Online Toxicity Against Syrians in Turkish Twitter: Analysis and Implications

This study examines the portrayal of Syrians on Turkish Twitter between January and August 2021 through a big data analysis of over 30,000 tweets. We employ the concept of online toxicity to differentiate between disinformation and hate speech and explore how they are embedded in the negative debates about Syrians on Twitter. Through opinion analysis, the study recognizes disinformation and hate speech patterns within the tweets and questions the role they play in boosting anti-Syrian narratives, as well as the main actors behind them in the Turkish Twittersphere. The findings indicate that discourse regarding Syrians on Twitter was overwhelmingly negative, with both disinformation and hate speech playing a significant role. Furthermore, a considerable portion of the disinformation tweets can be traced back to opposition political actors, highlighting how negative sentiment on Twitter was not only expressive of a generalized public resentment against Syrians but instrumentalized for political purposes. Overall, the paper demonstrates how Twitter contributes to public debate about Syrians in Turkey, reproducing nationalist narratives and serving political agendas.

Keywords: Twitter, refugees, big data, disinformation, hate speech, online toxicity.

The Syrian refugee crisis constitutes a major consequence of the continuous armed conflict in Syria. With Turkey hosting more than 3.5 million Syrian refugees (UNHCR, 2021), the country's anti-Syrian sentiment has been growing (Karakas, 2021). This is largely the result of the politicization of the question of Syrians' status in Turkish society especially with the changes in the political scene. After the 2017 referendum, Turkey witnessed the emergence of two major pre-electoral alliances, the first is the ruling People's

Alliance (Cumhur İttifakı) which mainly includes the Justice and Development Party (AKP), the Nationalist Movement Party (MHP). The second represents the opposition Nation's Alliance (Millet İttifakı) which consists of the Republican People's Party (CHP), Good Party, the Islamist Felicity Party (SP), Democrat Party (DP), the Democracy and Progress Party (DEVA) and the Future Party (GP) (Secen et al., 2023). Most of these opposition parties besides the newly emerged Victory Party (Zafer) employed promises to expel Syrians from the country as part of their election campaigns (Farooq, 2021). These arguments became the focus of many debates across social media, particularly Twitter. Syrian refugees have been misrepresented and become a target of online toxicity in terms of disinformation, hate speech, and digital racism (Ozduzen, 2020). Beyond the psychological harm, such online toxicity has fueled physical and verbal hate crimes against Syrian refugees (Özerim & Tolay, 2021). An example of such attacks was the August 2021 violent riots against Syrians in the Altındağ neighborhood in Ankara (BBC-News, 2021).

Relevant research in media studies has investigated online discourses against refugees, particularly in the aftermath of the 2015 "refugee crisis" at Europe's borders (Erdogan-Ozturk & Isik-Guler, 2020); Most of these studies have focused on the social aspects of digital racism and affiliated it with Turkish populism (Özerim & Tolay, 2021). Others have integrated quantitative analysis of a single hashtag that was trending in a specific context, such as the Taksim Square Protests (2019), and investigated the relevant tweets (Ozduzen, 2020). This body of research has highlighted the circulation of hateful discourses and "online toxicity" (Kim et al., 2021). against refugees, in Turkey and beyond.

Nevertheless, online toxicity has not so far been adequately conceptualized in its different components or the social and political actors instigating it. In this paper, we seek to explore the prominent discourses about Syrians on Turkish Twitter, how negative sentiment against them relates to disinformation and hate speech, and the actors behind disinformation and hatred campaigns. Our findings demonstrate that Twitter is not only reflective of a generalized negative sentiment in Turkish society against Syrians but, rather, reproduces and amplifies toxic phenomena such as disinformation and hate speech. This is motivated by the fact that with more than 18 million active users in Turkey, Twitter has grown popular among Turkish political leaders as they represent 40% of the top 15 Twitter users1. Therefore, we found that online toxicity directed against the Syrian population is often fabricated and affiliated with a specific political agenda related to the debates between the ruling coalition and the opposition. We argue that these hate speech campaigns on social media are being strategically instrumentalized by identifiable political actors rooted within particular social and spatial contexts.

