İçtimaiyat
Sosyal Bilimler Dergisi | Journal of Social Sciences
DergiPark AKADEMİK doi Foundation akademik intihal.net E-ISSN 2602-3377

Identifying Domain-specific Opinion Leaders in Twitter (X): An Optimized Approach*

Yazar(lar) / Author(s)
Nurullah Karaca
Phd Student, Yıldız Technical University, Cultural Studies Program, İstanbul-Türkiye
e-mail: nurullahkaraca@gmail.com.
Assoc. Prof. Dr. Onur Güneş Ayas
Yıldız Technical University, Sociology Department, İstanbul-Türkiye
e-posta: gunesayas@gmail.com.
(Sorumlu Yazar/Corresponding author)

Makale Bilgileri/Article İnformation
Tür-Type: Araştırma makalesi-Research article
Geliş tarihi-Date of submission: 01. 01. 2024
Kabul tarihi-Date of acceptance: 30. 04. 2024
Yayım tarihi-Date of publication: 31. 05. 2024

Hakemlik-Review
Hakem sayısı-Reviewers: İki Dış Hakem-Two External
Değerlendirme-Rewiev: Çift Taraflı Kör Hakemlik-Double-blind

Etik/Ethics
Etik beyan- Ethical statement: Bu çalışmanın hazırlanma sürecinde etik ilkelere uyulmuştur.
Benzerlik taraması- Plagiarism checks Yapıldı-İntihal.net-Yes-İntihal.net
Etik bildirim- Complaints
ictimaiyat@alparslan.edu.tr.
ictimaiyatdergi@gmail.com

Çıkar çatışması-Conflict of interest
Çıkar çatışması beyan edilmemiştir.
The Author(s) declare(s) that there is no conflict of interest

Finansman-Grant Support
Herhangi bir fon, hibe veya başka bir destek alınmamıştır.
No funds, grants, or other support was received.

Abstract

The role of opinion leaders on Twitter to analyze and understand the diffusion of information cannot be overlooked. The coming of the information age and the advent of social networking platforms have not eliminated opinion leadership but rather led to the emergence of its new forms. In this line, this paper deals with organically identifying opinion leaders on Twitter, based on the list feature. The method relies on the meta-data of Twitter lists, containing semantic cues to infer the topical expertise of its members. Based on the studies that have already shown the effectiveness of this method, this paper further illustrates in detail how the method can be employed flexibly to identify highly influential Twitter users in any specific domain. In this regard, the method can be adapted to different research questions, allowing researchers to apply it to suit their specific objectives and data. This paper also presents a novel approach as to how influential Twitter users identified through Twitter lists can be ranked. The ranking index proposed is attentive to both vertical (public perception and engagement) and horizontal (peer perception) dimensions of information diffusion.

Anahtar Kelimeler: Opinion leaders, Diffusion of innovation, Social networks, Twitter (X).

Twitter (X)'da Alana Özgü Kanaat Önderlerinin Belirlenmesi: İyileştirilmiş Bir Yaklaşım

Öz

The historical development, functions and current position of the mukhtar institution (mukhtarship) is an important topic for addressing the attitude of the traditional administrative approach to technology and changes in service delivery. In the first part of this study historical origins and the purpose of its establishment are examined. The second section discusses the concept of e-government and its role in service delivery processes. E-government is a digital transformation tool that enables governments to provide services to citizens faster and more effectively. These technological advancements enable the citizens to use the services which were provided by the mukhtarship institution before. Citizens can easily use many of the services offered by mukhtars online. In this respect, both time and financial savings are ensured. In the last section, in the face of developing technology a discussion on the future necessity of the mukhtar institution is being discussed. As a result, considering the developing technology and social needs, the necessity of transforming and modernizing the mukhtar institution is emphasized. This not only allows the state to provide better services to its citizens, but also helps it use resources more effectively. In this context, it is recommended that the headmen's offices in neighborhoods and villages be reviewed, and certain changes be made.

Keywords: Kanaat önderleri, Bilginin yayılımı, Sosyal ağlar, Twitter (X).

Atıf- Citation (APA)
Karaca, N., Ayas, O. G. (2024). Identifying domain-specific opinion leaders in Twitter (X): An Optimized approach. *İçtimaiyat,* 8(1), 65-81. https://doi.org/10.33709/ictimaiyat.1404626

* This work is derived from a dissertation in part-fulfilment of a PhD in Cultural Studies degree at the Yıldız Technical University, under the supervision of Onur Güneş Ayas. (Bu makale, Yıldız Teknik Üniversitesi Sosyal Bilimler Enstitüsü Kültürel Çalışmalar Doktora Programı kapsamında Doç. Dr. Onur Güneş Ayas'ın danışmanlığında yazılmakta olan doktora tezinden türetilmiştir.)

1. Introduction

According to the "We Are Social 2023" report, about 60% percent of the world population actively uses social networking platforms ("Digital 2023: Global Overview Report," 2023). The extensive reach and availability of these platforms make them valuable sources of information to analyze and understand for researchers across almost all disciplines in social sciences. One may claim that social networking does not fully reflect the complexities and nuances of everyday life. Nonetheless, it does not operate outside the realm of real life, however peculiar some of the characteristics of online behavior. Thus, analyzing individual and collective online behavior, and investigating the formation and diffusion of ideas in social networking platforms provides insights into social, cultural, and political dynamics.

Among social networking platforms, Twitter[a] stands out as being primarily an information network — a platform for the dissemination of information and public discussion (Kwak et al., 2010; Myers et al., 2014). The top Twitter management has long promoted the platform as the center of (online) public discussion, defining it “the global town square,” extending “all those capabilities of [the Greek] Agora” by providing “multidirectional [and] unfiltered” (Costolo, 2013) setting for communication “with greater accessibility and velocity” (Dorsey, 2018). Following his acquisition of the platform, Elon Musk similarly emphasized the role of the platform as “a common digital town square” and stressed its significance “to the future of civilization” (Musk, 2022). Likewise, while some commentators called for social media to be defined as critical infrastructure mainly due to the concerns related to disinformation, which they consider a threat to democracy (Heumann, 2018; Sharpe, 2020), the US government has already taken steps toward this direction (CISA, n.d.-a, n.d.-b).

