A GRAPH BASED APPROACH FOR EFFECTIVE INFLUENCER MARKETING

Salman Ansari, Muhammad Ahsan Tahir and Shahwaiz Bukhari

Techlets Pvt. Ltd.

ABSTRACT

We present a novel graph-based approach to find the optimal set of influencers from a large pool of influencers. The goal is to select minimum number of influencers that can reach the desired audience. In order to find such a set, one has to compute the reach of all possible combinations of available influencers resulting in complexity of order O (2^n). Our proposed greedy approach selects the pair of influencers that results in highest reach at every iteration reducing the complexity to O(n^2). Our work is complimented with analysis of 550 Instagram influencers and over 100,000 post. After the analysis, we concluded that influencers who prefer quality over quantity receives better engagement. Influencers sharing 3 posts per week and posts with caption length of over 500 characters relatively received better engagement numbers.

KEYWORDS

Graph Analytics, Influencer Marketing, Social Media Influencer, Greedy Algorithm

1. INTRODUCTION

With 3G/4G telecom services now common and offered at among the world's most affordable rates in Pakistan, 67 million locals have mobile broadband internet access. According to a Gallup & Gillani poll, 48% of internet users consume online content daily[1]. One content producer category is the "Social/Digital Influencer" who are followed by 'fans' numbering from a few thousands (micro-influencers) to millions (celebrities). Their digital content, whether pictures, videos and live streams, are delivered on digital and social channels – Instagram, Facebook, YouTube, Beigo, TikTok – and cover a wide range of genres from lifestyle, fashion, health to entertainment and comedy.Brands can use social and digital influencers to launch and manage effective organic marketing campaigns. Pakistan's SMEs, comprising 98%+ of all companies, actively use influencer marketing as a cost-effective strategy to engage and sell in/directly to Pakistan's online population[2], an estimated "trade" value at PKR 5 billion*/year[3]. For sellers, influencers represent an impactful and efficient outreach channel as they affect purchase behaviors and choices, brand perceptions and even the national narratives of ~30% of Pakistan's population[4].

A key challenge is to optimize influencer marketing budgets by cost-efficiently identifying the most *optimal set of influencers* who can effectively reach desired audiences, from prospects, buyers to promoters. To only consider an influencer's number of followers is misleading as there is an *underlying overlap* between the followers of influencers who have contextual similarities – that is two influencers with the same city, age group, same genre/category and content themes are likely to appeal to the same people following them both. To solve this overlap problem, we are focusing on one of the most influencer-driven social networks, Instagram. We propose a Greedy Algorithm to rationalize the minimum number of influencers to attain *the best* reach.

DOI: 10.5121/ijdkp.2019.9402

Our proposed solution is supported by analyzing500+ influencer profiles and over 100,000 Instagram posts to establish trends and benchmarks for influencers. For example; the average post's engagement is 3.28% of total followers. Based on our findings those influencers who consistently had a higher engagement rate, appeared to emphasize on quality over quantity, as posts had detailed and meaningful caption length swith frequency of up to three posts/week. The key constraint to our approach is to have a comprehensive list of followers of all the influencers as overlap cannot be found between the influencers without the complete list of all the users following the influencers.

2. RELATED WORK

Influencer marketing is described as the "art and science of engaging people who are influential online to share brand messages with their audiences in the form of sponsored content" [5].Social media influencers who are responsible for influencer marketing are the third-party endorser responsible for shaping attitude of the audience towards the brands through videos, blogs and social media [6].

To solve our defined problem, we explored graph-based solutions as Influencers-Followers relationships that can best be explained in graphical representation with nodes as Influences/Followers and edges as interactions between them. We examined several different techniques and tried to model our problem on them. Bhamaikar and Rao [7] worked on identifying cliques using degree and connectivity constraints, we tried to model our problem and modify their approach to create and analyze cliques of followers to find the true reach of all influencers but realized that clique detection was not the right representation of the problem we were working to solve. Leskovec, Lang, Dasgupta, Mahoney [8]employed approximation algorithms to identify community structures in large social networks, we intended to employ their approach to estimate true reach of Influencers by identify communities of followers. Leskovec, Backstrom and Kleinberg [9] suggested to monitor information flow using content of the information like hash tags or memes. Sawhney, Prasetio and Paul [10] combined graph structure and semantic understanding of the text for community detection. We worked to identify overlapping followers by monitoring the content they post on their profiles but it is too much of an overhead to monitor the content of all followers of the influencers.

To the best of our knowledge and research, we were **unable to find any body of work directly related to our problem**. We therefore propose novel approach to solve the problem of finding effective influencers by introducing the "Greedy Algorithm".

3. KEY DEFINITIONS:

3.1. Influencers:

Instagram users with a substantial following, engagement rates, working with brands or creating their own content for the purpose of advertisement, promotions and outreach.

3.2. Overlapping Followers:

The followers of an influencer who also follows other influencers. If user 'A' is following two influencers IA and IB than 'A' is considered the overlapped follower of IA and IB.

3.3. Reach:

The maximum number of users who can be reached with a selected influencer. For any influencer, reach is the total number of its followers.

3.4. True reach of influencers:

The number of distinct users who can be reached with selected influencers. For two influencers IA and IB,it is calculated as:

$$TrueReachofIAandIB = (ReachofIA + ReachofIB) - OverlappingFollowers$$

3.5. Engagement:

The total activity received on any post, defined as:

$$Engagement = SumofLikes + SumofComments$$

3.6.Influencer Node:

The nodes in the graph representing influencers.

3.7. Follower Node:

The nodes in the graph representing followers.