The contribution of our paper is, thus, twofold. Empirically, we unpack the concept of online toxicity to illustrate the distinction but also the interplay between online hate speech and disinformation. Methodologically, we employ a novel framework that utilizes a machine learning-based sentiment analysis model proposed by Author (2018) and Natural Language Processing (NLP) techniques to collect and investigate more than 30K Turkish tweets during the period January-August 2021 and recognize both hate speech and disinformation tweets, along with their source geolocations. Overall, the paper provides research insights both into discussions on the mediation of forced migration as well as academic debates on online hate speech and disinformation.

Syrian Refugees on Social Media

The role of social media platforms in migration discourses has been underscored by ambivalence. On the one hand, social media has been seen as providing space for alternative representations of the hostility afforded to migrants by mainstream media. While media coverage has been largely negative, if not overtly hostile, against migrants and refugees in different national contexts (Chouliaraki & Stolic, 2017), social media has allowed for a space for alternative representations, where solidarity for migrants and refugees can be expressed, challenging dominant policies and mainstream coverage (Siapera, 2019). Furthermore, social media platforms have provided refugees with a ‘voice’ (Georgiou, 2018) within the digital media space, as well as opportunities for social support and resilience (Udwan, Leurs, & Alencar, 2020) and enabled tracing their migration route

[1] https://www.statista.com/statistics/1318367/turkey-twitter-accounts-by-follower-numbers/

patterns even when based on the approximate locations in geotagged tweets (Hübl et al., 2017).

On the other hand, significant research has illustrated that dominant discourses on social media have echoed the hostile mainstream media coverage. The hashtag #refugeesnotwelcome, for example, was employed in different European contexts to reproduce a broader rhetoric of exclusion, constructing refugees as outsiders and threatening criminals (Kreis, 2017). Although refugee narratives on social media are far from monolithic, processes of othering of refugees are not only present but even more dehumanizing than mainstream media representations, as hashtags such as #rapefugees indicate.

These broader themes are echoed in the portrayal of Syrians in Turkey. Hate speech against Syrians in Turkish media was already present in 2015 (Az et al., 2017) and has been increasing since. Syrians have been negatively stereotyped in terms of their culture, as well as being an economic burden to Turkey, in ways that have contributed to expressions of aggression and violence against them (Alp, 2018). The least negative coverage, according to (Şen, 2017) can be found in Hürriyet and HaberTürk, both newspapers supporting the governing Justice and Development Party (AKP), and evidently following its more open policy towards refugees.

In social media, negative stereotypes were further amplified. According to (Aydınlı, 2020), Syrians have been criticized for not fighting for their country and for financially benefiting from Turkey, while also being constructed as a threat to the local population. Disinformation shared on social media platforms further reiterated the myth of perceived privileges enjoyed by Syrians. Twitter has become a forum for “digital racism”, allowing for the racialization of refugees and the circulation of xenophobia (Ozduzen, 2020). Overall, the diverse rhetorical practices of othering Syrians on Twitter have included their construction as criminals and cowards and traitors of their homelands, invaders in Turkey, possible terrorists, fake refugees, and “lesser” Muslims than Turkish people (Erdogan-Ozturk & Isik-Guler, 2020).

These negative portrayals have served broader populist and nationalist discourses, reaffirming relations of inclusion and exclusion and ultimately the national narrative (Aridici, 2022). The hashtag *#ÜlkemdesuriyeliİstemiyorumIstemiyorum* whose English translation is (#IdontwantSyriansinmycountry) on Twitter has been used to express and reinforce ideas of a strong and exclusionary nationalist identity (Erdogan-Ozturk & Isik-Guler, 2020; Özerim & Tolay, 2021). Studying Turkish Twitter in the aftermath of Erdogan's statement about giving Syrians citizenship rights in the summer of 2016, (Bozdağ, 2020) found that even the minority of Tweets supporting citizenship for Syrians adopted a nationalist frame, either doing so based on the imagination of Islamic brotherhood or emphasizing the importance of security investigations for those considered for citizenship. Similarly, her research on how Syrians were represented during COVID-19 both in mainstream and online news media in Turkey, (Yücel, 2021) found that Syrian refugees were largely ignored or "symbolically annihilated" in media coverage of the pandemic, even though they faced the pandemic's greatest socioeconomic consequences. The few news stories that mentioned them, focused on Turkey's "generous" policies while criticizing European immigration policies.