While researchers are actively examining a multitude of social media-related topics, including its effects on mental health, relationships, information dissemination, activism, and privacy, there seems to be less of an emphasis on opinion leaders on Twitter. One reason appears to be that dealing with big data — accessing and analyzing it — may require a certain level of proficiency in data-related tools, concepts, and software skills. To this end, this article presents a step-by-step and detailed explanation as to how Twitter lists can be utilized to identify and rank domain-specific opinion leaders in the network.

2. Background

The reflections of digitalization and digital transformation in academic research are gradually gaining more visibility. The emergence of computational science and digital humanities are two examples to illustrate the phenomenon. Referred to as “computational turn,” this shift refers to a phenomenon beyond the infiltration of digital technologies when analyzing data. It is already affecting methods and approaches concerning the collection as well as the nature of data, including “the epistemologies and ontologies that underlie [research programs]” (Berry, 2011, p. 1). Grey asserts that we are on the threshold of an exploratory science (eScience), a fourth

[a] As of July 2023, Twitter rebranded as X and concurrently updated the familiar "tweet" label to the more universal “post.” However, despite the rebranding, we have opted to stick with the original name Twitter, and continue using the term "tweet," since our study was conducted prior to the change, they are still widely known as such.

scientific paradigm, after the computational, theoretical (models and generalizations), and empirical (descriptions of natural phenomena) paradigms (Hey et al., 2009, p. xviii).

The main drive behind this shift is the availability and the deluge of big data. This phenomenon appears to have led to “a new form of empiricism,” which posits that data can speak for itself (Kitchin, 2017, p. 30).

Proponents of this perspective herald the end of theory, marked by the emergence of data-driven science that replaces knowledge-driven science. Despite its appeal, Kitchin asserts that the fallacious reasoning of this stance undermines its validity. Above all, big data is prone to sampling bias due to its representational nature, that is, a sample of a larger reality it aims to capture. Besides, discovering patterns “is not the endpoint, but rather a starting point” (Kitchin, 2017, p. 31). In this regard, an analysis based on social theory and in-depth contextual understanding is essential, for data is generated within the scope of social, cultural, and political contexts. Further, making data intelligible is subject to human bias and framing (Kitchin, 2017, p. 31).

It should also be noted that while big data garners considerable attention, it is crucial to recognize that small data is likely to play a significant role, at least in the near future. Despite its limited scope compared to big data, small data has a long history and proven track record of development across many institutions with established methodologies and modes of analysis. Small data has long enabled researchers to explore in detail the "individual, nuanced and contextual stories" as well as "the varied, contextual, rational and irrational ways in which people interact and make sense of the world" (Kitchin, 2017, p. 34). Most likely, the integration of both forms of data will be prevalent in future studies.

Everything considered, it is important to acknowledge that both big data and small data come with their own set of strengths and limitations. In an attempt to define the characteristics of big data vis-à-vis small data, Laney outlined three key aspects. Commonly known as 3Vs, these aspects include volume (exponential growth of data), velocity (its generation in real-time), and variety (diverse arrays of structured and unstructured big data types) (as cited in Kitchin, 2017, p. 27). Other scholars have extended these aspects to include veracity (whether big data is accurate or reliable), virtue (ethical concerns), and value (whether big data increases our understanding) (Sloan & Quan-Haase, 2017, p. 6). Building upon those considerations, it becomes clear that big data presents a dual landscape of possibilities and challenges. Undoubtedly, big data promises to uncover novel patterns, correlations, and trends; however, an attentive consideration of certain conditions and limitations peculiar to big data is required.

When discussing social media big data, specifically focusing on Twitter data, Kitchin sheds light on several noteworthy challenges and restrictions. Among these are:

- *Sampling bias* (Users of a given social media platform may not equally represent segments of society.)
- *Fake data* (Fake accounts or hacking may lead to the presence of false information.)
- *Algorithmic manipulation* (Visibility and popularity of posts on social media are heavily influenced by ranking algorithms, rather than the interest of the users.)
- *API design* (The data extracted are structured by APIs, which means that not all relevant data may be retrieved.)

- *Disparity of data* (Different methods of API access may generate varying results, leading to disparities in the collected data.)
- *Fidelity* (Social media posts often represent a curated version of individuals' lives, views, and attitudes, rather than a true reflection.)

Given these intricate challenges outlined by Kitchin, researchers examining Twitter data should approach their work with heightened awareness and rigorous methodological scrutiny to ensure credible findings, ethical responsibility, and minimal distortion brought about by biases, fake data, algorithmic manipulations, and other potential contributing factors.

3. Opinion Leaders

Social media functions as a public sphere. Its influence permeates the digital realm, dramatically altering personal, societal, and organizational interactions. However, certain individuals stand out from the crowd, attracting substantial exposure, interaction, and followings. Depending on the context, designated with titles such as influencers, thought leaders, or opinion leaders. Being a vital component of the platforms, they have become one of the central foci of academic inquiries focused on analyzing the underlying social dynamics of the platforms. Illustratively, an investigation led by the Pew Research Center in 2021 demonstrated that the majority of tweets (97%) were produced by a quarter of Twitter users (McClain et al., 2021).

Even though its roots can be traced back to the sociology of Gabriel Tarde, the notion of opinion leadership was brought to the attention of researchers in the 1940s and has since become one of the key cornerstones in communication studies and diffusion research. Tarde aimed to outline "the laws of imitation," a concept that contemporary diffusion researchers now refer to as "adoption" (Rogers, 2003, pp. 40–42). Lazarsfeld et al.'s pioneering work on American voters underscored the significance of interpersonal relationships in influencing people's political choices, extending Tarde's insights into social dynamics and decision-making processes. In their study, they discovered that interpersonal relationships are a key component in people's (political) choices. Previous studies, such as the hypodermic needle model, emphasized the direct and significant influence of the mass media on shaping people's opinions. These studies recognized that people tend to passively consume media content. A key finding of Lazarsfeld et al.'s study in the context of communication research was that "ideas often *flow* from [mass media] to the opinion leaders and *from* them to the less active sections of the population" (Lazarsfeld et al., 1960, p. 151). According to them, what separates opinion leaders from non-opinion leaders is firstly they engage in the field they have knowledge or expertise are much more and secondly, they function as a bridge between the media of communications and the masses (Lazarsfeld et al., 1960, p. 152).

Rogers (2003) in his widely influential book, *Diffusion of Innovations,* recognized the significance of opinion leaders in the diffusion networks and programs of innovations. He defined opinion leadership as "the degree to which an individual can influence informally other individuals' attitudes or overt behavior in a desired way with relative frequency" (Rogers, 2003, p. 362). He recognized the significance of interpersonal relationships and the role of opinion leaders in the context of both communication process and diffusion networks, channels of communication through which ideas, innovations, and behaviors spread. Nonetheless, he challenged the two-step model, proposed by Lazarsfeld et al. (1960) mainly on two grounds.

