# WayPop Machine: A Wayback Machine to Investigate Popularity and Root Out Trolls

Tuğrulcan Elmas
*LSIR, Distributed Information Systems*
*EPFL*
Lausanne, Switzerland
tugrulcan.elmas@epfl.ch

Thomas Romain Ibanez
*LSIR, Distributed Information Systems*
*EPFL*
Lausanne, Switzerland
thomas.ibanez33@gmail.com

Alexandre Hutter
*LSIR, Distributed Information Systems*
*EPFL*
Lausanne, Switzerland
alexandre.hutter@gmail.com

Rebekah Overdorf
*Faculty of Law, Criminal Justice and Public Administration*
*University of Lausanne*
Lausanne, Switzerland
rebekah.overdorf@unil.ch

Karl Aberer
*LSIR, Distributed Information Systems*
*EPFL*
Lausanne, Switzerland
karl.aberer@epfl.ch

***Abstract*—Contrary to celebrities who owe their popularity online to their activity offline, malicious users such as trolls have to gain fame on social media through the social media itself. The exact reasons that a certain user has become popular are often obscure especially when the popularity was gained illicitly through means such as fake amplification of content. In this paper, we develop a methodology for uncovering why an account has become popular and present an open source tool that encapsulates this methodology. This tool aims to aid others in uncovering malicious accounts which have artificially gained many followers and to distinguish such accounts from those which gained followers and popularity honestly.**

***Index Terms*—trolls, osint, tool, data visualization, twitter, social media, social cybersecurity, online social networks**

## I. Introduction

In recent years we have seen a marked increase in disinformation including as part of a strategy of so-called hybrid warfare. Adversaries within and from abroad do not only spread misleading content from their own social media channels but also employ fake accounts, that are sometimes colloquially named "trolls". Trolls may impersonate other users, pose themselves as a citizen of the target country or act like an authority within the target country [1]. As opposed to celebrities offline, trolls have to gain fame through the platform itself. To do so, they can author viral tweets that attract public attention or use malicious strategies such as buying retweets [2] or followers [3].

Malicious strategies such as misleading repurposing [4] that employ deletions further complicate this picture. In such a scheme, a user buys a popular account (e.g., a meme page) and then purges all the tweets posted by the previous owner and changes the name and handle of the account, making it a different person entirely (e.g., a political troll) but leaving its followers intact. Users may also delete their tweets for other reasons, making them not accessible to observers. In each of these cases, an observer of the account may struggle to understand why it is popular, either because the tweets that made the account gain followers are so far in the past that they are not easily visible without delving deep into the user's timeline or they are deleted altogether and the account now represents a new identity.

Recent studies propose social media data for open source intelligence, OSINT, for tasks such as the detection of cybersecurity threats [5] and malicious actors [6]. While social media data is indeed useful for these tasks, it is also prone to manipulation by malicious attribute changes and deletions. External researchers looking at the up-to-date version of popular accounts cannot understand why they are popular, as social media platforms such as Twitter do not provide historical data (e.g., follower count of an account at a specific date) or deleted data (e.g., past names of an account). This posits a challenge when using social media for open-source intelligence. As such, OSINT researchers may need to rely on retrospective datasets that are not subject to change.

For Twitter, the Internet Archive provides an excellent dataset to facilitate open-source intelligence using retrospective social media data. The dataset "Twitter Stream Grab" stores the 1% random sample of all tweets and is freely available online [7]. However, processing this data is also a challenge: it consists of terabytes of raw data that is tedious to effectively store and parse.

In this paper, we present an OSINT tool, WayPop, that efficiently employs the Twitter Stream Grab. Our tool aims to aid others in uncovering trolls that have artificially gained many followers and to distinguish such accounts from those which gained followers and popularity honestly. It features the follower growth of users in the past, their viral tweets, their deleted tweets, and any change to their profiles. To the best of our knowledge, this paper is the first to focus on historic follower growth and on detecting the reason that an account is popular. WayPop is publicly available on GitHub at https://github.com/tugrulz/WayPop.

