

Interactive Bubble Graph To Visualize Instagram Influencers' Engagement Rates

Ahmed Al Hasani and Alex Costinescu

INFO 5602

ABSTRACT

With the rise of social media stars, influencer advertising has become a sizable industry. Due to the prevalence of purchased likes, comments, and followers it has become important for advertisers to verify that accounts are actually good targets for advertising campaigns, typically through the metric of engagement. This project follows several Omani accounts and finds engagement scores for likes, watches, and comments over a period of time. These engagement scores will be used by an Omani company to determine which accounts are good potential advertising targets. We found that a high engagement rate for comments was rare and like a sign of highly dedicated followers while low engagement for likes and watches, lower effort metrics, likely signify accounts that would be less valuable for advertisers.

Keywords: Instagram, influencers, advertisement, engagement

1 INTRODUCTION

The popularity of social media platforms, such as Instagram, has given birth to a new profession - Influencers. These are accounts which have many, typically loyal followers who, as the name suggests, they have influence over. This is of particular interest to advertisers, because it gives them very specific access to target groups for their ad campaigns and a trusted spokesperson to present their product to these groups.

However, purchased followers and engagements have also become prevalent. This has made it important for advertisers to check if influencer accounts have real followers - typically through the metric of engagement. Engagement refers to a group of related metrics that measure the ratio of a given account's followers that respond to their posts. Likes, watches, and comments are typical responses that are tracked to determine an account's engagement. For example, an account with high engagement would consistently have likes and watches close to their follower count and comments proportional to that count. It is rare to see high engagement in comments because of their relatively high effort nature in comparison to likes and watches. In contrast, a low engagement account would have little to no comments and like/watch amounts significantly lower than their follower count would suggest.

Engagement is also important to track over time, because it is possible for accounts to buy likes, watches, and comments for periods of time. An account whose engagement spikes for a period of time has likely purchased followers.

Our goal was to determine which influencer has appealing metrics if we would place an ad in profile, using the visualization tool we developed. We tracked 28 accounts of varying follower count for

16 days, logging the number of posts they made per day, their follower counts, as well as likes and comments and watches for video posts.

2 RELATED WORK

There are several tools that are related to our project. The first one is Gapminder's visualization tools. The tool was bought from Trendalyzer, the information visualization software for animation of statistics built by Hans Rosling. The most prominent visualization in the tool is the bubble animated graph, which was showcased by Rosling, in his famous Ted Talk "The Best Stats You've Ever Seen". The tool we developed depicts most of Rosling's visualization tool¹.

The second related work is Upfluence. Upfluence provides services for marketers to advertise effectively on Instagram. They have tools that match a brand with its suitable influencers to reach the right target customers. Tools include a dashboard with various visualization tools that aggregate engagement metrics across the group of influencers related to the brand. They allow users to compare several influencers based on the type of content uploaded to their accounts².

Lastly, Phlanx is a software that measures an Instagram User's engagement rate. The application's free version accepts a username as an input, and return an overall engagement score/percentage³.

These online Instagram tools inspired us to build a visualization tool that will compare several influencers based on a certain engagement rate (comments, likes, watches per day). We opted to gather data on a daily basis rather than measure a user's total engagement.

For context on what drives real engagement on Instagram posts, Maria Hellberg's work in *Visual Brand Communication on Instagram: A study on consumer engagement*, is useful. Through a series of in-depth interviews with young, urban females she found that content that "engages personally and evokes emotions drives higher engagement." This can be extrapolated to other groups and points to why brands are so concerned with engagement: high engagement accounts are more likely to appeal to their followers emotions and thus have a stronger pull on their followers' buying decisions. As Hellberg points out, this is especially important on Instagram where users are less receptive to explicit advertisement and instead seek out content and creators that fit their tastes⁴.

The importance of engagement does not just stop with ensuring ads reach their target audiences. Brands also have to carefully vet accounts to avoid spending money on fake influencers or

influencers with inflated follower counts. CHEQ reported that “fake influencer marketing will cost advertisers \$1.3 billion in 2019.” This is over 15% of the reported \$8.5 billion total spent on the influencer advertising industry. Even if an accounts followers are real, many followers are so-called “zombie accounts” - accounts of users who are no longer on the platform or who have otherwise stopped using that account. These factors make it especially important for brands who are looking for a good return on investment to verify that influencers have genuine and consistent engagement⁵.

To understand what fake followers and fake influencers look like, mediakix’s article *How to Spot Fake Followers on Instagram* provides the common signs to look out for. Fake followers can be spotted by their lack of profile pictures and posts, gibberish usernames, and high following to follower ratios - A typical user account on Instagram usually has a 1:1 ratio between accounts they are following and followers while bot accounts have an average ratio of 41:1. On the other hand, fake influencers can be spotted using engagement rates, checking followers for bot accounts, and by looking for generic/unrelated comments on their posts. Fake influencers also do not typically have accounts on other social media, which is common for real influencer accounts⁶.