Firstly, the communication process does not necessarily involve the two steps. It might involve a single step, wherein the mass media can exert its influence over an individual. In some cases, on the other hand, a communication process that involves several stages may take place (Rogers, 2003, p. 304). Another shortcoming of the model, according to him, was that it failed to acknowledge the role of diverse communication sources or channels at different stages in the process of communication and decision-making (Rogers, 2003, p. 305).

Opinion leadership differs among individuals, comprising four elements: scope, domain knowledge, behavior, and time. The influence spans locally or globally, with experts categorized as having single or multiple-field knowledge. Personalities and conveyed ideas determine if the behavior is constructive or destructive, while influence length marks another distinction - short vs. enduring (Bamakan et al., 2018, p. 205). In the context of human communication and dissemination of ideas and attitudes, and interpersonal relationships, one of whose aspects is opinion leadership, the concept of homophily was introduced to explain the nature of social influence in communication research. Homophily refers to the principle that people are most likely to interact with those who are alike (Rogers, 2003, p. 305). According to McPherson et al. (2001), the homophilous character of social influence entails the localization of information, be it cultural, behavioral, or material.

The concept of heterophily, on the other hand, represents a pattern of social interaction between individuals with dissimilar characteristics or attributes. Unlike homophily, it introduces possible complications of cognitive dissonance, misunderstandings, and frustration. Even so, according to Rogers, heterophilous communication has a significant potential for conveying information and diffusion of innovations among various segments of society. Novel ideas oftentimes enter a system through the members of higher status. While they tend to be more innovative and have more exposure to new ideas, they can hardly channel these novel ideas and innovations to other individuals (Rogers, 2003, p. 307). This is where opinion leaders come into the picture. To ensure the successful dissemination of novel ideas and innovations created by change agents to wider audiences, they leverage the power of heterophilous communication and tackle its challenges.

4. Twitter Lists

Several studies have shown that Twitter lists are valuable resources for inferring the fields of interest and topical expertise of Twitter users (Dongwoo et al., 2010; Ke et al., 2017; Sharma et al., 2012; Wu et al., 2011; Yamaguchi et al., 2011). Twitter allows its users to curate lists without enforcing any rules. The motivation behind curating Twitter lists, one can surmise, is to organize certain Twitter users based on a category, theme, or topic. The meta-data of the lists (title and description) usually contain information from which the “latent characteristics” of its members can be inferred (Dongwoo et al., 2010, p. 1). For instance, a Twitter list entitled “technology,” or includes a tech-related term in its meta-data signal that (at least some of) its members have tech-related backgrounds (Ke et al., 2017, p. 3). The frequency with which a user is a member of Twitter lists with similar meta-data increases the likelihood of his or her topical interest or expertise.

Making use of Twitter lists to identify domains of expertise can be more effective compared to other approaches. Twitter profile bios ideally contain information about individuals and organizations. However, relying on the profile information may be misleading, for it may not necessarily and accurately reflect the domain of expertise of a Twitter user (Sharma et al., 2012, p. 2). In some instances, users choose not to provide any information on their profile bio, leaving it blank. Tweets, on the other hand, are rich and the ultimate source from which topical expertise can be identified. However, since “tweeting” practice does not follow a strict guideline in terms of the language used (multiple styles) and the content created (multiple topics). Further, tweets mostly contain daily conversation, making it difficult and time-consuming to infer topical expertise (Sharma et al., 2012, p. 2).

The power of Twitter lists comes from its being a tool for social annotation. By curating Twitter lists, users classify (directly) and label (indirectly) the other users in the network, based on how they perceive them, rather than strictly adhering to prescribed labels. For the researcher, this means “crowdsourcing” the task inference of “who is who” on the network (Sharma et al., 2012, p. 2), saving both time and effort. Relying on the social annotation of Twitter users through the lists can be limited in certain instances, since the information (categories and labels) they provide cannot be always regarded as consistent. Then again, following users on Twitter and curating them in a Twitter list are two different actions. Based on observation, we can assert that curation is a more deliberate and selective process. By curating Twitter lists, users create timelines (feeds), ideally more focused than the two default timelines, “For you” (algorithmic timeline) and “Following.” As Wu et al. (2011) succinctly express, “the classification of users … effectively exploit the ‘wisdom of crowds’ … both in terms of their importance to the community (number of lists on which they appear), and also how they are perceived (e.g. news organization vs. celebrity, etc.)” (2011, p. 707). Thus, with the utilization of Twitter lists, researchers can harness the collective intelligence of the crowd, and identify opinion leaders and their public perception effectively.

5. Motivation

Although the above-cited studies have convincingly illustrated its effectiveness for inferring the topical expertise of Twitter users, few studies focused on Twitter lists can be used to identify individuals in a specific domain (of expertise) in detail. A study on the identification of scientists on Twitter, conducted by Ke et al. (2017), stands as a good exception, as it not only highlights the effectiveness but also provides practical guidance on the application of the approach. Still, Ke et al. (2017), employed the approach to identify scientists across 28 disciplines. Informed by this study, we utilize the method to identify topical experts in the domain of technology, i.e., the influential users on Turkish Twitter who engage in discussions related to technological progress.

Twitter serves as a platform for opinion leaders to gain significant following and social influence. Regarded as possessing expertise, credibility, or authority in particular fields or areas of interest, opinion leaders on Twitter frequently shape and steer conversations, contribute to certain narratives, and influence public discourse. Their influence is not limited to their followers, for the platform enables non-followers to engage with the content they create through retweets, quotes, replies, hashtags, and the timeline algorithm, thus potentially impacting public discourse on a broader scale. In this regard, Twitter opinion leaders are likely to engage not in the two-step flow but in a complicated and multi-step flow process (Park, 2013, p. 1642), and reach a wider audience compared to non-digital communication settings thanks to the architecture of the platform.

While many studies on Twitter in connection with homophily have suggested that the platform facilitates echo chambers in which users are exposed to opinions aligned with their own, a growing body of research is challenging this hypothesis.