IEEE/ACM ASONAM 2022, November 10-13, 2022
978-1-6654-5661-6/22/$31.00 © 2022 IEEE

## II. Related Work

### A. Social Media as an OSINT Source

Social media is proven to be a vital OSINT source in tasks such as detecting cybersecurity threats and malicious actors. CyberTwitter is a perfect example of a system for the former task. The system proposed extracts intelligence about cybersecurity threats and vulnerabilities in real-time from social media using semantic web methods [5]. Similarly, Dionisio et al. use deep neural networks [8] and Khandpur et al. combine query expansion with event detection [9] for this task. Meanwhile, Tundis et al. use social media data to assess the validity of cyber threat intelligence sources [10].

OSINT using social media aids identification of malicious actors such as terrorists [11], trolls [1], disinformation spreading groups [12] as well as operations that otherwise may not be available to public knowledge such as astroturfing [13] and government censorship [14]. Past research proposes data analysis techniques and tools facilitating it for this task. For instance, Osteritter et al. propose a dynamic network analysis to characterize malicious actors (e.g., trolls) in public discourse on social media [6]. Similarly, our work proposes a temporal analysis of changes in follower counts and attributes of social media accounts to root out such malicious actors. Our contribution is a tool that facilitates such analysis.

### B. User Analysis Tools

In this work, we propose a tool that allows retrospective analysis of users. Several web applications provide a summary (e.g., number of tweets, followers, hashtags, and their interactions) using the recent tweets of a given user such as foller.me, accountanalysis.app, twitonomy.com, followerwonk, or of the authenticating user such as Twitter's own analytics tool. However, they use up-to-date statistics which are extracted using only the most recent tweets due to API limitations, contrary to our approach which uses a retrospective dataset. Additionally, our tool is the first to provide follower growth, change of attributes, and deletions to the best of our knowledge.

Some interfaces raise awareness for the authenticated users of the tool. For example, WDTKAM [15] shows its users the personal information they disclose on the web through their posts on Twitter. Gao et al. [16] proposed a tool to show its users their own biases in order to mitigate selective exposure and burst filter bubbles. Others [17], [18] provide information about suspicious profiles.

## III. System Overview

### A. Architecture

We used Django for the server-side (backend) programming and MongoDB for managing the database. It uses the NoSQL paradigm to store hundreds of gigabytes of data and efficient querying [19]. Although the underlying database has a compressed size of 115GB, the queries using user id, screen name, or tweet ids are all instantaneous. We used Bootstrap CSS for the front-end programming. Finally, we used D3 (Data-Driven Document) for the visualizations. It is a javascript library that allows for creating dynamic and interactive plots in the browser.

### B. Data Structure and Processing

As we argued previously, this tool requires a retrospective Twitter dataset, i.e., one collected in the past. To this end, we utilize archive.org's publicly available Twitter Stream Grab dataset [7]. Due to the enormous size of this dataset (4.67 terabytes) and ethical considerations (see §VII), we limit the tool to popular users with more than 5,000 followers, and who were recently active (in December 2020 in our case). Although we utilize this particular dataset, this tool is flexible enough to accommodate other retrospective datasets.

We further took two important processing steps to decrease the size. First, because each raw tweet object contains both the tweet and the posting user's information, we split each object into a tweet and a user object. Secondly, we only kept the relevant attributes for each type of object:

**Tweet Object:** Tweet id, creation date, user id of the poster, the tweet's content in textual form, its public metrics, its source, and, if applicable, its deletion time.

**User Object:** User id, account creation date, the current and historical screen names (profile handles), names, descriptions, followers, statuses, friends, and favorites counts.

Each user has multiple *data points* that indicate their past profile attributes (e.g., description) and the tweet associated with them. The past attributes are stored in a history array within the user object.

All processing was performed on a single machine using an AMD Ryzen9 3900x 12 core processor with 32 GB of memory. The final uncompressed sizes of the files for the users and the tweets were 71 GB and 200 GB, respectively. The entire process took 30 days.

