TikTok's Users Analysis in the Context of US Politics

Advanced Topics in Computer and Networks Security

Author: Marco Bellò
Author: Alberto Morini
Supervisor: Alessandro Galeazzi

18/06/2024

Abstract

Introduction

TikTok is one of the most prominent social networks currently available, boasting over 100 million users in the USA alone [14]. Approximately 60% of young adults (aged 18-24) and nearly all children (aged 5-15) use TikTok daily [18]. This significant influence attracts not only advertisers but also politicians: the Democratic Party began organizing paid influencers as early as the 2020 United States election (for instance, presidential candidate Michael Bloomberg engaged in paid partnerships on social media [10]), while Republican and Conservative hype houses campaigned on behalf of political candidates. In Germany, the political party CSU invited influencers to political events and has recently started creating influencer-like social media posts on platforms such as TikTok [10].
This led to the creation of influencer-driven marketing firms, which now claim to control vast, immediately-deployable stables of small-scale influencers for various campaigns. These "nano" and "micro" influencers differ from the conventional image of an influencer: they are everyday people with captive, intimate social media audiences who represent demographics particularly appealing to U.S. political campaigns, such as Latinos in South Florida, Black voters in Atlanta, and college-educated women in the Rust Belt [12]. Political influencers often do not have an institutional background, in fact most of the times their notoriety and fame is platform-built [11].
Despite this political engagement, TikTok has attempted to market itself as a platform for everything but politics: since 2019 the company has banned paid political advertising, stating that "the nature of paid political ads is not something we believe fits the TikTok platform experience." Nevertheless, many creators regularly use the platform to disseminate political messages and viewpoints without disclosing whether the content is sponsored or not [3, 13].
Given these dynamics, there is considerable value in studying user interactions on the platform. This work aims to do so, focusing on the context of the 2024 U.S. political election. The study will examine user movements using a social graph, analyze user similarity through cosine similarity, infer political affinity, measure engagement, and visualize the impact of publishing a video on followers and comments.

1 Data Gathering

Considering the context (US elections) it is imperative to have users data divided between left and right leaning, so a group of *super-users* was selected using various sources. *Super-users* are defined as follows:

- Influencers: people whose notoriety is platform-built, without a background in institutions of entertainment;
- Politicians;
- Newspapers or news sites (i.e. The Washington Post).

The complete followers list, with relative sources (missing if selected arbitrarily by the authors), is the following:

- Left-Wing:
  - @aocinthehouse [15, 17]
  - @bernie [15, 17]
  - @teamkennedy2024 [15, 17]
  - @cnn [8, 2]
  - @thedailybeast [8, 2]
  - @democracynow.org [unofficial]
  - @huffpost [8, 2]
  - @vox [8, 2]
  - @newyorker [8, 2]
  - @nytimes [8, 2]
  - @washingtonpost [8, 2]
  - @msnbc [8, 2]
  - @harryjsisson [7]
  - @underthedesknews [7]
  - @chrisdmowrey [7]
  - @rynnstar [18]
  - @ginadivittorio [6]
  - @genzforchange [9]
  - @repbowman [13]

- Right-Wing:
  - @dailymail [8, 2]
  - @dailywire [8, 2]
  - @thesun [8, 2]
  - @notvictornieves [7]
  - @clarksonlawson [7]
  - @studentsforlife [4]
  - @babylonbee [4]
  - @real.benshapiro [11]
  - @theisabelbrown [3]
  - @alynicolee1126 [3]
  - @itsthemandrew [3]

Only five of each groups had been selected to data gathering, due to time constraints. All data has been gathered using TikTok's official APIs (https://developers.tiktok.com/doc/overview/)

1.1 TikTok's APIs

2 Privacy and Echo chambers

To help visualize all the gathered data, to see the presence of echo chambers and to infer political views of users, data regarding followers of super-users has been used to create a social network graph and a cosine similarity matrix.