In particular, Törnberg's (2022) study argues that social media platforms do not necessarily cause ideological polarization, but rather provide a setting where pre-existing divisions can manifest themselves more vividly. In line with this perspective, Colleoni et al.'s (2014) study draws attention to the multi-purpose uses of Twitter, highlighting the platform's capacity to support diverse modes of engagement, communication, and interaction. When it is viewed as a social medium, it appears to have homophily is a prevalent phenomenon in the platform. However, when it is viewed as a news medium, regardless of social ties, the degree of homophily appears to be much lower, while public discussion becomes much more visible (2014, p. 328). In the two scenarios, different dynamics can be observed regarding the influence of opinion leaders. In the first scenario, opinion leaders may be considered intermediaries to reinforce certain opinions through which they contribute to higher levels of homophily, fueling polarization. This is particularly relevant when the discussion revolves around political issues. In the second scenario, on the other hand, opinion leaders may act as translators of certain trends and developments.

6. Identifying Turkish Technology Opinion Leaders on Twitter

Several approaches have been developed to identify opinion leaders in social networks, both non-digital and digital environments, mainly for marketing and academic purposes. These approaches range from descriptive methods, such as observations, surveys, and interviews, to machine learning algorithms. While descriptive/conventional approaches have the advantage of simplicity, they tend to be ineffective in big social networks and pose risks such as experience bias (Bamakan et al., 2018). On the other hand, the bulk of non-conventional approaches on the other hand require a good working knowledge of statistics and probability, and in some cases solid foundation of software development skills. This could be one reason why novel (digital) research methods remain relatively uncommon in social sciences. A notable exception appears to be social network analysis. A multitude of research has used the approach to study and analyze the relationships, patterns, interactions as well as structures of communication, detecting communities and information diffusion networks. Simply put, social network analysis attempts to discover important nodes in a social graph through centrality measures.

In this study, an optimized approach has been employed to identify opinion leaders based on Twitter lists. Twitter allows the curation of user-generated lists, allowing the content generated by members of these lists to be displayed in a separate timeline. The metadata (title and description) of Twitter lists includes information that gives valuable insight into the topical expertise of members added to them. Making use of this information has great potential for researchers to infer who-is-who on Twitter. Developed by Sharma et al. this approach offers a "more accurate and comprehensive characterization of Twitter users" compared to the other approaches such as the profile information created by Twitter users themselves (2012, p. 533).

The metadata of Twitter lists allows the researchers to acquire valuable information indirectly by inferring the area of expertise or interests of users on the platform. It allows arguably a more informative and comprehensive method of inference than relying solely on the profile bios. However, Twitter lists also have certain shortcomings. Even though the metadata of a Twitter list gives valuable insight, the topicality and expertise of the users added do not necessarily match. One reason is that one's interest does not always entail his or her expertise in the field. While they show genuine interest in the field they express their opinions to their audiences, ideally, these opinions often reflect in extensive knowledge. They articulate complex issues and provide insights into the workings of a field more clearly and understandably for the public. Their interest is reminiscent that of academic ones and has a different nature compared to that of the general public. However, this is not to be confused with their presentation of the content they create. In this regard, the style and the language manifest in their content are greatly accessible to the general public, unlike that of the academician. That is why, the data collected via Twitter lists should be the initial step of identifying the influential users, namely, opinion leaders.

6.1. Data Collection

Through its API (Application Programming Interface), Twitter enables researchers, among many other professions, to retrieve and analyze Twitter data. In this study, mainly *List*, *User*, and *Tweet* endpoints are used to access Twitter data (X Developers, n.d.-c). It is important to note that the API is subject to changes and modifications and for that reason, it's essential to stay updated with any modifications or adjustments made by the platform. Although this study has utilized both the first and second versions of the API, it is worth noting that the first version is planned to be discontinued. During this research project, API access was acquired via the Academic Research program (now defunct) provided by Twitter. To interact with Twitter API and to make use of Twitter data more efficiently, several software tools and libraries have also been used during this study, most of which are built on top of the Python language (Python Software Foundation, 2021). The two main libraries[b] used are Tweepy (Harmon & other contributors, 2021) and Pandas (The pandas development team, 2010/2021). While Tweepy facilitates a more effortless and convenient way to access the Twitter API, Pandas is used to manage data efficiently. For the retrieval of the tweet data another Python library, *twarc* (Documenting the Now Project, 2013/2021) was utilized.

6.2. List-based Identification of Technology-related Turkish Twitter Accounts

In this section, we outline the processes of identifying eligible candidates, providing the rationale behind each decision. The first stage of the process involves defining certain constraints. By defining and implementing specific constraints, the results obtained through the API will be refined. What follows is a detailed explanation of the steps as to how Twitter lists are utilized to identify a set of influential Turkish Twitter accounts, which will be further refined to obtain a more focused and representative sample in the next sub-section.

6.2.1. Defining the Constraints

At the outset of the data collection process, it is of utmost significance to have adequately defined constraints that are going to determine the course of action in the process. The

[b] In the context of computing, a library is a collection of pre-written code that is served to be used in software projects. A library often contains commonly needed functions or routines and is called from the other piece of software to extend its functionality. A crucial component of modern software development, libraries provide sets of commonly needed functions to be reused across many projects (“Library (Computing),” 2023).

constraints used in this study are as follows. It should be noted that each research is likely to have unique constraints.

1. *Twitter lists that only contain the keywords defined* (*"tekno" and "tech"*) *in their title are considered.* The chosen keywords keyword enable the retrieval of lists containing not just "teknoloji" or "technology" but also variations such as "tekno-bilim," "teknolojik." The number of keywords used can be adjusted according to the research needs.
2. *Twitter users with fewer than 100,000 followers are to be filtered out.* This constraint is imposed to decrease the volume of data, and, by extension, the manual workload in the step of expanding the seed set. It is also considered appropriate for the intention to identify potentially the most influential users, rather than all the Twitter users in the specified domain.
3. *Only the Turkish-tweeting users who are creating content in the domain of technology are to be taken into consideration.*

6.2.2. Setting the Seed Set

As part of the preparation for the process, an initial (sample) set of individuals was identified. The individuals in the initial seed set were handpicked based on their representativeness of the broader population that is sought to be obtained in the end.

Although all the members of the set were selected through observation, they were carefully chosen to conform to the above-mentioned constraints. At this point, it should be noted that all the data retrieved is from November 2021, unless stated otherwise.