## IV. Features

The tool has several features that provide historical information about a Twitter account. The end-user designates a Twitter account by its screen name or Twitter id. If the account exists in the database, the tool will redirect the user to the page where the account pane and the summary pane summarize the descriptive statistics of the user.

### A. Account Summary

The first page the end-user will be redirected to includes both the account pane and the account summary. The account pane shows up-to-date information about the user profile via Twitter API. If that is unavailable (i.e., the user has been suspended), it shows the most recent profile found in the retrospective dataset. The information provided includes the name, screen name, description, home page, location, account creation date, and whether the account is currently suspended, as seen in Fig. 1. The user can also export the given user's data, consisting of the tweet object and user object described in §III-B.

Next, we show the summary of the account. The summary includes the daily rhythm of the user, the number of tweets,

retweets and replies, the users the account has retweeted and mentioned, and the sources of the tweets (i.e., the app that is used to post the tweet). Compared to the accountanalysis.app, which shows statistics from the last 200 tweets of a given user, our tool shows the statistics using all the retrospective data. That way, the user can see statistics of old or deleted tweets.

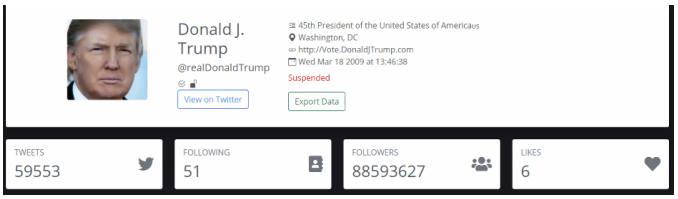


Fig. 1. The account pane shows the attributes of the Twitter account. If the account is still active, it shows the up-to-date information collected using Twitter API. Otherwise, it uses the most recent data in the dataset.

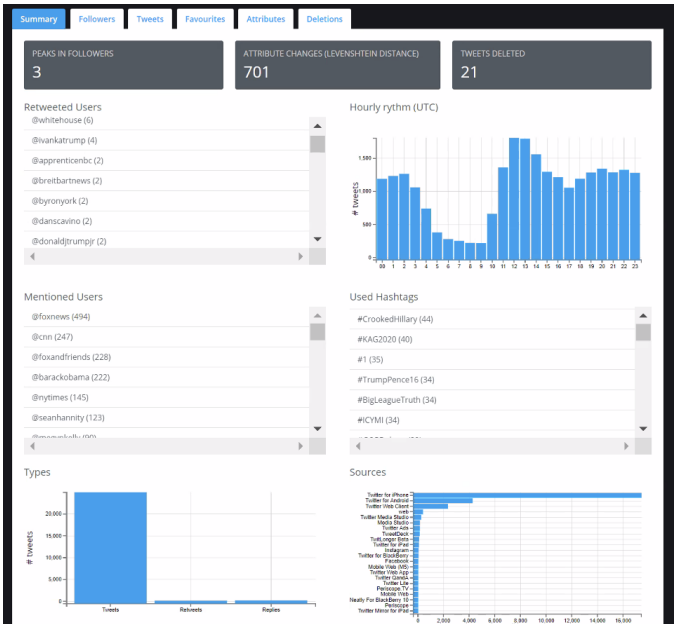


Fig. 2. The summary including statistics of a Twitter account (`@realDonaldTrump` in the example). The tool provides information on the daily rhythm of the user, the number of tweets, retweets, and replies, the users the account has retweeted and mentioned and the sources of the tweets (i.e. the app that is used to post the tweet) using the retrospective dataset.

### B. Follower Growth

The key to the tool is historical follower growth, which attempts to help us understand why a user is popular. It does so by visualizing the follower growth in a chart in which the x-axis denotes the time, and the y-axis denotes the follower count if the corresponding data point is available. Hovering on the data point shows the corresponding tweet for which the chart shows the time and follower count. Some data points have unusual follower growth due to corresponding tweets that went viral. We highlight these data points in the chart. We describe the methodology to select data points to highlight in the next section.