4. PROBLEM STATEMENT

Marketing agencies can select from their in-house directory of influencers talent to be allocate to a marketing campaign. As one Instagram user can follow several users on the platform, there is a high probability that there will be overlapping of followers between the influencers (a follower following more than one influencer), hence the actual reach of the selected influencers will be far less than the sum of total number of their followers.

Every brand wants to optimize marketing spend. They want to continuously improve reach metrics by using the right combination of influencers to maximize reach using minimal budget spend.

Our sample dataset [Section:Dataset] had an average overlap rate of 11% between followers of two influencers. As the number of selected 'similar' influencers increased, this overlap rate increased making it an even bigger issue in terms of increasing unique reach. To as certain this problem's magnitude, we considered the following example:

A marketing agency selects four influencers (IA, IB, IC, ID) registered with them, but their total in-house database could be in the millions. The dynamics of these influencers are illustrated in Tables 1 showing the followers, Table 2 tallying the true reach and Figure 1 which shows the graphical representation.

Table 1: Sample Four Influencers and their followers

Influencer	Total Followers	Followed By
IA	6	a,b,c,d,e,f
IB	4	a,b,g,h
IC	4	c,d,g,h
ID	3	a,i,j

Table 2: True reach for each possible follower combination

Selected Influencer	Sum of Followers	True Reach
IA	6	6
IB	4	4
IC	4	4
ID	3	3
IA, IB	10	8
IA, IC	10	8
IA, ID	9	8
IB, IC	8	6
IB, ID	7	6
IC, ID	7	7
IA, IB, IC	14	8
IA, IB, ID	13	10
IB, IC, ID	11	8
IA,IB,IC,ID	17	10

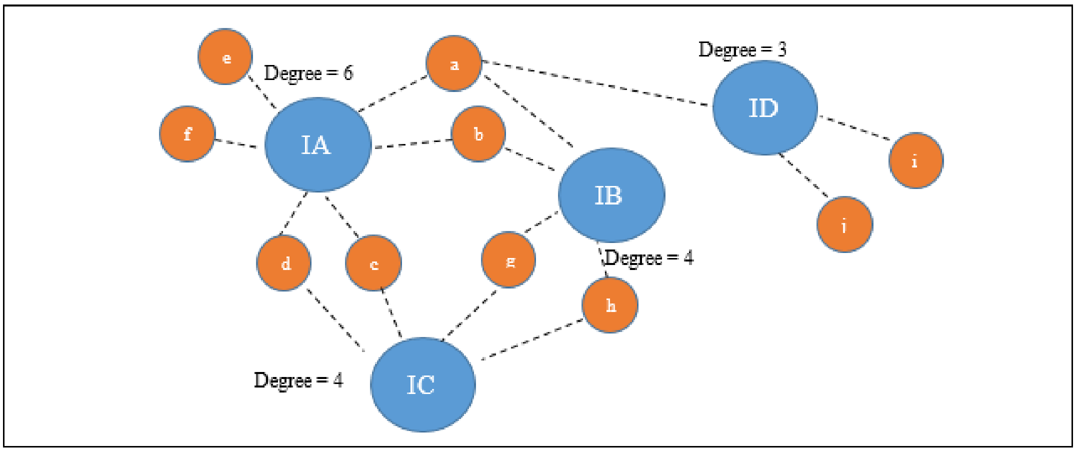

Figure 1 – Graphical data representation.

Larger nodes represent influencer-nodes while smaller nodes represent follower-nodes.

From the above scenario, influencer IC's followers can all be reached with influencers IA and IB so the addition of influencer IC does not increase the total reach. The impact of a brand's selection constrained by budget if they limit by having:

(1) Two influencers: choosing IA and ID will be more economical over the obvious choice of IA-IB or IA-IC, as the true reach of IA-ID is 8 which is the same as the true reach of IA-

IB and IA-IC and more than the true reach of IB-IC that is 6, as ID has less number of followers than IB and IC and should charge less than the other two (Ref Figure 1 and Table 2), an improved reach by 33%.

(2) Three influencers: the combination of IA-IB-ID will be more efficient than IA-IB-IC despite the sum of total followers of IA-IB-IC (14) being more than the total followers of IA-IB-ID(13) as the *true reach* of IA-IB-ID is 10 while that of IA-IB-IC is 8, an improved reach of 25%.

In the real world, the graph of all the available influencers and their followers, contains millions of nodes and edges so to find the best solution, one will have to explore all the possible combinations resulting in the complexity of O (2^n).

5. OUR SOLUTION

To solve the problem of reaching the maximum audience by selecting specific influencers from a set of influencers, we propose a Greedy Algorithm. The heuristic is to keep combining pair of influencers from a set of available influencers that generates the maximum reach until the required reach or set of influencers is obtained. We used **SNAP-Stanford Network Analysis Project**[11] to generate and manipulate the graph.

5.1. Parameters

The proposed algorithm allows users to set following parameters, or alternatively the algorithm can select a default value.

5.1.1 Target Reach

Total audience the brand wants to target. The set of influencers under consideration should have total number of unique followers equal to or greater than the targeted reach.

5.1.2 Macro Influencer Threshold

Minimum number of followers required by the influencer to be considered a Macro Influencer. The system uses a default value of 300,000.

5.1.3 Micro Influencer Threshold

Minimum number of followers required by the influencer to be considered a Micro Influencer. The system uses a default value of value is 100,000. Any influencer having followers between Micro Influencer Threshold and Micro Influencer Threshold is considered Micro Influencer.

5.1.4 Number of Macro Influencers

The maximum number of macro influencers who can be part of selected influencers.

5.1.5 Number of Micro Influencers

The maximum number of micro influencers who can be part of selected influencers.

