not entirely accurate) assumption, we can consider that the set "A" under the influence of the "IMP" function provides an approximation of the number of individuals impacted without double-counting. This means that shared followers can be taken into account by considering their ratio to the total number of followers and recalculating the total impressions of the advertising users.

For example, when dealing with two advertising users, we can write these relationships as:

$$A = \{a_1, a_2\}$$

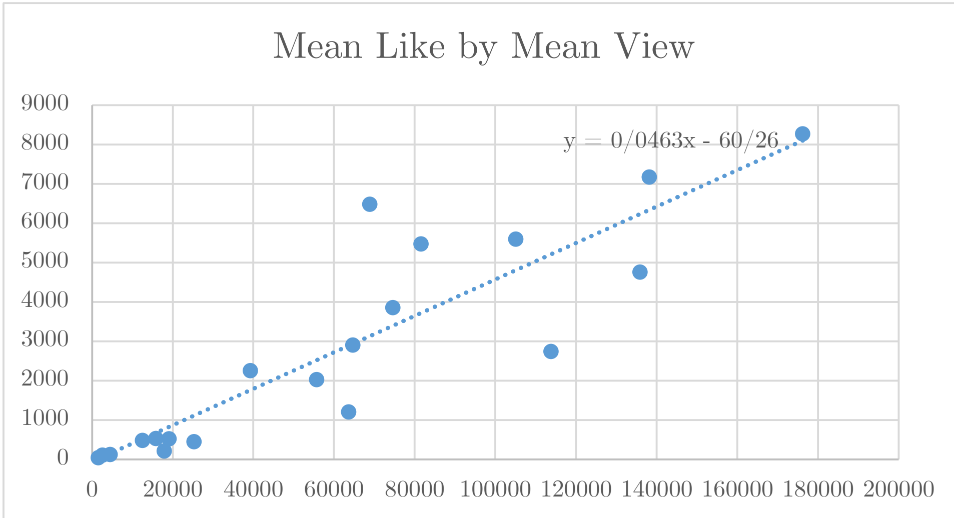
$$IMP(A) = \left(1 - \frac{\left(NumberOfMutualFollowers(a_1, a_2)\right)}{NumberOfFollowers(a_1) + \ NumberOfFollowers(a_2)}\right) \times \left(imp(a_1) + imp(a_2)\right)$$

In this context, "$NumberOfMutualFollowers(a_1, a_2)$" represents the number of mutual followers between users "a_1" and "a_2," and "$NumberOfFollowers(a_1)$" and "$NumberOfFollowers(a_1)$" are the total followers of users "a_1" and "a_2," respectively. This equation attempts to account for the overlapping impressions by adjusting the overall impact calculation.

Estimating the IMP and imp Functions

What ultimately matters to us is the visibility of a post. If we equate this with the level of impressions, we can assume that the visibility of a post is directly related to the amount it is displayed in the feeds of other users, and on average, this measure is directly related to the number of likes. To investigate this relationship, a larger dataset is required, but by examining 10 to 15 tweets from 22 users, we were able to plot the following graph.

It's worth noting that although the study involved 22 users, two users were excluded due to significantly deviating view counts from the average. Consequently, the final graph reflects this data.



In this model, it appears that the number of views can be approximated by a linear relationship with a certain degree of error associated with the number of likes.

C Final Optimization Function

Ultimately, to find the optimal set of users to whom advertising should be outsourced, we can maximize the following function:

$$\mathcal{L}(A) = IMP(A) - \lambda \sum_{a_i \mid a_i \in A} cost(a_i)$$

The optimization variable of this function is, in essence, the selected set of target users. For example, from a larger pool of influencers, we need to choose 'n' users to whom advertising will be outsourced. Our variable is the selection of 'n' users from the set of influencers.

In this function, "IMP(A)" represents the overall impact of the selected users on the network, and the second term, involving the summation, calculates the total cost of advertising for the selected users. The parameter "λ" represents a weighting factor that controls the trade-off between maximizing impact and minimizing advertising costs. The optimization process aims to determine the set of users that maximizes this combined objective.

This optimization problem can be solved using various optimization techniques, such as linear programming, integer programming, or other mathematical optimization methods, to identify the optimal set of users for advertising based on the defined criteria.

3 Workflow

Given the initial scarcity of information regarding these models and how they operate, it is advisable to begin with a series of investigations into the mathematical functions that underlie these models. In summary, the following steps can be proposed:

1. Extraction of Price Tariff Information:

- Collect data relevant to existing price tariffs within the community from various individuals with different numbers of followers.

- Refine the variables affecting prices, such as the number of followers, the number of posts, and similar factors.

2. Extraction of Price Function:

- Analyze the gathered data to estimate the price function based on the influential variables.

- Use mathematical models or statistical methods to model and estimate the price functions.

3. Extraction of Influencer Impact Information and Estimating More Accurate Relationships:

- Collect information related to the impact of each influencer from the data available on social networks.

- Analyze the data to estimate the relationship between the number of followers, the number of likes, the number of views, and the influencer's impact more accurately.

4. Presentation of a Precise Proposed Model to Achieve a More Accurate Solution:

- Develop an optimization model that leverages the price functions and impact functions.

- Determine optimal values for model parameters to solve the optimization problem and select the best set of influencers for advertising.

These initial steps are logical and sound, but throughout the project, it may be necessary to delve into the details of each stage and employ more advanced data analysis and modeling techniques. Additionally, careful consideration of various settings and conducting a variety of experiments will likely be required to arrive at an optimal model for advertising on social networks.

4 Challenges

- An important point to consider in this problem is that, due to the existence of well-known tools within social networks for promoting posts and pages, influencers may have relatively lower appeal.

- Additionally, it should be noted that the reach and visibility of a post are highly dependent on the subject matter being advertised. For instance, advertising a home appliance store located in the United States to Persian-speaking users may not be very effective. This aspect has not been incorporated into the model, and future work may involve introducing a coefficient into the optimization function to match the ad's theme with user interests, requiring additional processes.

- In the final optimization function, we must be mindful that selecting an unknown number of users from a large set and computing the cost for each of them will incur significant computational costs, making it a non-polynomial problem. Alternative solutions can be proposed based on the problem-solving assumptions, meaning that instead of introducing an exact solution (which would be a list of users), an approximate solution regarding numerical features of users will be provided.

- If advertising pricing for each page is done fairly, it can be said that there won't be much difference between using different users with a limited budget. The difference arises from the fact that there is a gap between the user's expected outcome from the ad and the actual outcome. Therefore, the main challenge lies in accurately estimating the price and the actual view count for each page.

- To assess the effectiveness of each influencer user, it is necessary to extract a series of additional relationships that determine the impact of their activities in the network, in addition to the simple mathematical relationships derived. However, this model does not consider these additional relationships.