*1) Detecting peaks in follower growth:* We provide a method to help the end-user easily infer which tweets cause any unusual follower growth (i.e. viral tweets). We compute the data points for which there is a **peak** (i.e. unusual growth within a small time period) in the users' follower growth. To do this, we predict the follower count in the target time using linear interpolation and compare the prediction with the actual value. If the actual value is higher by a dedicated margin, we consider the value as a peak. If the actual value is not available, we use the linear interpolation of the closest point to the desired point.

Let $f$ be the logarithm of the follower count at time $x$. Consider the time $x-T$ and $x+T$ where $T$ equals an arbitrary duration.

The rate of change in follower counts between $x$ and $x-T$:

$$s(x) = \frac{f(x) - f(x-T)}{T}$$

We then compute the linear interpolation of the follower count $E(x)$ at time $x+T$ using s(x): $E(x) = T \cdot s(x) + f(x)$. Finally, we compare interpolated value with the actual value $f(x+T)$: $\Delta(x) = f(x+T) - E(x)$.

We consider $x + T$ a peak if $\Delta(x)$ is larger than $max(m1, \Delta(x)_{max} * m2)$ where $m1$ and $m2$ are both between 0 and 1. $m1$ is used to set a minimum follower growth for all users while $m2$ is to ensure highlighting only the significant follower growths for users with many unusual follower growths. After manual inspection of a few users, we set $T =$ 4 days, $m1$ to 0.02, and $m2$ to 0.3. For reference, consider a user who gains one follower every day. After reaching 200 followers, they became viral with a single tweet and gain 800 followers in a single day, making their follower count 1,000. At that time, $\Delta(x)_{max}$ will be 1.6 so $\Delta(x)_{max} * m2$ will be 0.48 when $m2 = 0.3$. In this case, the time when the user has accumulated 200 followers will be highlighted only if they gained 77 followers the day before.

### C. Tweets and Favorites

Tweets and Favorites panels complement the others by providing information on the historical growth of tweet and favorites counts respectively. Displayed to the right of the chart are the most engaged tweets in both panels. Engagements are the sum of retweets, quotes, and favorites. They facilitate providing a more complete picture of popularity.

### D. Change of Attributes

Our tool provides the given user's past attributes on this panel: name, screen name, and description field. It indicates the attribute changed and the time the change is first observed. We use Levenshtein distance to quantify the extent of the change. The visualization is in the form of a graph. The x-axis indicates the time and the y-axis is the sum of the Levenshtein distances between pairs of attributes. The user can hover on the bars indicating changes to see the previous and the new text of the attribute, as Fig. 6 shows.

### E. Deletions

On this panel, we show the number of deleted tweets of the user over time. Sudden peaks in the number of deleted tweets indicate purging behavior, similar to sudden drops in likes. Our tool also shows the text of the deleted tweets when the user hovers over the chart as seen in Fig. 3.

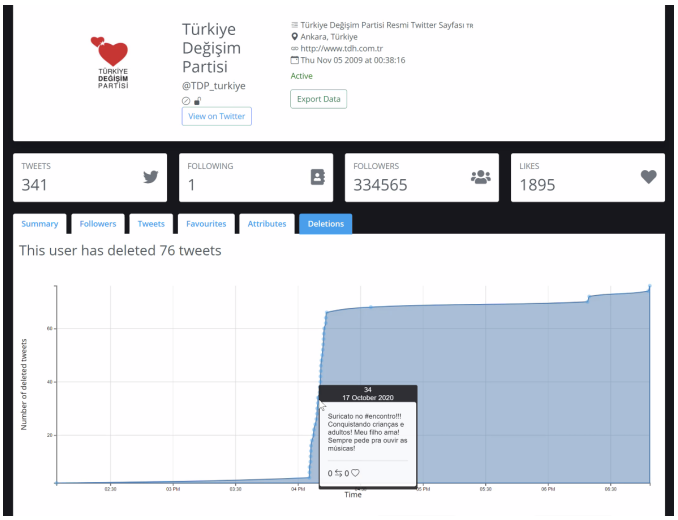


Fig. 3. The deletion statistics of the old account of Juliana Knust, now owned by a political party. 76 tweets were purged on October 17, 2020.