5.2. Methodology

The goal is to manipulate the influencer graph so that **the sum of degrees of the influencer nodes produces the true reach of the influencers**. To obtain the true reach of two influencers, a two-step process was repeated until two nodes are selected:

(1) After selecting the first influencer node, we removed all the nodes representing followers (follower nodes) of the selected influencer from the graph.

(2) The degree of all the influencer-nodes that were connected with the removed follower nodes will decrease ensuring that overlapping follower nodes between the selected influencers are considered only once. (Ref: Figure 2(a) and Figure 2(b))

Note, before removing the nodes from the graph, the degree of the influencer under consideration is stored for future processing.

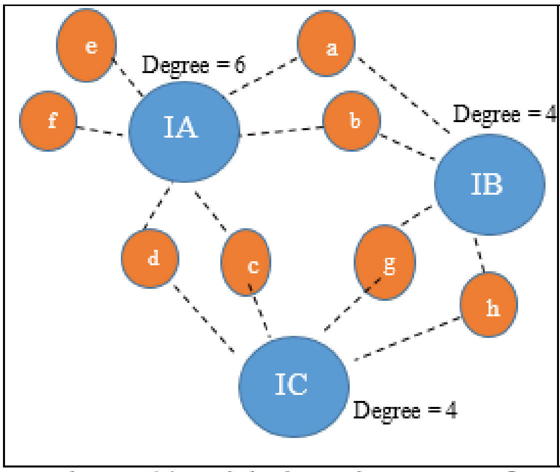

Figure 2(a): Original Graph – Degree of IA:6, IB: 4, IC: 4

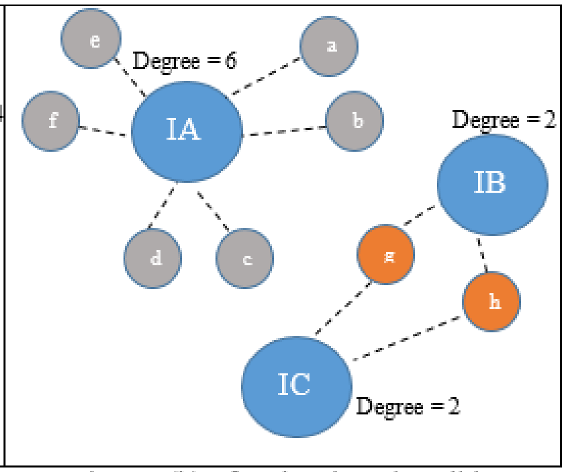

Figure 2(b): If IA is selected – All its follower nodes will be removed from the graph, reducing the degree of IB and IC to 2 from 4.

With reference to the example in Figure 2, the influencer IA degree that is 6 is stored and a, b,c,d,e,f are removed from the graph. After removing the nodes, the sum of degree of influencer IA and any other influencer will result in the true reach of the both influencers, as the overlapping nodes are only considered while storing the degree of influencer IA.

5.3. Algorithm

The first phase of our algorithm is to pre-process and transform the data into a structure that can generate a graph. We create an Influencer dictionary as keys and a complete list of all their followers as the value.

For the example in Figure 2A, our follower Map will be:

followerMap = {'IA':[a,b,c,d,e,f], 'IB':[a,b,g,h], 'IC':[c,d,g,h]}

Once we have the follower Map an undirected graph as represented in Figure(1) is created, using SNAP.

Code	Description
1. completeGraph = snap.TUNGraph.New() 2. for i in followersMap: 3. if not completeGraph.IsNode(i): 4. completeGraph.AddNode(i) 5. for j in followersMap[i]: 6. if not completeGraph.IsNode(j): 7. completeGraph.AddNode(j) 8. completeGraph.AddEdge(i,j) 9. FOut = snap.TFOut("completeGraph.graph") 10. completeGraph.Save(FOut) 11. FOut.Flush()	1. Creates undirected graph 2. Visits all influencers in our mapping 3. Checks if a node representing influencer exists in graph or not 4. Adds a new node 5. Visits all followers of an influencer 6. Checks if a node representing follower already exists in graph. 7. Adds a new node for follower 8. Adds an edge between Influencer and follower 9-11.Saves the initial state of the graph for later use.

The main algorithm is summarized in the following steps:

(1) Select an influencer node (i)
(2) Store the degree of influencer node (i) and remove all its follower nodes (followers) from the main graph, degrees of all the influencer nodes sharing the removed follower nodes will be updated
(3) Visit every other influencer node and add the degree of each influencer node in the degree of influencer node (i) and find the pair that results in highest reach, i.e., has the highest sum of degrees. Record the value of this reach and the selected pair.
(4) Restore the graph's initial state and perform steps 1 to 3 for each influencer node.
(5) Select the pair that results in highest reach. The reach is stored, and all their follower nodes are removed from the graph. The graph's initial state is updated with this state, i.e., after the removal of follower nodes of selected influencer nodes.
(6) Repeat steps 3 to 5 until the targeted reach is achieved or all influencer nodes have been selected.

5.3.1 Finding the first influencer pair

After creating the graph in Figure 1, the first step is to find the initial two influencer nodes that will result in highest reach is shown in Figure 3a.

