Finding Least Number of Social Media Influencers that can reach to maximum people

Sangeeth Sajan Baby, Siji Rani S

Department of Computer Science, Amrita Vishwa Vidhyapeetham

September 20, 2021

Outline

Introduction
- Background
- Existing Algorithm
- Problems of Existing Algorithm

Related Works
- Related Works in Detail
- Overall Impression and Research gap
- Our Problem

Proposed Methodology

Dataset and Platform

Validation

Summary

Background

Most advertisements on social media platforms are concentrating on the number of followers. But this is an inefficient way such that the follower overlapping may occur. In this research work, we are aiming to develop a system that reduces the number of influencers and maximize the number of people. This system helps to reduce the advertising cost and management costs significantly. Two important questions are,

- Why the advertising companies need to find the most Influential people?
- How they find the most influential people?

Existing Algorithm

Most of the existing algorithms are finding the the follower count from the graph and design the marketing according to the output obtained.

1. Select the Network for Analysis
2. Find the Influential node with Degree Centrality or K-Core Centrality or Closeness Centrality
3. Locate Multiple Influencers nodes by Independent Cascade Model or Linear Threshold Model
4. Sort the List of influential Nodes in descending order of Follower count.

Reference: [3][4]

Problems of Existing Algorithm

The existing methods is not considering the factors such as Language, Age of the Followers, Personal Interests etc. Problems are not limited to this but,

1. Not taking Care of the Context of Marketing
2. Not considering External Factors
3. Not considering the follower count initially
4. Computational Complexity

Related Works in Detail

A Graph Based Approach For Effective Influencer Marketing[2]

- They will select the network and then analyse the engagements such as Likes and Comments.
- Find the combination of influencer node in which they can reach up to maximum people.
- Drawback: It recommends the Influencer which may be out of scope for the Product to be marketed.

Authors: Salman Ansari, Muhammad Ahsan Tahir and Shahwaiz Bukhari

Code

```
1.  shortlistedInfluencers = []
2.  visitedNodes = []
3.  for i in followersMap:
4.    FIn = snap.TFIn("completeGraph.graph")
5.    completeGraph = snap.TUNGraph.Load(FIn)
6.    audienceA = completeGraph.GetNI(i).GetDeg()
7.     for node in followersMap[i]:
8.         if node not in followersMap:
9.    completeGraph.DelNode(node)
10.   visitedNodes.append(i)
11.    for j in followersMap:
12.       if j in visitedNodes:
13.           Continue
14.    audienceSize                                =
       audienceA+completeGraph.GetNI(j).GetDeg()
15.       if audienceSize>maxAudience:
16.    maxAudience = audienceSize
17.    selectedInfluencerA = i
18.    selectedInfluencerB = j
19. print          "Max          Audience          Possible:
      "+str(maxAudience)
20. shortlistedInfluencers.append(selectedInfluencerA
      )
21. shortlistedInfluencers.append(selectedInfluencerB
      )
```

Figure: Pseudo Code for the process.

Related Works in Detail

Maximizing the Spread of Influence through a Social Network[3]

- Focuses on the set of users that should be targeted when promoting or marketing a product
- Uses the Linear Threshold Model(LTM) and Independent Cascade Model(ICM).

- The major aim of this research is to maximize the spread of Influence and It succeeds in it.

Authors: David Kempe, Jon Kleinberg, Eva Tardos

Related Works in Detail

Machine Learning Techniques for Brand-Influencer Matchmaking on the Instagram Social Network[4]

- Initially crawl the user data and mine associated data such as hashtags, captions etc.
- Use tensorflow to identify objects
- Make predictions of the Inlfuencers
- Drawback: Not mentioning about the potential reach

Authors: Taylor Sweet, Austin Rothwell, Xuan Luo

Related Works in Detail

Finding Influential Users in Social Media Using Association Rule Learning[5]

- Finding the most common behaviour of users using ARM
- Will the find the topic is interested or not and find the relationship
- Drawback: Interest of a group of users not personal preferences

Authors: Fredrik Erlandsson, Piotr Bródka, Anton Borg, Henric Johnson

Overall Impression and Research gap

In each of the researches they are giving some accurate results according to their objectives and aims. But still there are some research gaps existing. In this thesis we can address those gaps,

- Considering the external Factors
- Considering the Context of Marketing
- Follower count analysis
- Addressing the computational complexity

Our Problem

- Maketing agencies may end up in wrong Influencer
- Aim is to find the minimum number of effective influencers

The audience reach can be mathematically expressed as,

$$Maximum reach of Influencers A and B = Reach of A + Reach of B - Common Followers \quad (1)$$

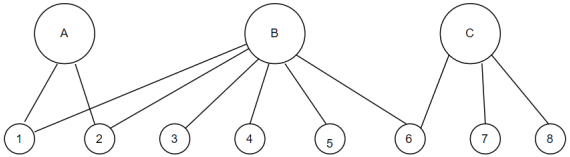

Figure: Sample Network

- A,B,C are influencer nodes and 1-8 are follower nodes
- Our aim is to find the reach of the Influencer

Influencers	Total Followers	Maximum Reach Reach
A	2	2
B	6	6
C	3	3
A,B	8	6
A,C	5	5
B,C	9	8
A,B,C	9	8

- Here B.C and A,B,C have the highest number of followers
- It cannot be concluded that we can use either B,C or A,B,C
- B,C is the correct answer since it has the least number of followers

Basic Steps

1. Load the Network for analysing
2. Apply various Constraints
3. Classify the nodes based on the degree as Follower nodes and Influencer nodes
4. Continue the step 3 until every node is classified
5. Create a list of most influential nodes.