Of the aforementioned \$8.5 billion social media marketing industry, Instagram accounts for over \$1.7 billion spent in 2019. According to mediakix, this is expected to grow to \$2.3 billion in 2020. In their report, they also show that Instagram ranks as the most important channel for marketers with 89% of those surveyed saying that it is “strategically important to their influencer marketing strategy” and 69% saying that it will be the channel they will spend the most on in 2019. This provides context why our project will be focusing on Instagram, as it is currently the most important channel by a wide margin⁷.

In order to display our data in the most optimal way, we decided that an interactive visualization would be necessary. This would allow users to filter data to their specific needs and allowing them to focus on in certain data points to see exact numbers. For this, we found Brodeck et. al.’s work, which dives into interactive visualizations useful⁸.

3 PROJECT DESCRIPTION

Before building our tool, we built several data mining scripts that were responsible for collecting the data needed for our tool. The scripts were invoked automatically 7:30 pm Mountain Standard Time for 16 days in August and September, At that time, the local time in Oman was 5:30 am. This specific time was important, because we believed the lowest engagement changes would happen at that time, which would allow the crawler to collect the data for the duration it needed (roughly 10 minutes) without risking losing important data while crawling due to low activity rates at that time.

Once we had collected enough data, we performed several aggregation operations to find the total engagement rate across the 10 most recent posts only collected on that day. Once the aggregation was completed, we would save the final dataset in a table indexed by the name of the user. The table’s columns include: watches, comments, likes, and total followers.

The data crawled and analyzed is from predetermined user accounts on Instagram that vary in type of content (pictures, videos, news) and in follower size. The dataset included 28 user accounts. Additionally, we grouped accounts based on similar attributes. There are five groups in total:

- News Accounts (5 accounts)
- More Than 200k Male (2 accounts)
- More Than 200k Female (9 accounts)
- More Than 300k Male (10 accounts)
- Others (2 accounts)

Groups differ in gender and follower size. We hypothesized, before creating the tool, different genders might interest different viewers on Instagram. We wanted to explore if gender would play a role in the engagement rate, which will be discussed in the results section.

The Others group describes two accounts which are not run by professional influencers, but rather by athletes who use Instagram as a way to update their followers about their career developments. The News group includes users with a high frequency image uploads to their accounts. Their uploads address only recent news, and since there are multiple Instagram accounts with a high follower count of such nature, we believe it is more suitable to group them separately. The remaining groups are influencer accounts split by gender and follower count.

We then proceeded to build a visualization tool to show a group of Instagram users engagement rate per day. The tool includes two graphs. A bubble graph, and a line graph, and both are connected to each other. We also included several interactive tools, which include:

- **Slider:** Changes the date of the data displayed in the bubble graph
- **Tap Tool:** Clicking on a bubble displays the selected account’s engagement plotted against time on the line graph.
- **Hover Tool:** Hovering over a bubble or point on the line graph displays more information about that data point.
- **Selection Tool:** Click on the tabs at the top selects which type of engagement to display. Using this, users can select between comments, likes, and watches.

We wanted to use a bubble graph, just as the one used in Gapminder (with the slider), because it encodes several data attributes seamlessly and with appealing visuals to engage viewers. Additionally, with the set up mentioned above, we knew that the bubble graph would be visually appealing, because we tested Gapminder’s animated bubble graph before building ours using Bokeh. The bubble graph communicates:

- Each user’s follower count (bubble radius)
- Each user’s engagement rate (y-axis)
- Each user’s post counts (x-axis)
- The date of the shown data (slider and text in the background of the graph).
- Each user’s influencer group (color)

The graph itself encodes five data attributes. It achieves this with an organized and an uncongested view, which enables effective

portrayal of what the data is communicating to the viewer. In order to make the data more easily and quickly digestible, the x and y axis have been scaled logarithmically while the radius of bubbles is scaled by dividing by 150 to reduce their overall size.

We added a line graph because we wanted to provide a way for viewers to investigate a particular user's engagement rate over time, without manually tracking that information by moving the slider, and manually writing down the information they need, and then changing the slider again and memorizing/writing the information they see again. The line graph communicates:

- Engagement rate (y-axis)
- Time (x-axis)

The line graph is connected to the bubble graph through the tap tool. If a viewer taps on a bubble (an Instagram Influencer), then the line graph changes to display that particular account's information.