Code	Description
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()	1. Defines List to store the selected influencers 2. Defines List to record the visited influencer 3. Visits all influencer nodes in our graph (i) 4-5 Restores previous state of the graph 6. Stores the degree of Influencer node under consideration (i) 7-9.Removes all follower nodes of the Influencer (i) from the graph 10 Marks influencer under consideration (i) as visited. 11-18 Visits all other influencers (j) and checks which influencer combined

Code	Description
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(selected InfluencerB)	with the influencer (i) results in maximum reach. 14 Adds the degree of second influencer (j) and degree of first influencer (i) to see, if they result in maximum reach. 15-18 If the degree of (j) and (i) is greater than previously recorded maximum, stores the information 20-21 The selected nodes are added in the list of shortlisted Influencers

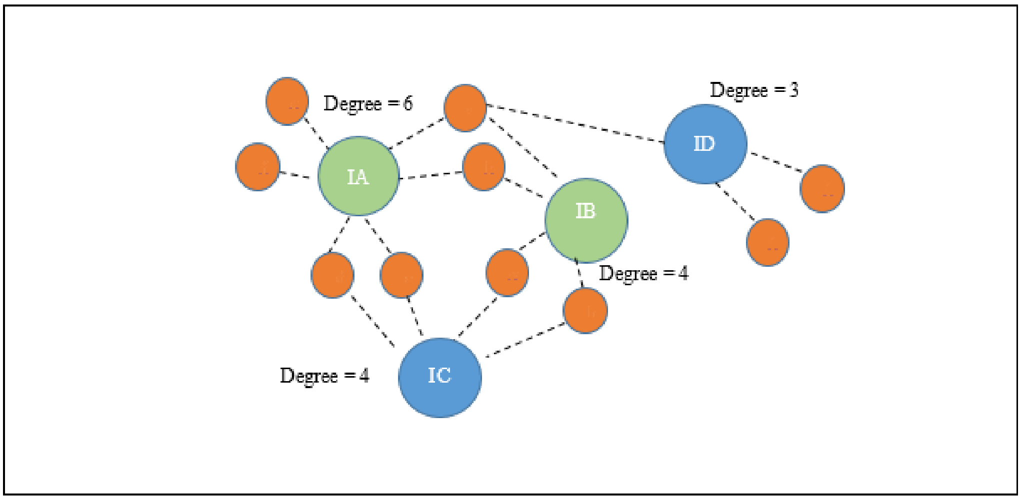

Figure 3(a) – As influencers IA and IB result in the highest reach, they are selected.

5.3.2 Update Graph States after Initial Influencer Selection

Once the initial nodes are selected, the sum of their degrees is stored as max Audience, the maximum Audience that can be reached after selection of these two influencer nodes. All the follower nodes are removed from the graph and the graph's initial state is updated to this new state as shown in Figure 3(b).

Code	Description
1 for i in shortlisted Influencers: 2 for node in followers Map[i]: 3 if complete Graph. Is Node(node) and node not in followers Map: 4 complete Graph. Del Node(node) 5 del followers Map[i] 6 FOut = snap.TFOut("complete Graph. graph") 7 complete Graph. Save(FOut) 8 FOut. Flush()	1 Visits both shortlisted influencers(i) do 2-5 2 Visits each neighboring node of shortlisted influencer(i) 3-4 Checks if the neighboring node is not influencer, then remove from graph. 5 Removes the shortlisted influencer from the followers Map too as it has already been shortlisted as a potential influencer. 6-8 Update the initial state of the graph is this modified state.

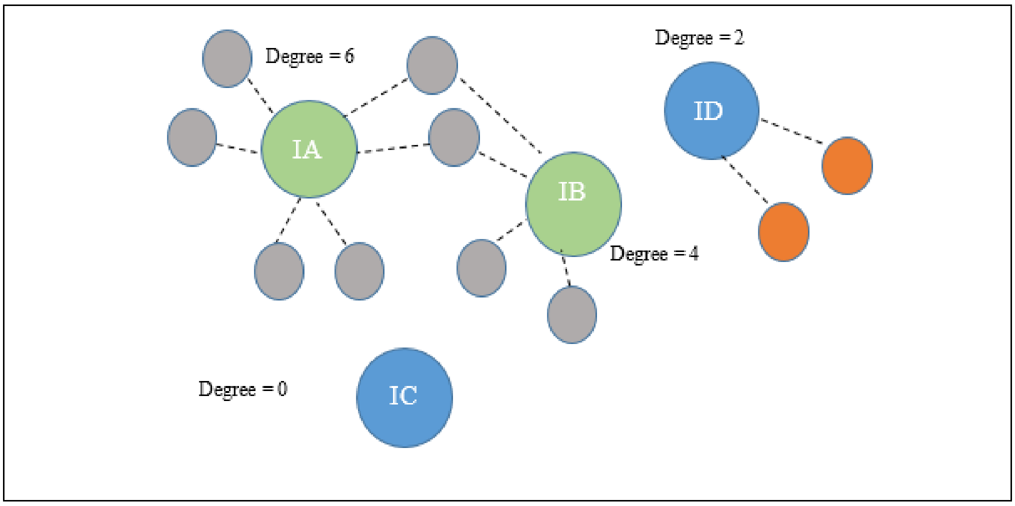

Figure 3(b) – As influencers IA and IB result in the highest reach, they are selected

5.3.3 Iteratively selecting influencers to achieve desired reach

Once the initial pair of influencers is selected, the merged pair is compared with every other remaining influencer to measure which of the remaining influencer, if merged with selected influencers, will result in maximum reach. If such an influencer is found, it is merged with already merged influencers. The process continues until the desired reach is achieved or all influencers have been visited, as shown in Figure 3(c).