Table 1: The initial seed set of influential Twitter Users

Name	Twitter Handle	Followers
M. Serdar Kuzuloğlu	mserdark	1.35M
Cem Say	say_cem	251K
Serkan İnci	srkninci	217K
Alphan Manas	alphanmanas	120K
Sedat Kapanoğlu	esesci	117K

Although the size of the initial seed set is not likely to affect the end result, the process will eventually branch out (snowball) to discover other similar Twitter users. Nonetheless, choosing at least one influential Twitter user who meets the third criterion assures the method to work. The rationale behind creating a sample set containing multiple influential Twitter users is to increase the likelihood of retrieving more Twitter users from Twitter lists. Besides, the popularity of the members of the initial seed set most likely guarantees firstly that a good number of Twitter lists will be retrieved in the early phases of the process and secondly that the Twitter lists retrieved will have many other users in the same field/domain, and that the iterative process will be completed sooner than later.

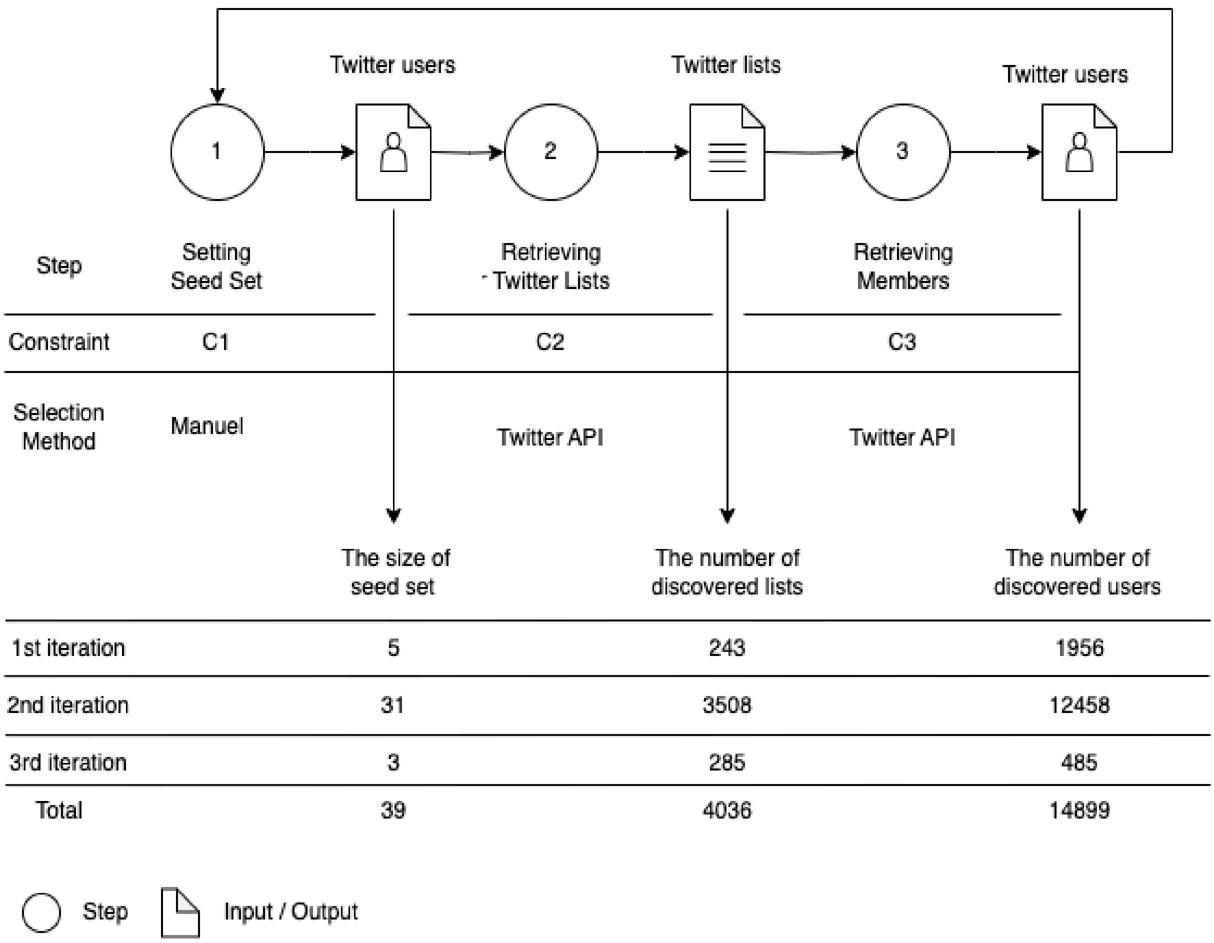

Figure 1: Workflow to identify opinion leaders

6.2.3. Retrieving Lists

Having the constraints defined and the initial seed set determined, first, all the Twitter lists to which the members of the initial seed were retrieved via "lists/memberships"[c] from the Twitter API. The retrieved data (the collection of Twitter lists) are stored in the same location. Next, members (Twitter users added to these Twitter lists) of the retrieved Twitter lists are iteratively collected. In this step, certain conditions are enforced. Twitter list retrieved is first checked whether its owner's account is publicly accessible, for the members of the Twitter lists whose owners have private accounts cannot be retrieved. Then, whether the lists retrieved are already stored is checked in order not to allow duplications. Finally, whether the title or the description of a list matches with the predetermined keywords (the first criterion) is checked. Provided that the Twitter lists that match the last condition are stored.

[c] See Twitter API documentation: (X Developers, n.d.-a) See Tweepy documentation: (Harmon & other contributors, 2009)

6. 2. 4. Retrieving Members

The last step of the process deals with the retrieval of the members (Twitter users) of the Twitter lists created in the previous step. The endpoint used in this step is “lists/members."[d] All the Twitter lists collected are iterated to retrieve unique users added to these lists provided that the user has more than 100k followers (the second constraint). The iteration continues until all the collection is exhausted. The decision tree for this subprocess is quite simple for there are only two conditions to check: 1) whether the Twitter user has more than 100k followers; and 2) whether the user data is already stored. If the user data is not stored, it is stored with additional information, frequency value. This value is not retrieved through the API but is calculated during the process to be used later in the ranking/scoring part. In order to calculate the frequency value (how many times a Twitter user appears in Twitter lists retrieved in the process), the first time a user, as a member of the lists, is returned, it is stored with a frequency value of 1. If a user is already stored (which means the user is added to multiple lists), the frequency value is incremented by 1.

6.2. 5. Expanding the Seed Set

Up to this point, two different data were programmatically obtained from the Twitter API: 1) A collection of Twitter lists (based on the first criterion) to which the members of the initial seed set are added. 2) A collection of Twitter users (based on the second criterion) from the first data (the collection of Twitter lists).