## V. Case Studies

We describe three case studies in which our tool facilitates analysis of follower growth and repurposing of accounts.

### A. MKBHD

Marques Brownlee is a tech Youtuber with 16 million subscribers. He also has 5.8 million followers on Twitter, and 3.8 million followers on Instagram. As he is popular on multiple platforms, it is unlikely that he gained popularity through malicious activity such as misleading repurposing. Thus, we present him as a negative example here. Our tool shows that `@MKHBD` has minimal changes to its description field, which means that does not use a repurposed account. Next, we analyze his follower growth per time, as shown in Fig. 4. The tool identifies three points of unusual growth. We read the user's most popular tweets and found that the user announced "giveaways" in these periods. For instance, on the 5th of December, 2016, which is the first peak point, he published a Youtube video "100 Oneplus 3T Giveaway!" and asked the viewers to follow his Twitter account, which earned him a high number of followers within a short period. Thus, our tool found that his popularity was partly due to growth strategies such as giveaways.

### B. Semih Mahcupadis

Semih Mahcupadis was a fictional character that only existed on Twitter. He claimed to be from the sociology department of Bogazici University although there was no such person on the university's website. DFRLAB found that he

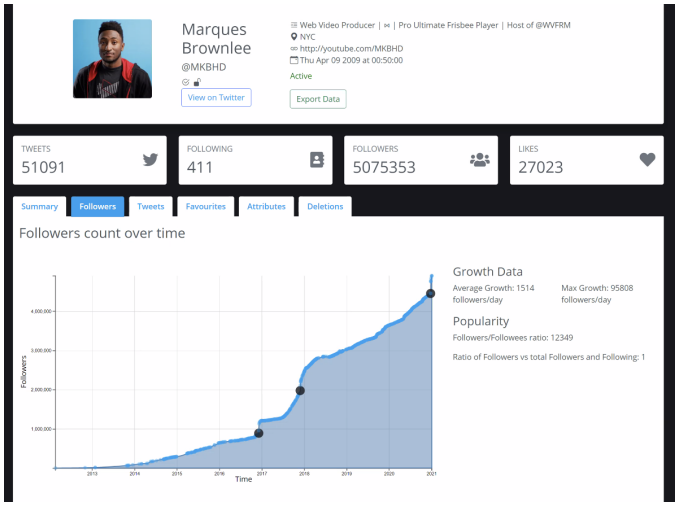


Fig. 4. The follower growth of the account `@MKBHD`. Our tool identified three points of unusual follower growth, all are giveaways in exchange for follows.

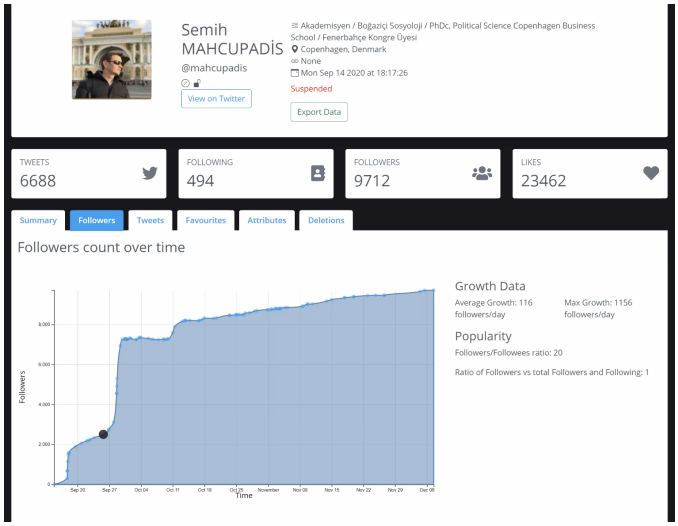


Fig. 5. The follower growth of the account `@mahcupadis`. It had an unusual growth after a viral tweet that spurred nationalist rhetoric.

spurred nationalist rhetoric by discrediting Yetvart Danzikyan, a Turkish-Armenian journalist, during Karabakh War in 2020. We analyzed this account using our tool. Although we could not find evidence of misleading repurposing, we found that the account grew its follower count from 2,200 to 7,800 on September 27, 2020, by a single viral tweet which became the focus of the report by DFRLAB [20]. In this tweet, he claims to be an Armenian and he calls on his fellow Armenian to stop being woke. The tweet received 4,000 retweets and 40,000 likes. The account was suspended after our analysis.