Code	Description
1 while len(followersMap)>0: 2 currentAudience = maxAudience 3 if currentAudience>desiredReach: 4 break 5 audience=0 6 for i in followersMap: 7 FIn = snap.TFIn("completeGraph.graph") 8 completeGraph = snap.TUNGraph.Load(FIn) 9 audience = completeGraph.GetNI(i).GetDeg() 10 audienceSize = currentAudience+audience 11 if not audienceSize<maxAudience: 12 maxAudience = audienceSize 13 selectedInfluencer = i 14 shortlistedInfluencers.append(selectedInfluencer) 15 for node in followersMap[selectedInfluencer]: 16 if completeGraph.IsNode(node) and node not in followersMap: 17 completeGraph.DelNode(node) 18 del followersMap[selectedInfluencer] 19 FOut = snap.TFOut("completeGraph.graph") 20 completeGraph.Save(FOut) 21 FOut.Flush()	1 Continues until all influencers have been explored 2 Stores the audience that can be reached with current set of influencers 3-4 Checks if desired reach can be achieved with current set. 5 Audience – to store followers of the influencer that will be explored 6 Explores all other influencers 7-8 Restores the state of graph 9 Gets the audience of the node that is being explored 10 Getsthe audience size that can be reached by adding

	this influencer with pre-selected set of influencers 11-13 If the influencer being checked is higher than the previously recorded reach, then update with the new influencer 14 Once all influencers areprocessed, the influencer with thehighest reach is added to the list of selected influencers. 15-18 Removes all the follower nodes of the selected influencer from the graph and removes it from followers Map 19-21 Updates state of the graph with this modified state.

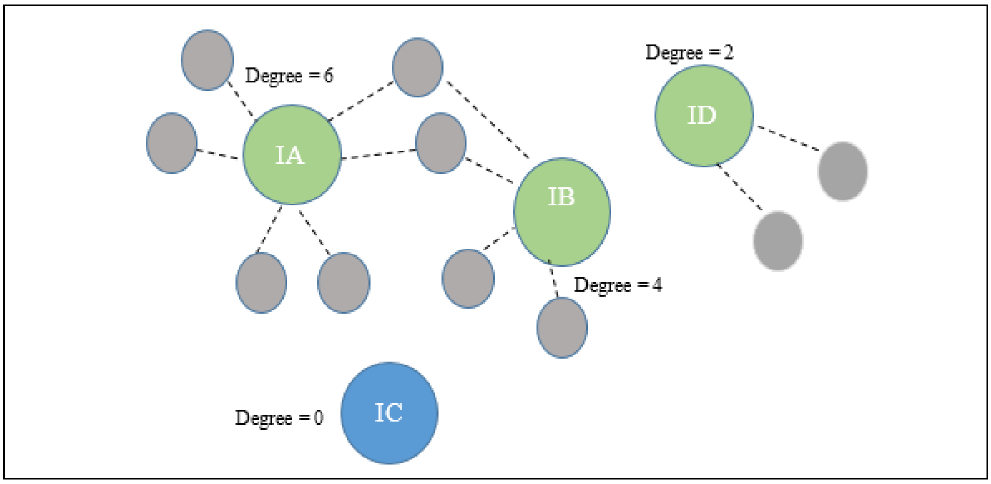

Figure 3(c) – influencer ID combined with influencers IA and IB results in a higher reach than influencer IC combined with IA and IB

Once the script is completed, the variable shortlisted Influencers will have the complete list of all the shortlisted influencers. This general algorithm, run within the script, is further updated to cater for useful parameters, for example:

(1) The number of followers and influencer should have to be categorized as either a macro-or micro-influencer.
(2) The total number of macro- and micro-influencers who can be part of the final list of shortlisted influencers.

These parameters are important to optimize the results as per the brand's budget. View the complete python script at: bit.ly/2X9lalS

5.4. Implementation with High Performance Computing

In order to make our proposed algorithm computational feasible, High Performance Computing (HPC) environment can be employed. We designed a complete pipeline using Nifi to ingest data from our database into Hadoop and calculate the reach of all influencer pairs that is needed at each iteration in parallel. Based on the results of these parallel computations, the influencer is selected that results in highest reach when combined with set of selected influencers see Figure (4).

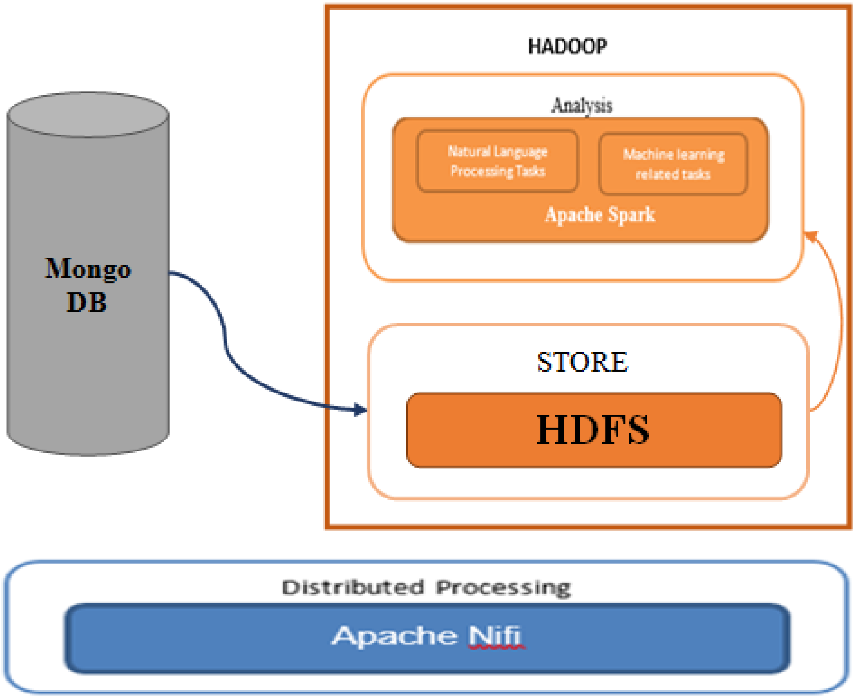

Figure 4 – High level architecture of HPC

5.5. Dataset

We identified 500+ influencers from Pakistan with 5,000 to 300,000 followers and collected their publicly available data which includes:

(1) Public profiles including biography, number of followers and number of people they follow
(2) Publicly available profile feeds for the last six months, including the caption of the posts and number of likes and comments received on each post
(3) Followers list.