In this step, the second data, the collection of influential Twitter users, is manually inspected whether they meet the third criterion (The influential users who are creating content in the domain of technology). While the members of the initial seed set were considered epitomic examples of influencers/opinion leaders in the domain of technology, not all the members of this collection can be considered as such. In this regard, the aim of this step was twofold: 1) to identify the other influencers/opinion leaders; 2) to expand the seed set with these new Twitter users to reach more Twitter lists, and by extension, to identify more users. To this end, the second data (the collection of Twitter users) were inspected manually. Manual inspection mostly relied upon the descriptions of user accounts, which is commonly known as the profile bio. Often, the personal timelines are visited to gain more information, since Twitter profile bios do not necessarily yield information as to the interest or expertise of a Twitter user, and might even be blank. On the other side, the public perception of an opinion leader holds greater significance than how they define themselves.

Having inspected the collection of Twitter users, 31 more influential Twitter users were determined. Besides, the initial seed set was expanded by these Twitter users. At the same time, a new set from the newly discovered users was created to be used in the next iteration of the process to branch out new Twitter lists. The rationale to exclude the initial seed set in the second iteration of the process was 1) because all the Twitter lists to which the members of the initial seed set are already retrieved and stored; 2) to speed up the process by minimizing API requests due to the API rate limits.

---

[d] See Twitter API documentation: (X Developers, n.d.-b). See Tweepy documentation: (Harmon & other contributors, 2009).

The cycle (setting seed set, retrieving of Twitter lists, retrieving of Twitter users) explained above was repeated three times. In the third iteration, the number of Twitter users and lists decreased dramatically, which is taken as an indication that the data saturation point was reached.

Table 2: The number of identified Twitter users

Cycles	Seed Set	Twitter Lists	Twitter Users
1st cycle	5	243	1956
2nd cycle	31	3508	12458
3rd cycle	3	285	485
Total	**39**	**4036**	**14899**

In the course of the entire process, a total of 4036 Twitter lists that matched the first criterion were discovered. These Twitter lists contained 14,899 unique Twitter Users, 39 of which were manually selected based on the third criterion. These influential Turkish Twitter users were considered to be potentially shaping people's opinions on the domain of technology. Since the selection process was conducted manually, it should be noted that it was subject to certain biases or oversights. Many influential Turkish Twitter users were eliminated for it was observed that the content they produce was nuanced, that is, targeting not the general public, but mostly specific communities such as fintech, cryptocurrencies, digital marketing, etc. Since the aim is to identify opinion leaders that potentially influence the Turkish general public, those who were not creating content in the Turkish language were also filtered out. The final list of the selected influential Twitter users, comprised of 28 people, sorted by the list frequency, is given below:

Table 3: Overview of the Identified Twitter Users

	Name	Twitter Handle	Followers	List Frequency
1	M. Serdar Kuzuloğlu	mserdark	1.353.882	174
2	Tansu Yeğen	TansuYegen	445.911	101
3	Hakkı Alkan	hakki_alkan	100.815	84
4	Burak Buyukdemir	burakbuyukdemir	169.767	50
5	Sina Afra	SinaAfra	415.468	46
6	Selçuk Bayraktar	Selcuk	1.781.461	44
7	Alphan Manas	alphanmanas	120.331	41
8	Burak Bayburtlu	burak	191.136	37
9	Mustafa Varank	varank	790.757	32
10	Cem Say	say_cem	251.780	31

11	Sedat Kapanoğlu	esesci	117.267	30
12	Haluk Bayraktar	haluk	325.197	24
13	Tevfik Uyar	tevfik_uyar	105.345	24
14	Özgür Alaz	ozguralaz	229.756	23
15	Nevzat Aydın	zagortenay76	811.331	19
16	Nazım Salur	NazimSalur	100.244	15
17	Ismail Demir	IsmailDemirSSB	266.189	14
18	Yusuf Akbaba	ssysfakb	111.469	14
19	Said Ercan	saidercan	1.039.093	11
20	Dr. Taylan Yıldız	taylanyildiz	258.516	10
21	Levent Erden	leventerden	322.466	9
22	Cenk Sidar	cenksidar	103.845	9
23	Orhan Bursalı	ORHANBURSALI	247.243	7
24	Serkan İnci	srkninci	217.710	5
25	Vahap Eren	VahapErenTR	139.291	4
26	Immanuel Tolstoyevski (Fularsız Entellik)	imTolstoyevski	116.395	2
27	Kıvanç Özbilgiç	kivancozbilgic	210.335	2
28	Mesut Sevgili	mesut_sevgili	271.591	1

6.3. Measuring the Impact: An Updated Approach

The most manifest indicator of a Twitter user's popularity is the number of followers. Yet, relying solely on this data is most likely not efficacious to determine a Twitter user's impact/influence on the network. To measure impact, most social media analytics companies appear to be using a variety of formulas called engagement rate. While some of these formulations take into consideration the number of followers, what is common in all of them is that these formulas advocate that the engagement metrics, "the number of favorites, retweets, and mentions [a Twitter user's] tweets are generating" (“Twitter Engagement Calculator,” n.d.; “Twitter Engagement Metric,” n.d.), represent a better indicator than that of the number of followers. A variant of engagement rate, also called as “public engagement rate” (“X (Twitter) Engagement Rate Benchmark,” n.d.) is calculated by dividing the sum of retweets and likes by the number of followers.

For this study, however, a ranking index was developed to measure the impact of the selected influential Twitter users. Although the ranking index was developed to take into consideration the variables of the engagement rate, it offers a more advanced but uncomplicated approach to measure the impact based on which Twitter users can be ranked. The variables taken into consideration for the ranking index are as follows.

1. The number of followers
2. The number of engagements (likes, retweets, mentions and comments)
3. The frequency of occurrence in Twitter lists
4. In-group popularity (who is following whom in the group)

As the first two of these variables are widely used in the engagement rate formulations to measure the impact of a Twitter user, the index developed employs two more variables, which were considered as they are as valuable as the first two. The frequency of occurrence (the third variable) of a Twitter user in Twitter lists is valuable for the addition of a Twitter user into a Twitter list is a more conscious and deliberate act when compared to the following behavior. Although the data collected indicate that many Twitter users who cannot be considered influencers in the field of technology have been added to the Twitter lists promoted (in its title or description) as related to the field of technology, that does not comprise the idea Twitter lists are a great resource for identifying Twitter users in a given field. Besides, Twitter lists indirectly classify, albeit not strictly, Twitter users, the frequency value indicates a different kind of popularity. Then again, we can surmise that Twitter's lists feature is mostly used by tech-savvy users.