### C. TDP

Türkiye Değişim Partisi (Party for Change in Turkey), TDP, is a political party in Turkey that is found in 2020. It has 330,000 followers on Twitter and 108,000 followers on Instagram but does not have a Facebook page or a Youtube

channel. The disproportionate number of followers between the platforms and its low public support raises suspicions over its Twitter account. Additionally, the account is created in 2009, 11 years before the party's foundation. Media speculated that the account was likely compromised and sold [21].

We investigated the party's Twitter account using our tool. We found that the account originally belonged to Juliana Knust (`@jujuknust`), a Brazilian actress. We found that the account gained most of its followers between 2012-2016. However, as seen in Fig. 3, on the 17th of October, 2020, 76 tweets from the account are deleted within the same day, which is equal to all tweets from the account that is stored in our dataset. On the 18th of November, 2020, it changed its screen name to `@TDH_TR` and its name to "Türkiye Değişim Hareketi" as Fig. 6 shows. Those activities may signify that the account was hacked, stripped of its tweets, and sold to its current owner. It is still online.

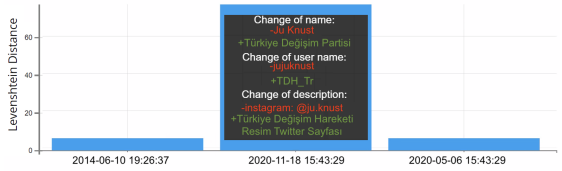


Fig. 6. The dramatic change of attributes of the old account of Juliana Knust, which is now owned by a political party.

## VI. Limitations

Our tool is biased towards popular users in terms of preciseness because those who are retweeted show up more often in the 1% sample, as retweets contribute to the sampling. Since we already limit the tool limited to popular users, this limitation does not introduce a critical problem, but still has to be acknowledged. Our tool is also limited to social media data and cannot account for external factors such as events, e.g., Donald Trump having an unusual follower growth when he officially became the president, or manipulation, e.g., users buying growing followers using bots. We advise researchers to consider those factors when using our tool.

## VII. Ethics

Our tool only uses the data provided by Twitter and the public data provided by the Internet Archive. Although we made the source code of the tool public, we do not share the data to stay in line with the Twitter TOS. We also realize that our tool could be misused to target regular Twitter users. Although we cannot completely eliminate this dual-use issue, we limit our tool to inspect only popular users. who had at least 5,000 followers. This threshold is borrowed from Twitter itself, which uses the same to reveal unhashed versions of the accounts involved in information operations. We also restrict our data to accounts active on the platform in order to respect users who have decided to no longer use the platform. The tool will be available and usable only as long as the Twitter Stream Grab is publicly available.

## VIII. Conclusion and Future Work

In this study, we present an OSINT tool that employs retrospective social media data to root out accounts that may be popular by malicious means. In future work, we plan to develop a methodology to automatically detect viral tweets and other follower growth strategies to be more cost-effective.

## References


[1] D. Kim, T. Graham, Z. Wan, and M.-A. Rizoiu, "Tracking the digital traces of russian trolls: Distinguishing the roles and strategy of trolls on twitter," *arXiv preprint arXiv:1901.05228*, 2019.

[2] T. Elmas, R. Overdorf, and K. Aberer, "Characterizing retweet bots: The case of black market accounts," in *Proceedings of the International AAAI Conference on Web and Social Media*, vol. 16, 2022, pp. 171–182.