To keep our analysis generic and unbiased, we tried to select influencers from a wide variety of thematic areas and niches but despite best efforts, around 70% of our influencer are fashion bloggers.

6. RESULTS

Our objective was to measure the effectiveness of our greedy approach by analyzing the reach and overlap of the influencers selected by our algorithm. We were also interested in establishing the KPIs and benchmarks to measure the performance of Instagram Influencers. This section discusses all the results and observations that were made.

6.1. Followers Overlap

The average followers overlap between any two influencers in our dataset was around 11%. The highest overlap between two influencers was found to be 90% with followers 210,241 and 37,858. The lowest overlap in our dataset was 0.05% between two influencers with followers 86,485 and 11,630.

We used our proposed solution to find the set of influencers that can reach at least 1 million users from our dataset. We also set the constraint of having only 1 influencer having followers more than 200,000 in the final selected set. If our solution wasn't used and top influencers were chosen from the dataset the desired reach could have received with 5 influencers having followers *303060, 194519, 190430, 190343 and 186984* respectively. In reality the true reach of these 5 followers is only**724,302** instead of **1,065,336** (sum of reach of all influencers).

Our solution selected 6 influencers with followers *279065, 194945, 190343, 186984, 133920, 131886, 101426* respectively. The true reach of selected influencers is **1,035,474** and the reach is **1,218,568**(sum of reach of all influencers) [Table 3]. It should also be noted that the influencer with highest followers in our dataset having followers 303,060 was not selected by our proposed solution as it has quite a high overlap with the followers of other influencers in our dataset.

Method	**Reach of Selected Influencers**	**Followers Overlap**	**True Reach of Selected Influencers**
Selecting top influencers	1,065,762	341,460	724,302
Using Proposed Greedy Approach	1,218,568	183,094	1,035,474

Table 4 – Method is the technique used to select influencers, Reach of Selected Influencers is the sum of reach of all influencers, Followers Overlap is the number of users following more than one influencer and True reach of Selected Influencers is the number of actual unique users that can be reached with selected influencers.

6.2. Analysis of Influencer Profiles

We performed a detailed analysis of our dataset, with the objective to observe a relationship between influencers and their profile activity by analyzing the content they posted and establish benchmarks used to measure the influencers' effectiveness.

6.2.1Average Engagement of a Post

The average influencers engagement was 3.28% with only 29 out of 550 influencers had average engagement more than 10%, based on the influencers' profile feed.

$$Average\ Engagement = \frac{Engagement\ on\ post}{Total\ followers\ of\ the\ influencer} * 100$$

6.2.2 Post Frequency/Week

We calculated that most influencers do not post more than three posts in a week and of 550 influencers, about 25% or 140 posted more than five times/week i.e., more than one post daily as shown in Figure 5.

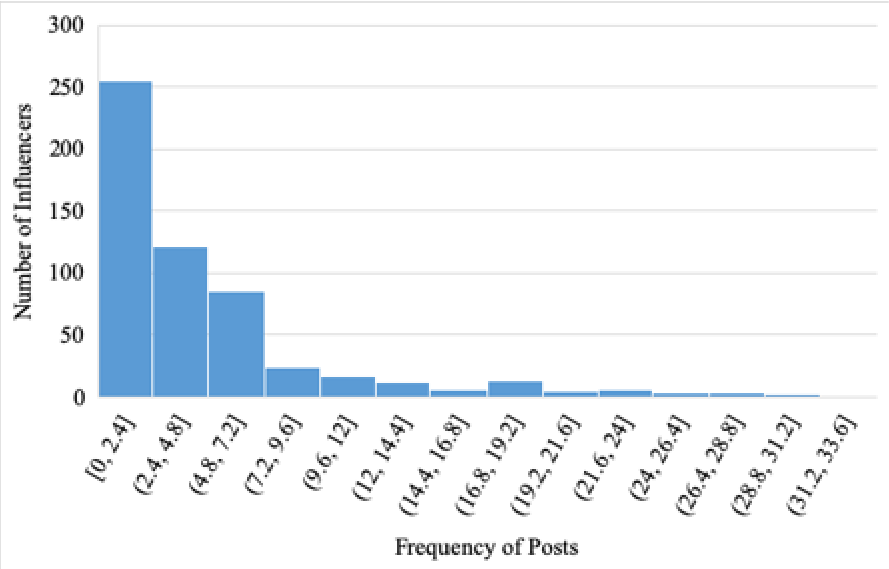

Figure 5 - Post frequency of Influencers

6.2.3Average Engagement vs. Post Frequency

To find the optimum number of weekly posts that an influencer should make, we calculated the average frequency that resulted with the highest engagement number. There was no conclusive trend, but the data showed that the probability of a higher engagement is with less number of weekly posts i.e., influencers who posted 2 to 3 times/ week received relatively better engagement rates than influencers who posted more frequently as shown in Figure 6.