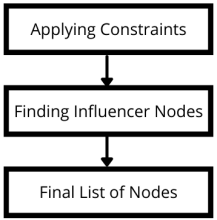

Figure: Major Steps

Applying various Constraints

UserId	UserName	Language
1	A	Malayalam
2	B	Hindi
3	C	French
4	D	Tamil

- Here we need to apply some constraints, such as Language = Hindi
- PySpark code for obtaining the list of nodes having the constraints is,

$$consdf = df.select('*').where(col("Language") == "Hindi") \quad (2)$$

Node Classification and Influencer List

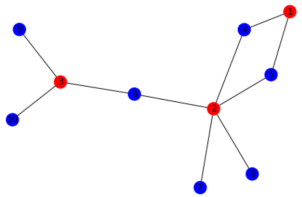

Figure: Node Classification

```
list0 = []
list1 = []
for node in G:
    k = G.degree[node]
    if(k>2):
        list0.append(node)
for node in list0:
    k = G.neighbors(node)
    for m in k:
        if m not in list1:
            list1.append(m)
print(list0)
print(list1)
```

Figure: Pseudo code for finding the list of Influencers

- Initially we will classify the Influencer nodes based on degree
- Then find the list of Influencers

Flow Chart

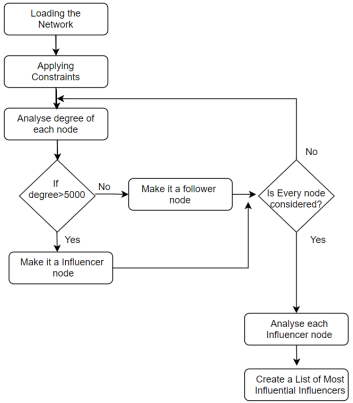

Figure: Flow Chart of Proposed Methodology

Dataset and Platform

Here in this methodology we are using the Twitter Data set Provided by SNAP by Stanford University: https://snap.stanford.edu/data/ego-Twitter.html

Dataset statistics	
Nodes	81306
Edges	1768149
Nodes in largest WCC	81306 (1.000)
Edges in largest WCC	1768149 (1.000)
Nodes in largest SCC	68413 (0.841)
Edges in largest SCC	1685163 (0.953)
Average clustering coefficient	0.5653
Number of triangles	13082506
Fraction of closed triangles	0.06415
Diameter (longest shortest path)	7
90-percentile effective diameter	4.5

Since we need to apply the constraints, we need to use Big data Analytics, so we are using PySpark through Data Bricks for this purpose.

Validation

```
[2, 3]
[4, 5, 6, 7, 8, 9, 10]
```

Figure: Output for Sample program

- The predicted output for the sample program, which is exactly similar to the theoretical output
- Input: Marketing Constraints such a Language, Temperatures
- Output: Least Number of Followers that can reach to maximum users under the constraints.

Software and tools used for this project will be NetworkX and Neo4j. The framework will be ASP.NET or Django.

Timeline

1. Abstract- 28 August 2021
2. First Presentation - 19 August 2021
3. Sample Program Implementation - 15 October 2021
4. Final Implementation - 30 October 2021
5. Sample Report -
6. Final Report -

Summary

Marketing agencies can easily select a person from a pool of social media influencers for promoting their products . But the real problem is with selecting the influencer. This project can easily address by this problem by taking care of,

- Getting the most Influential Users
- Providing maximum reach
- Cost effectiveness

Again one of the major issue that this project might face is the complexity. For the best complexity we need to use the Grover's Algorithm which can reduce the complexity form $O(n^2) to O(\sqrt{n})$.

Future Works

One of the major issue that this project might face is the complexity. For the best complexity we need to use the Grover's Algorithm which can reduce the complexity form $O(n^2) to O(\sqrt{n})$.

References

1. Twitter Data Set: https://snap.stanford.edu/data/ego-Twitter.html
2. A GRAPH BASED APPROACH FOR EFFECTIVE INFLUENCER MARKETING
3. Maximizing the Spread of Cascades Using Network Design: https://arxiv.org/pdf/1203.3514.pdf
4. Maximizing the Spread of Influence through a Social Network: https://www.cs.cornell.edu/home/kleinber/kdd03-inf.pdf
5. Finding Influential Users in Social Media Using Association Rule Learning: https://arxiv.org/pdf/1604.08075.pdf
6. Machine Learning Techniques for brand-influencer Matchmaking on the Instagram Social Network: https://arxiv.org/pdf/1901.05949.pdf

Thank You. Questions?