As for the in-group popularity (the fourth variable), it was considered that the Twitter user who is producing content in a given field is most likely being followed by other Twitter users in the same field. Although these might not be true in every case the data collected might not have absolute accuracy and there is more than one way to follow Twitter users (e.g., Twitter lists), it was observed that the ranking index produced meaningful results.

It should be emphasized that all the formulations proposed to calculate the influence of a Twitter user are based on publicly available data (through the Twitter API), which is relatively limited. In this respect, such formulations are far from being scientifically sound-proof. The data collected on social networking platforms might not accurately reflect or represent individual motivations, preferences, and attitudes. Be that as it may, it was convincing that the index generated meaningful and measurable insight from the data collected in the process.

The top 10 Turkish Twitter users who exhibit greater impact factors according to the ranking index developed are given below. These individuals can be considered more influential than the rest of the selected group of influential Twitter users in shaping the opinion of the Turkish general public on the domain of technology.

Table 3: The Ranking of Identified Twitter Users (Opinion Leaders)

	Name	Ranked	List Frequency	Followers	Engagement	In-Group Popularity
1	M. Serdar Kuzuloğlu	1	1	2	9	1
2	Selçuk Bayraktar	2	6	1	1	4
3	Tansu YEĞEN	3	2	6	2	5
4	Mustafa Varank	4	9	5	5	4

5	Nevzat Aydın	5	15	4	7	4
6	Cem Say	6	10	13	4	5
7	Sina Afra	7	5	7	22	4
8	Hakkı Alkan	8	3	27	16	6
9	Alphan Manas	9	7	21	17	3
10	Nazım Salur	10	16	28	21	2

Ranked by the impact factor calculated by the index, the table also provides sub-rankings for each variable for a detailed comparison. The sub-ranking indicates the relative strengths and weaknesses of each opinion leader in relation to the other variables encompassed by the index. The index basically calculates the overall ranking of each variable. It should be noted that all of them are equally weighted.

The index indicates that the first 4 opinion leaders consistently perform well in each sub-ranking. Among the top 10 in each sub-ranking, these opinion leaders appear to have greater reach or respect among Turkish Twitter users, ranging from average people to tech-savvy users and the other opinion leaders identified during this study. In terms of list frequency, on the other hand, 8 opinion leaders were ranked among the top 10. This demonstrates that the list frequency variable correlates closely with the final list, regardless of the rankings. Although this does not prove the efficiency of the ranking index, it can be surmised that the list frequency is a potentially strong indicator to assess to understand the popularity of Twitter users. However, a dedicated and comparative study is needed for a comprehensive evaluation. In terms of the number of followers and engagements, 6 out of the top 10 opinion leaders were already in the top 10 in both categories. However, the other 4 opinion leaders were picked out by the index for they ranked high in the category of the list frequency and in-group popularity, although they have significantly fewer followers and lower engagement performance. Consequently, contrary to conventional expectations, it was observed from the result of the index that there is a highly inverse relationship between variables (followers and engagement vis-à-vis list frequency and in-group popularity). The variables of list frequency and in-group popularity appear to balance the variables of followers and engagement, the two key values used to calculate engagement rate. At this point, it is crucial to note that the engagement rate formulation yields divergent outcomes than that of the index. For instance, the highest-ranked opinion leader in the index (Serdar Kuzuloğlu) falls in the 24th position in this formulation.[e] This is because Kuzuloğlu has a significant following, yet his tweets receive limited engagement compared to the other opinion leaders among 28 influential Turkish Twitter users picked in the previous part. Such a finding can be interpreted as a limitation, if not ineffectiveness, of the engagement rate method.

It is also noteworthy that the in-group popularity variable, which indicates the level of recognition and respect in the domain, and the final result of the index completely correlates. That is, irrespective of their rankings, the top 10 performers in both in-group sub-ranking and

[e] The formula is as follows: total number of engagements / (total) number of tweets / total number of followers * 100 (“Twitter Engagement Calculator,” n.d.).

the index are identical. From the inverse relationship between the two groups of variables mentioned above and the complete correlation of in-group popularity with the index, it can be inferred that popularity does not necessarily entail recognition and respect among opinion leaders (and likely tech-savvy Twitter users who create Twitter lists). Since opinion leaders are in a unique position to diffuse innovations and new ideas, they are expected to possess a combination of persuasive communication skills (rhetoric) and substantial domain-specific knowledge (expertise) to interpret trends and developments, ideally in a balanced manner. The outcome of the index is considered satisfactory in this aspect.

7. Conclusion

As we navigate the ever-evolving landscape of social media, comprehending the role of opinion leaders becomes increasingly crucial. This research sheds light on how seemingly simple tools like Twitter lists can effectively unveil hidden networks of influence. Recognizing and engaging with these individuals is essential not only for brands but also for researchers who seek to understand the social and cultural dynamics of the digital age, where the transformation of public space and discussion is taking place.

Unlocking the latent power of Twitter lists and a metrics-backed ranking algorithm, this study demonstrated an efficient technique for identifying domain-specific opinion leaders. By leveraging the unique insights provided by Twitter lists and a data-driven ranking index, this research offers a scalable method. The approach proves invaluable for refining research strategies and conducting deeper analyses of public opinion in the digital age. The method shows promise for further exploration across a variety of social media platforms and within different domains, aiming to deepen our understanding of influence dynamics. Although the demonstrated method exhibits considerable efficacy, there is still room for improvement. Enhancements can be pursued by exploring additional types of measurements, becoming more meticulous with data organization, and monitoring changes in social media trends. To better understand influence dynamics, researchers can expand the method’s scope by considering insights drawn from related fields or platforms.

Bibliography

Documenting the Now Project. (2021). *DocNow/twarc* (2.8.0) [Python]. Documenting the Now. https://doi.org/10.5281/zenodo.593575. (Original work published 2013)

Dongwoo, K., Yohan, J., Il-Chul, M., & Oh, A. H. (2010, April 10). Analysis of Twitter Lists as a Potential Source for Discovering Latent Characteristics of Users. *Workshop on Microblogging*. CHI '10: Proceedings of the SIGCHI Conference on Human Factors in Computing Systems, Atlanta, CA, USA.