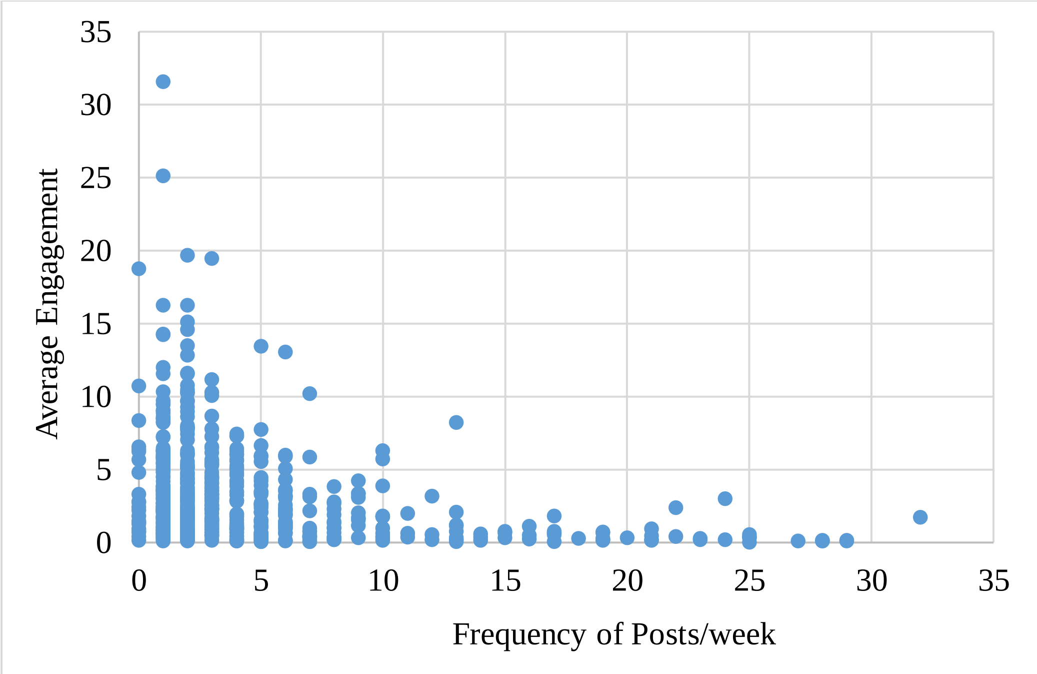

Figure 6-Average Engagement vs. Frequency of Posting/week

6.2.4Average Engagement vs. Caption Length

Posts with longer captions are more likely to get higher engagement than the posts with shorter caption based on analyzing 100,000+ posts by550 influencers. Around 35% of the posts analyzed had caption length less than 200 characters, around 45% had lengths between 200 and 600 characters and only 20% had caption length of more than 600 characters, as illustrated in Figure 7.

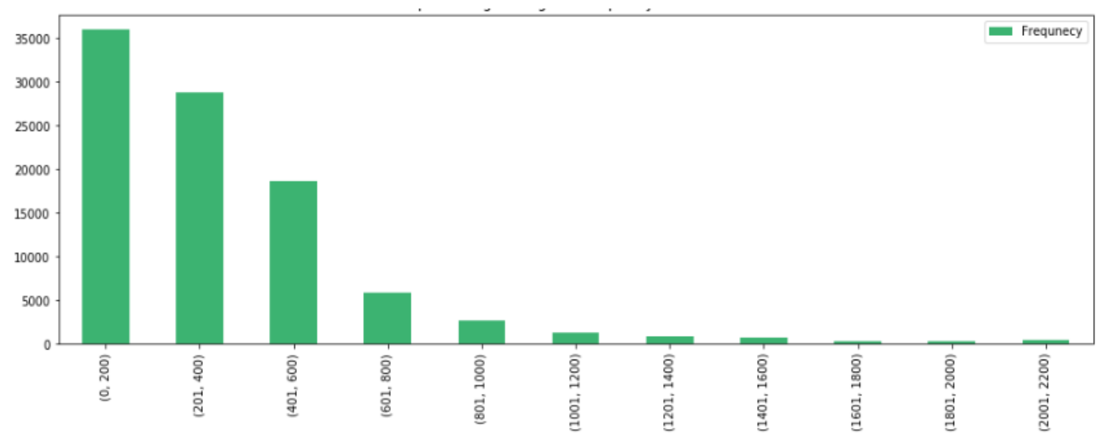

Figure 7: Caption Length Range vs. Frequency of Occurrence

The analysis shows posts with longer captions are more likely to get good engagement. Based on the analysis, every post with higher than the 5% engagement rate has a caption length longer than 500 characters, as shown in Figure 8.