What these studies illustrate is that Twitter discussions are not only expressive of a general negative sentiment but also of its instrumentalization within specific nationalist discourses and agendas. Therefore, it is important to study online toxicity not only in relation to hate speech but also disinformation. Although the concept has been used in a variety of ways and to different ends by communication scholars (Masullo Chen et al., 2019), following the definition provided by (Kim et al., 2021), we understand online toxicity as the political comments "expressing disrespect for someone by using insulting language, profanity, or name-calling; by engaging in personal attacks; and/or by employing racist, sexist, and xenophobic terms". The multidimensionality of the concept, which includes threats, obscenity, insults, hate, as well as harassment, and "socially disruptive persuasion, such as misinformation and radicalization" (Sheth, Shalin, & Kursuncu, 2021), allows the accounting for the variety of ways in which negative sentiments and hostile discourses against Syrians are expressed on Turkish Twitter. What we mainly focus on in this paper is the interplay between what can be understood as hate speech and instrumental practices of disinformation, both important components of online toxicity. This focus allows us to move beyond depoliticized notions of "uncivil" or "impolite" speech (Masullo Chen et al., 2019), to explore how online toxicity can be purposefully employed as a political tool.

Despite the lack of a universally accepted definition of hate speech (Alkiviadou, 2019), a widely used conceptualization is (Nockleby, 2000), which describes the concept as "any communication that disparages a person or a group on the basis of some characteristic such as race, color, ethnicity, gender, sexual orientation, nationality, religion, or other characteristics". The circulation of hate speech on social media platforms has been a concern for media scholars, especially in relation to far-right discourses in different contexts (e.g., Vidgen & Yasseri, 2020). Calls for a regulatory framework to tackle the ease and speed with which hateful content circulates online (Alkiviadou, 2019) have been met slowly and reluctantly by the adoption of self-regulation practices by the different social media platforms. Twitter itself has a "hateful conduct policy"[2] under its defined rules and policies, which proclaims to not allow users to "promote violence against or directly attack or threaten other people based on race, ethnicity, national origin, caste, sexual orientation, gender, gender identity, religious affiliation, age, disability, or serious disease". Understanding online toxicity against Syrians as hate speech allows us to detect the type of hateful Twitter comments that can be seen not only as discriminatory and racist but also as inciting violence against them.

The second dimension of online toxicity we consider in this paper is that of disinformation, namely the dissemination of deliberately falsified information (Wardle, Derakhshan, & others, 2018). There is a "discursive affinity" between hate speech and disinformation, as their tactics and aims often align when the target of spreading falsified information is to instigate and promote negative sentiment against specific groups (Hameleers, van der Meer, & Vliegenthart, 2021). "Hate groups" and ideologues, for example, such as white supremacists, are among the dominant groups spreading disinformation online (Marwick & Lewis, 2017). At the same time, however, it is important

[2] https://help.twitter.com/en/rules-and-policies/hateful-conduct-policy

to distinguish between the two concepts, hate speech and disinformation, as hate speech is not always the result of disinformation and disinformation is not exclusively aiming to spread hate. Exploring them separately allows us to identify the nuances of Twitter debates, as well as trace the production and circulation of online toxicity to the communicative agency of strategic deceptions deployed by influential political actors. On the other hand, we believe that the process of producing and spreading online toxicity can be further investigated when aligned with the geographic context. Such a context forms an important aspect to be considered while conducting refugee-related studies (Hübl et al., 2017). Consequently, we seek to address the following research questions:

1. What are the prominent discourses being circulated about Syrians on Turkish Twitter?
2. How does the negative stance against Syrians on Twitter relate to disinformation and hate speech?
3. Which actors are behind the online smear campaigns against Syrians?

Methodology

In seeking to identify patterns of online toxicity and disinformation against Syrian refugees in the Turkish Twittersphere, we conducted a big data study. We collected the tweets using the trending hashtags related to Syrians in Turkish Twitter to first identify the negative views within the sample through a sentiment analysis process, and then explicate instances of disinformation or hate speech among them, while also considering the associating geolocation information for each tweet category. To these ends, we employed a framework in which a large-scale collection of Turkish tweets was scraped and then subjected to Natural Language Processing (NLP) and Machine Learning (ML) techniques to extract the views in these tweets and identify patterns of disinformation and hate speech.