2.1 Data Preparation

First of all a list of super-users's followers has been gathered via TikTok APIs in the form of a JSON file, with the following structure (we can ignore "videoID" and "videoDate"):

```
[
    {
        "influencer": "ith-super-user Name",
        "videoID": "videoID",
        "videoDate": "videoDate",
        "followerList": [
            "follower1",
            "follower2",
            "follower3",
            "follower4",
            "follower5",
            "follower6",
            "follower-k"
        ]
    }
]
```

For better understanding here follows a portion of the real JSON data used:

```
    {
        "influencer": "huffpost",
        "videoID": "7354208741996186911",
        "videoDate": "2024-04-05 11:46:08",
        "followerList": [
            "mathieucambet",
            "raphclp",
            "jennet153"
        ]
    },
    {
        "influencer": "huffpost",
        "videoID": "7354208741996186911",
        "videoDate": "2024-04-05 20:46:08",
        "followerList": [
            "doodlegolden0",
            "evanroyalaug",
            "cshanebritt",
            "kabed70"
        ]
    },
```

JSON data then gets imported to *Social_Graph.r* to analyze:

```r
data <- fromJSON(paste(readLines("data.json")))

left_influencer_names <-  # vector of strings with left
                          # super-user names
right_influencer_names <- # vector of strings with right
                          # super-user names

# data.frame used to calculate all the graphs and tables
full_total <- data.frame(
  influencer = data$influencer,
  followerList = I(data$followerList)
)

full_influencer_names <- union(left_influencer_names,
                               right_influencer_names)
```

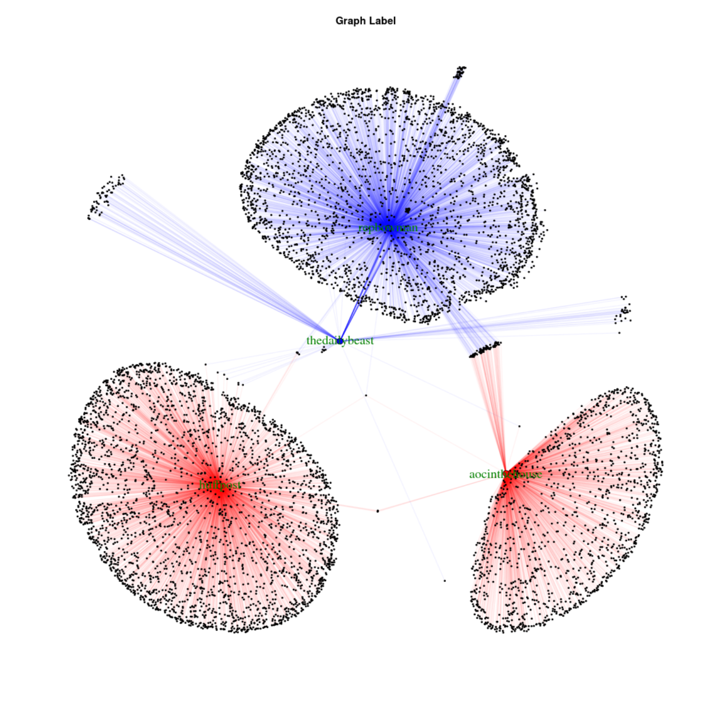
Now we have three data structures to work with: two vectors with all super-user's names and a `data.frame` that stores all super-users and their gathered followers, like so:

influencer	followerList
alynicolee1126	character(0)
alynicolee1126	c("elsee30", "renetheriot171")
alynicolee1126	character(0)
alynicolee1126	markbutler5636
alynicolee1126	c("frazierbklima", "electricmaster57", "ravirquaddnos", "gabbertour", "darrenricks87", "kory_sprinkle")
alynicolee1126	character(0)
alynicolee1126	c("addie.thorp", "caliwren", "truth.be.told", "idgafboutya501", "rowdy_chav", "thesalvatoresandmore._", "monahuluta", "recoveringhedonist", "pureblood_k")
alynicolee1126	c("_christian_l11", "robertdonald98", "mesa11964", "red_tacoma_4.0", "sladethekoolaid")

2.2 Social Graph

Broadly speaking, a social graph is a graph that represents social relations between entities, were vertices (or nodes) represents users and edges represents relations between such users. It is a model of representation of a social network, and has been referred to as "the global mapping of everybody and how they're related".
To give a brief example: if Alice and Bob are friends on a social network, in a social graph they would be represented each as a node, and there would be an edge between them.
The term was popularized at the Facebook F8 conference on May 24, 2007, when it was used to explain how the newly introduced Facebook Platform would take advantage of the relationships between individuals to offer a richer online experience [16].