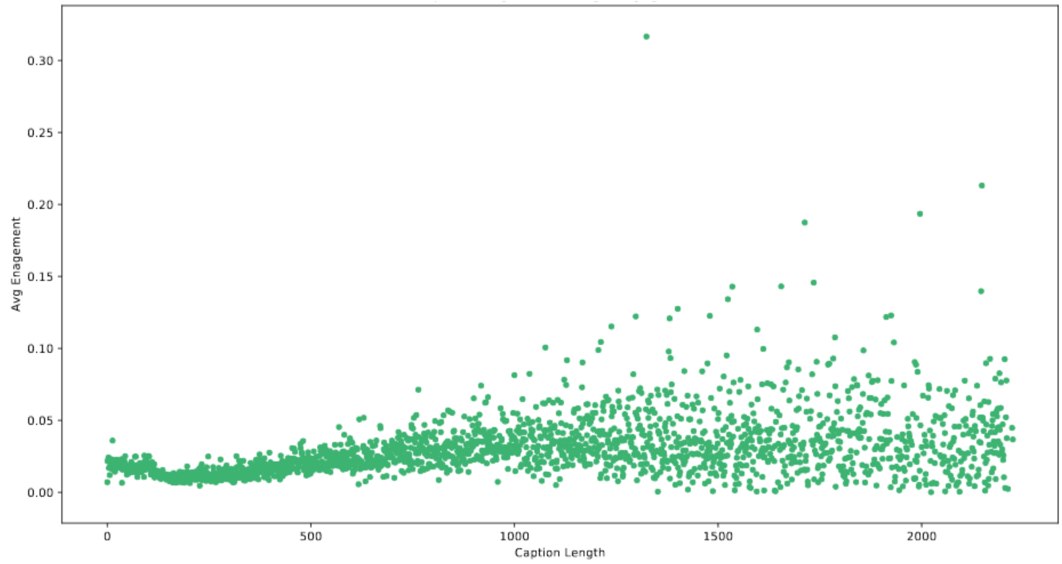

Figure 8: Caption Length vs. Average Engagement

6. Constraints And Limitation

The key constraint to our approach is to have a comprehensive list of followers of all the influencers as overlap cannot be found between the influencers without the complete list of all the users following the influencers. In our case, Instagram does not support any method to automatically download or acquire the list of all the followers, so, it becomes an additional cumbersome task to obtain the complete follower lists of all the influencers registered with the agency. In order to remain up to date, these lists need to be updated rather frequently too.

7. Future Work

During this work, we proposed the Greedy Algorithm approach to calculate the optimal set of influencers to maximize reach. This work can be extended to find the optimal set of influencers that maximizes impressions, without considering reach, thereby enabling brands to reach the same audience set multiple times via different influencers. In generating the optimal influencer set, users could only restrict the number of macro- and micro-influencers. To better optimize brands' budgets, the influencers sets can be further processed to include influencers' compensation rates.

8. Conclusion

Our main goal was to propose a Greedy Algorithm for marketing agencies performing influencer marketing, to choose an optimal set of influencers to achieve higher reach by analyzing all the possible sets of their registered/accessible influencers. We showed the problem's complexity to find the best solution i.e., the influencers sub-set who are likely to generate the maximum reach as O (2^n). Our proposed solution finds the optimal solution in O (n^2).

Further, after analysis of 500+Instagram influencers and 100,000+ posts over six months, we found the average influencer engagement rate is 3.28% of their followers. The data also showed that quality content was more important than quantity to achieve higher engagement rates as influencers who posted up to three times per week and posts with caption lengthsof 500+ characters had higher engagement numbers.

9. ACKNOWLEDGEMENTS

The work presented is part of project "Big Data Analytics & Artificial Intelligence for Consumer Profiling in Pakistan" fully funded by Ignite (formerly National ICT R&D Fund). We would also like to extend our gratitude to anonymous reviewers for their valuable insights and recommendations.

REFERENCES

[1] Gallup & Gilani Pakistan, "48% internet users in Pakistan claim to use social networking sites to access news at least once a day", Gallup Pakistan, https://gallup.com.pk/48-internet-users-in-pakistan-claim-to-use-social-networking-sites-to-access-news-at-least-once-a-day/, Accessed: 6-June-2019

[2] Pakistan Telecommunication Authority, "Telecom Indicators", Pakistan Telecommunication Authority, https://www.pta.gov.pk/en/telecom-indicators ,Accessed: 6-June-2019

[3] Fahad Khan Niazi, "Policy & regulatory bottlenecks for digital financial services in pakistan", Karandaaz, https://karandaaz.com.pk/wp-content/uploads/2019/06/Policy-Brief-New-1.pdf ,Accessed: 6-June-2019

[4] Influencer Marketing Hub, "The State of Influencer Marketing 2019 : Benchmark Report", Influencer Marketing Hub, https://influencermarketinghub.com/influencer-marketing-2019-benchmark-report/, Accessed: 6-June-2019

[5] Sammis, K., Lincoln, C., Pomponi, S., Ng, J., Gassman Rodriquez, E., & Zhou, J. (2015). Influencer Marketing for Dummies. Hoboken, New Jersey: Wiley

[6] Freberg K., Graham K., McGaughey K., Freberg L. (2011). Who are the social media influencers? A study of public perceptions of personality, Public Relations Review.

[7] AvinashBhamikar A., Ramchandra Rao P., (2012). Detecting Cliques Using Degree and Connectivity Constraints, International Journal of Data Mining & Knowledge Management Process

[8] Leskovec J., J. Lang K., Dasgupta A., Mahoney M. (2011) Community Structure in Large Networks: Natural Cluster Sizes and the Absence of Large Well-Defined Clusters, International Mathematics.

[9] Leskovec J., Backstrom L., Kleinberg J. (2009), Meme-tracking and the Dynamics of the News Cycle, International Conference on Knowledge Discovery and Data Mining

[10] SawhneyKartik, Presetio Marcella Cindy, PualSuvadip (2013), Community Detection Using Graph Structure and Semantic Understanding of Text, Knowledge-Based Systems

[11] Leskovec J., Rock S. (2016), SNAP: A General-Purpose Network Analysis and Graph-Mining Library, ACM Transactions on Intelligent Systems and Technology (TIST)

AUTHORS

Salman Ahmad Ansari Working as a Data Scientist in Techlets Pvt Ltd., Pakistan. Area of Interests: Machine Learning, Graph Theory, Data Analysis, Social Media Marketing and Natural Language Processing.

Ahsan Tahir CEO at Techlets Pvt Ltd, Pakistan. Vast Experience in Big Data Analytics

Syed Shahwaiz Bukhari Data Associate at Techlets Pvt Ltd., Pakistan. Area of Interests: Big Data Analytics, Deep Learning and Data Analysis.