Dorsey, J. [@jack]. (2018, September 5). *In any public space, you'll find inspired ideas, and you'll find lies and deception. People who want to help others and unify, and people who want to hurt others and themselves and divide. What separates a physical and digital public space is greater accessibility and velocity.* [Post]. X. https://twitter.com/jack/status/1037339409862025216

Harmon, R., J. & other contributors. (2009). *Tweepy Documentation* [Computer software]. https://docs.tweepy.org/en/stable/api.html#tweepy.API.get_list_members

Harmon, R., J. & other contributors. (2021). *Tweepy* (4.1.0) [Python]. https://doi.org/10.5281/zenodo.7860636

Heumann, S. (2018, October 12). Why Social Media Platforms Should Be Treated as Critical Infrastructures. *Election Interference in the Digital Age*. https://medium.com/election-interference-in-the-digital-age/why-social-media-platforms-should-be-treated-as-critical-infrastructures-6a437a127ff7

Hey, T., Tansley, S., & Tolle, K. (Eds.). (2009). Jim Gray on eScience: A transformed scientific method. In *The fourth paradigm: Data-intensive scientific discovery*. Microsoft Research.

Ke, Q., Ahn, Y.-Y., & Sugimoto, C. R. (2017). A systematic identification and analysis of scientists on Twitter. *PLOS ONE*, *12*(4), e0175368. https://doi.org/10.1371/journal.pone.0175368

Kitchin, R. (2017). Big Data – Hype or Revolution? In A. Quan-Haase & L. Sloan (Eds.), *The SAGE handbook of social media research methods* (pp. 27–40). SAGE.

Kwak, H., Lee, C., Park, H., & Moon, S. (2010). What is Twitter, a social network or a news media? *Proceedings of the 19th International Conference on World Wide Web*, 591–600. https://doi.org/10.1145/1772690.1772751

Lazarsfeld, P. F., Berelson, B., & Gaudet, H. (1960). *The people's choice: How the voter makes up his mind in a presidential campaign* (2nd ed). Columbia University Press.

McClain, C., Widjaya, R., Rivero, G., & Smith, A. (2021, November 15). The behaviors and attitudes of U.S. adults on twitter. *Pew Research Center: Internet, Science & Tech*. https://www.pewresearch.org/internet/2021/11/15/the-behaviors-and-attitudes-of-u-s-adults-on-twitter/

McPherson, M., Smith-Lovin, L., & Cook, J. M. (2001). Birds of a Feather: Homophily in Social Networks. *Annual Review of Sociology*, *27*(1), 415–444. https://doi.org/10.1146/annurev.soc.27.1.415

Musk, E. [@elonmusk]. (2022, October 27). *Dear Twitter Advertisers https://t.co/GMwHmInPAS* [Post]. X. https://twitter.com/elonmusk/status/1585619322239561728

Myers, S. A., Sharma, A., Gupta, P., & Lin, J. (2014). Information network or social network?: The structure of the Twitter follow graph. *Proceedings of the 23rd International Conference on World Wide Web*, 493–498. https://doi.org/10.1145/2567948.2576939

Park, C. S. (2013). Does Twitter motivate involvement in politics? Tweeting, opinion leadership, and political engagement. *Computers in Human Behavior*, *29*(4), 1641–1648. https://doi.org/10.1016/j.chb.2013.01.044

Python Software Foundation. (2021). *Python Language Reference* (3.9.5) [Python]. https://www.python.org

Rogers, E. M. (2003). *Diffusion of innovations* (5th ed). Free Press.

Sharma, N. K., Ghosh, S., Benevenuto, F., Ganguly, N., & Gummadi, K. (2012). Inferring who-is-who in the Twitter social network. *ACM SIGCOMM Computer Communication Review*, *42*(4), 533–538. https://doi.org/10.1145/2377677.2377782

Sharpe, A. (2020, October 22). *Is Social Media Critical Infrastructure?* LinkedIn. https://www.linkedin.com/pulse/social-media-critical-infrastructure-alex-sharpe

Sloan, L., & Quan-Haase, A. (2017). Introduction to the Handbook of Social Media Research Methods: Goals, Challenges and Innovations. In *The SAGE handbook of social media research methods* (pp. 1–10). SAGE reference.

The pandas development team. (2021). *pandas-dev/pandas: Pandas* (1.3.3) [Python]. https://doi.org/10.5281/zenodo.3509134 (Original work published 2010)

Törnberg, P. (2022). How digital media drive affective polarization through partisan sorting. *Proceedings of the National Academy of Sciences*, *119*(42), e2207159119. https://doi.org/10.1073/pnas.2207159119

Twitter Engagement Calculator. (n.d.). *Mention*. Retrieved December 13, 2023, from https://mention.com/en/twitter-engagement-calculator/

Twitter Engagement Metric. (n.d.). *Klipfolio*. Retrieved December 13, 2023, from https://www.klipfolio.com/resources/kpi-examples/social-media/twitter-engagement-metrics

We Are Social. (2023). *Digital 2023: Global Overview Report*. https://wearesocial.com/uk/blog/2023/01/digital-2023/

Wu, S., Hofman, J., Mason, W., & Watts, D. (2011). *Who says what to whom on Twitter*. 705–714. https://doi.org/10.1145/1963405.1963504

X Developers. (n.d.-a). *API Reference* (1.1) [Computer software]. X Corp. Retrieved December 13, 2023, from https://developer.twitter.com/en/docs/twitter-api/v1/accounts-and-users/create-manage-lists/api-reference/get-lists-memberships

X Developers. (n.d.-b). *GET lists/members* (1.1) [Computer software]. X Corp. Retrieved December 13, 2023, from https://developer.twitter.com/en/docs/twitter-api/v1/accounts-and-users/create-manage-lists/api-reference/get-lists-members

X Developers. (n.d.-c). *Twitter API Documentation* [Computer software]. Retrieved December 13, 2023, from https://developer.twitter.com/en/docs/twitter-api

X (Twitter) Engagement Rate Benchmark. (n.d.). *Social Status*. Retrieved December 13, 2023, from https://www.socialstatus.io/insights/social-media-benchmarks/twitter-engagement-rate-benchmark/

Yamaguchi, Y., Amagasa, T., & Kitagawa, H. (2011). Tag-based User Topic Discovery Using Twitter Lists. *2011 International Conference on Advances in Social Networks Analysis and Mining*, 13–20. https://doi.org/10.1109/ASONAM.2011.58