[3] S. Cresci, R. Di Pietro, M. Petrocchi, A. Spognardi, and M. Tesconi, "Fame for sale: Efficient detection of fake twitter followers," *Decision Support Systems*, 2015.

[4] T. Elmas, R. Overdorf, Akgül, and K. Aberer, "Misleading repurposing on twitter," *arXiv preprint arXiv:2010.10600*, 2020.

[5] S. Mittal, P. K. Das, V. Mulwad, A. Joshi, and T. Finin, "Cybertwitter: Using twitter to generate alerts for cybersecurity threats and vulnerabilities," in *2016 IEEE/ACM International Conference on Advances in Social Networks Analysis and Mining (ASONAM)*. IEEE, 2016, pp. 860–867.

[6] L. Osterritter and K. M. Carley, "Conversations around organizational risk and insider threat," in *Proceedings of the 2021 IEEE/ACM International Conference on Advances in Social Networks Analysis and Mining*, 2021, pp. 613–621.

[7] A. Team, "The twitter stream grab. accessed on 2020-12-01," 2020.

[8] N. Dionísio, F. Alves, P. M. Ferreira, and A. Bessani, "Cyberthreat detection from twitter using deep neural networks," in *2019 international joint conference on neural networks (IJCNN)*. IEEE, 2019, pp. 1–8.

[9] R. P. Khandpur, T. Ji, S. Jan, G. Wang, C.-T. Lu, and N. Ramakrishnan, "Crowdsourcing cybersecurity: Cyber attack detection using social media," in *Proceedings of the 2017 ACM on Conference on Information and Knowledge Management*, 2017, pp. 1049–1057.

[10] A. Tundis, S. Ruppert, and M. Mühlhäuser, "A feature-driven method for automating the assessment of osint cyber threat sources," *Computers & Security*, vol. 113, p. 102576, 2022.

[11] M. C. Benigni, K. Joseph, and K. M. Carley, "Online extremism and the communities that sustain it: Detecting the isis supporting community on twitter," *PloS one*, vol. 12, no. 12, p. e0181405, 2017.

[12] T. Elmas, R. Overdorf, and K. Aberer, "Tactical reframing of online disinformation campaigns against the istanbul convention," *arXiv preprint arXiv:2105.13398*, 2021.

[13] T. Elmas, R. Overdorf, A. F. Özkalay, and K. Aberer, "Ephemeral astroturfing attacks: The case of fake twitter trends," in *2021 IEEE European Symposium on Security and Privacy (EuroS&P)*. IEEE, 2021, pp. 403–422.

[14] T. Elmas, R. Overdorf, and K. Aberer, "A dataset of state-censored tweets." in *ICWSM*, 2021, pp. 1009–1015.

[15] F. J. Gómez-Fernandez and F. Terroso-Sáenz, "Towards a web tool for the analysis of twitter profiling information," in *Intelligent Environments 2020*. IOS Press, 2020, pp. 391–399.

[16] M. Gao, H. J. Do, and W.-T. Fu, "Burst your bubble! an intelligent system for improving awareness of diverse social opinions," in *23rd International Conference on Intelligent User Interfaces*, 2018, pp. 371–383.

[17] R. Ram, Q. Kong, and M.-A. Rizoiu, "Birdspotter: A tool for analyzing and labeling twitter users," in *Proceedings of the 14th ACM International Conference on Web Search and Data Mining*, 2021, pp. 918–921.

[18] Q. Kong, R. Ram, and M.-A. Rizoiu, "A toolkit for analyzing and visualizing online users via reshare cascade modeling," *arXiv e-prints*, pp. arXiv–2006, 2020.

[19] L. P. ISSAC, "Sql vs nosql database differences explained with few example db," 2014. [Online]. Available: https://www.thegeekstuff.com/2014/01/sql-vs-nosql-db/

[20] DFRLab, "Suspicious twitter accounts claiming to be armenians from turkey spur nationalist rhetoric," *Medium*, 2021.

[21] Karar, "Tdp'nin twitter hesabının brezilyalı aktrise ait olduğu iddia edildi," *Karar*, 2020.