Employing a social graph has numerous advantages: it helps visualize all gathered data (all users and their relations), visualize the presence of echo chambers and give insights to analyze the network as a whole.
The graph that follows clearly demonstrate that: considering the gathered data, users do not interact with each other outside their communities, thus forming cliques that can easily be interpreted as echo chambers:



The aforementioned graph is actually a mockUp without all the complete data: due to hardware constraints, graph compilation is impossible at the moment of writing (04/06/2024).

Super-users can be easily identified: their nodes are bigger than the rest, they are labeled and, most notably, the are at the center of their respective sub-graphs.

All black vertices represents followers of the super-users, unlabeled for improved readability, and the edge color represent the political orientation of the super-user to whom they are connected to(red for right-leaning and blue for left-leaning super-users).

Lets clarify: if *@user-Alice* follows super-user *@aocinthehouse* (offical account of Congress member Alexandria Ocasio-Cortez), which is classified as a left-leaning super-user, the edge connecting them will be blue.

Conversely, if *@user-Bob* follows super-user *@thesun* (official account of UK's tabloid The Sun), which is classified as a right-leaning super-user, the edge connecting them will be red.

2.3 Cosine Similarity

Having a graphical representation of a network is certainly valuable: pictures are not only more effortless to recognize and process than words, but also easier to recall. When words enter long-term memory they do so with a single code. Pictures, on the other hand, contain two codes: one visual and the other verbal, each stored in different places in the brain (Paivio). The dual-coding nature of images allows for two independent ways of accessing visual memories, increasing the odds of remembering at least one of them. Adding illustrations to text, researchers have concluded, aids comprehension and learning [5].

That being said, it is still advisable to measure numerically the similarity between users.

Broadly speaking, there are two types of similarity measures between nodes in a network: edge similarity, that provides the index of intersection of node parents (which are, of course among the neighbors) of the nodes being compared, and global structure similarity, that aims to evaluate the similarity between two nodes in the context of the whole network. Regarding the latter, Salton Index, Jaccard Index, and Sorensen Index always have good performance, while cosine similarity computational complexity is very high when applied to very large volumes of data [1]: when the data is dense, the structure-based indices (like Salton's) can perform competitively good as Cosine index, while with lower computational complexity. Furthermore, when the data is sparse, the structure-based indices give even better results than Cosine index [19].

N, q	CN	Sal	Jac	Sor	AA	RA	Cos	PCC
N=10, q=0.2	48	45	46	46	47	45	44	46
N=10, q=0.4	108	99	99	100	107	102	97	99
N=10, q=0.6	169	153	153	153	169	163	152	147
N=10, q=0.8	227	202	201	201	226	220	200	194
N=20, q=0.2	54	50	50	50	53	51	48	54
N=20, q=0.4	119	110	109	110	118	114	108	112
N=20, q=0.6	185	167	167	167	184	179	166	166
N=20, q=0.8	245	216	215	215	244	238	215	217
N=50, q=0.2	62	58	58	58	61	58	58	63
N=50, q=0.4	131	123	123	123	131	127	122	128
N=50, q=0.6	200	182	181	181	199	195	181	186
N=50, q=0.8	261	232	231	231	259	254	231	243

The above table shows values regarding precision in inferring similarity between users: Salton index (*Sal* column) seems to perform the best [19], therefore it has been used in this work to measure similarity between super-users.

Salton index formula is as follows:

$$s_{xy} = \frac{|\Gamma(x) \cap \Gamma(y)|}{\sqrt{k_x \times k_y}}.$$

All index values are calculated for each couple of super-users, and are shown in the following table where, as mentioned before, blue represent left-leaning and red represent right-leaning super users:

	thedailybeast	huffpost	aocinthehouse	repbowman	newyorker	alynicolee1126	babylonbee	real.benshapiro	clarksonlawson	notvictornieves
thedailybeast	1.0000000000	0.0088749628	0.003280670	0.0030531062	0.0028821675	0.003546902	0.0023624420	0.0008080255	0.0000000000	0.0007953238
huffpost	0.0088749628	1.0000000000	0.002755189	0.0019942806	0.0023532799	0.001654876	0.0003674144	0.0007540000	0.0000000000	0.0007421476
aocinthehouse	0.0032806704	0.0027551892	1.000000000	0.0728348570	0.0038275628	0.001712851	0.0003802860	0.0019510368	0.0000000000	0.0007681470
repbowman	0.0030531062	0.0019942806	0.072834857	1.0000000000	0.0019429436	0.001594039	0.0003539074	0.0010894217	0.0000000000	0.0007148644
newyorker	0.0028821675	0.0023532799	0.003827563	0.0019429436	1.0000000000	0.001880989	0.0008352317	0.0008570222	0.0000000000	0.0004217752
alynicolee1126	0.0035469022	0.0016548763	0.001712851	0.0015940390	0.0018809890	1.000000000	0.0020557341	0.0021093667	0.0000000000	0.0020762086
babylonbee	0.0023624420	0.0003674144	0.000380286	0.0003539074	0.0008352317	0.002055734	1.0000000000	0.0014049602	0.0000000000	0.0036876667
real.benshapiro	0.0008080255	0.0007540000	0.001951037	0.0010894217	0.0008570222	0.002109367	0.0014049602	1.0000000000	0.0006124765	0.0052028285
clarksonlawson	0.0000000000	0.0000000000	0.000000000	0.0000000000	0.0000000000	0.000000000	0.0000000000	0.0006124765	1.0000000000	0.0036170919
notvictornieves	0.0007953238	0.0007421476	0.000768147	0.0007148644	0.0004217752	0.002076209	0.0036876667	0.0052028285	0.0036170919	1.0000000000

Salton index values range between 0 and 1, and the diagonal of the matrix represented in the table shows all values equal to 1 because, of course, a super-user is always identical to itself.

The table confirms numerically what could be seen in the social graph: super-users share very few followers, which means that each community is a form of echo chamber.

2.4 Privacy Inference

3 Engagement

4 Content Impact

5 Conclusions

References

[1] A. L. R. Ahmad Rawashdeh. Similarity measure for social networks – a brief survey. In J. H. K. Michael Glass, editor, MAICS 2015 Modern AI and Cognitive Science Conference, volume 1353, pages 153–159, 2015.

[2] allsides. Media bias ratings, 2024.

[3] M. F. Becca Ricks, Brandi Geurkink. These are not political ads: How partisan influencers are evading tiktok's weak political ad policies, 2021.

[4] F. N. Brian Flood. Tiktok has silenced 11 pro-free speech organizations while 'muzzling conservatives,' study finds, 2022.

[5] P. Dewan. Words versus pictures: Leveraging the research on visual communication. Partnership: The Canadian Journal of Library and Information Practice and Research, 10, 06 2015.

[6] S. Lai. Campaigns pay influencers to carry their messages, skirting political ad rules, 2022.

[7] F. N. Matteo Cina. Conservative tiktok influencers claim platform discriminates against right-leaning voices, 2023.

[8] mediabiasfactcheck. Media bias fact check, 2024.

[9] A. Nafees. Political tiktok, explained, 2024.

[10] M. Riedl, C. Schwemmer, S. Ziewiecki, and L. M. Ross. The rise of political influencers—perspectives on a trend towards meaningful content. Frontiers in Communication, 6, 2021.

[11] M. J. Riedl, J. Lukito, and S. C. Woolley. Political influencers on social media: An introduction. Social Media + Society, 9(2):20563051231177938, 2023.

[12] J. L. Samuel Woolley, Martin Riedl. Tiktok has a political influencer problem targeted at gen z voters, 2023.

[13] N. Scola. Will politicians ban their best way of reaching young voters?, 2023.

[14] Statista. Countries with the largest tiktok audience as of april 2024, 2024.

[15] Wikipedia. 2024 democratic party presidential primaries, 2024.

[16] Wikipedia. Social graph, 2024.

[17] Yougov. The most famous democrats (q1 2024), 2024.

[18] J. Zeng and D. V. Kaye. From content moderation to visibility moderation: A case study of platform governance on tiktok. Policy & Internet, 14(1):79–95, 2022.

[19] Q.-M. Zhang, M. S. Shang, W. Zeng, Y. Chen, and L. Lü. Empirical comparison of local structural similarity indices for collaborative-filtering-based recommender systems. Physics Procedia, 3:1887–1896, 08 2010.