We then added two extra tools. The hover tool and the select tool. The hover tool allows viewers to find where a bubble, or a point in the line graph, is located according the y-axis and x-axis for accurate measurements. Additionally, the hover tool allows viewers to connect bubbles to their corresponding influencer usernames on Instagram. Finally, the select tool allows viewers to interchange between different engagement metrics. The hover tool can also be used on the line graph to view the day and exact engagement value on that day.

4 RESULTS

There are two types of results to report. The first is the use of the tool we developed, and the ease of comprehension of the information that is visualized. The second is the engagement rate of the influencers in our tool, the group as a whole, and how engagement rate fluctuated with time.

The animated bubble graph, with the colored bubbles, helped us understand easily each influencer's engagement rate and post counts based on their location on the graph. We were also able to compare influencers together with ease based on their location.

The different bubble sizes did communicate a difference in follower counts, however, because the values were scaled, it was difficult to accurately measure the differences. The scaled x-axis and y-axis also resulted in this challenge. The viewer (in this case, our team) were able to conclude with confidence that one influencer has a higher engagement rate, a higher post count, or a higher follower count, but the exact difference cannot be deduced.

The connected line graph eased the viewer in investigating a particular influencer further, but it also included scaled values, therefore, it produces the same challenge as the animated bubble graph.

During Thursday's Presentation, we asked our classmates for their feedback, specifically, on what can be improved and what is difficult. Everyone agreed that the scaled axis prevented them from accurately measuring the differences between different values, and that they can only conclude that one bubble has either a higher engagement, or higher post count, but without knowing the exact difference magnitude. They also said interpretability can

be further improved if the tool visualized subgroups separately rather than clustering all groups together.

On the other hand, our tool helped us specify trends easily. It was clear, with the help of our visualization tool, that the News Group had higher likes and comments rates than other groups, but they had similar watches rate. Additionally, the "More Than 200k W" group had consistent likes rate, but low watches, while "More Than 300k M" had more watches than any other group, but overall a low post count. Therefore, the tool can help users in inferring trends easily. Visualizing the groups separately would definitely be more helpful however, by unscaling the axis and also investing a smaller number of bubbles.

5 CONCLUSION

Advertising and marketing on social platforms, such as Instagram, empowers many people to establish businesses based on their likeness and to earn an income easily. To establish a credible Instagram account with a significant following however, some Instagram users purchase followers and/or engagements to appeal to companies with a lucrative account that would seem likely to expose a big following to the advertised services or products. This practice has caused an annual loss of \$1.3 billion in capital spent on advertisements and marketing campaigns on Instagram.

There are several tools available online (free/paid) that allows marketing agencies to find influencers in the desired markets with appealing metrics. However, these tools measure historical data, we wanted to create a tool that will measure daily engagement changes, to provide users with a more accurate and precise idea of which Instagram Influencer is more suitable for the proposed advertisement.

We built a visualization tool similar to the one created by Hans Rosling' animated bubble graph, because it can encode several data attributes seamlessly. A user can understand how bubbles (influencers) compare to each other on various metrics with ease, however, the user may not understand the exact magnitude of difference because the data is scaled.

The visualized data is scaled so that the bubbles can be near each other, otherwise, most of the bubbles will be close to each other in a small congested subsection of the graph, while a few might hover around in the top right corner. For future changes, our classmates suggested we allow a user to visualize different groups and all the groups together. Therefore, we will enable the user to investigate a certain group's data meticulously and accurately, or all the groups together but with the scaled data.

REFERENCES

- [1] Gapminder Staff. "About Gapminder." Gapminder. <https://www.gapminder.org/about-gapminder/history/>
- [2] Upfluence Staff. "Our in-house team of experts helps you create content that matters." Upfluence. <https://www.upfluence.com/services>
- [3] Phlanx. "Instagram Engagement Calculator." Phlanx. <https://phlanx.com/engagement-calculator>
- [4] Maria Hellberg. "Visual Brand Communication on Instagram: A study on consumer engagement," Master's thesis, Dept. Marketing, Hanken School of Economics, Helsinki, Finland, 2015.
- [5] CHEQ Staff, "THE ECONOMIC COST OF BAD ACTORS ON THE INTERNET," University of Baltimore, Baltimore, Mariland, 2019.

- [6] mediakix Staff. "HOW TO SPOT FAKE FOLLOWERS ON INSTAGRAM." mediakix.
<https://mediakix.com/blog/fake-followers-on-instagram/>
- [7] mediakix Staff. "INSTAGRAM INFLUENCER MARKETING IS A 1.7 BILLION DOLLAR INDUSTRY." mediakix.
<https://mediakix.com/blog/instagram-influencer-marketing-industry-size-how-big/>
- [8] Dominique Brodeck, Riccardo Mazza, and Denis Lalanne, "Interactive Visualization - A Survey," University of Applied Sciences, Switzerland, 2009.