We mined the tweets during the period January - August 2021. Taking the anti-refugee riots in Ankara in August 2021 as the end point of our sample, we decided to mine tweets from January of the same year, to identify changes over the year. We mined our sample using Twitter API and harvested the tweets along with their associating metadata such as the username, location, number of followers, number of likes, number of retweets, etc. This enabled avoiding the limitations of data size, quality, and availability faced by previous studies which used off-the-shelf scraping software (Ozduzen, 2020; Özerim & Tolay, 2021). The collection process relied on top trending hashtags (Table 1) about Syrians due to the intensive engagement. This allows for better capturing of the main discourses circulated about Syrians, the most common portrayals of Syrian refugees, and the public stance toward them. Also, we tracked the most-liked and most-retweeted relevant tweets and scraped the tweet replies posted within their threads. Some of the collected tweets were harvested from the timelines of specific users (proponents and opponents of hosting Syrians) who are considered influencers in Turkey.

Table 1. The trending hashtags used to scrape the tweets.

Hashtag (Turkish)	Hashtag (English)
#SuriyelileriAlmayın	*#Don't Host Syrians*
#SuriyelilerinVatanıSuriyedir	*#Syrian's Home is Syria*
#UlkemdeMülteciİstemiyorum	*#I don't Want Refugees in My Country*
#ProvokasyonaGelme	*#Don't trigger Provocation*
#Suriyelileriİstemiyoruz	*#We don't want Syrians*
#KardeşimeDokunma	*#Don't Hurt My Brother*

As a result, the raw collected data was composed of 276,093 tweets. Except for the influential users - politicians, celebrities, and other public figures - the author names (usernames) of the tweets were masked throughout the analysis. The raw collected tweets were, then, subjected to several cleaning and normalization procedures such as excluding irrelevant (e.g., tweets containing Afghan-related keywords or hashtags) and objective (opinion-free i.e., news) tweets. We also applied a stemming technique using Zemberek stemmer (Akın & Akın, 2007) and reduced Twitter-inherited symbols, punctuation, and stopwords - Author (2018). While this step has been ignored in the literature (Assimakopoulos, Baider, & Millar, 2017; Kreis, 2017), we believe tweet cleaning and normalization are crucial to conducting an accurate opinion identification and thus reaching context-relevant conclusions. Consequently, we ended up with a collection of 33,100 Syrian-related, subjective tweets. These tweets were posted by 18,738 unique accounts during the timeframe of our study (January 1st - August 31st, 2021). These were both from international and Turkish locations as shown in Table 2.

Table 2. Statistics of the studied tweet collection.

Property	#Tweets/Accounts
Subjective Tweets	33, 100
Unique Accounts	18,738
Tweets having geolocations	12,177
Tweets having real geolocations	10,701

For sentiment analysis, unlike similar studies which relied totally on the author's judgment to identify the sentiment in the tweets (Kreis, 2017b), we employed a machine learning-based opinion analysis model for Turkish. This model was developed by Author (2018) and trained to recognize negative and positive opinions embedded in Turkish tweets based on specific combinations of linguistic and stylistic features. It should be noted that while previous research tends to use off-the-shelf sentiment analysis software as it is (Ozduzen, 2020), we opted to customize the model presented in Author (2018) in terms of enriching the training data with a manually annotated subset of the studied tweets. The manual annotation for positive/negative tweets was provided by two Turkish native speakers with an inter-rater agreement of 87%. Additionally, to handle a potential imbalanced distribution of the classes (positive/negative) among the tweets, we split the annotated tweet dataset into training, validation, and test sets using a stratified sampling technique to ensure that the training, validation, and test sets have approximately the same percentage of samples of each target class as the complete set (Zhao He and Chen 2016). To do so, we took a random 20% of the tweets from each class for the validation and test sets. This, on one hand, enhanced the generalization ability of the model to our collected tweets and on the other hand, enabled a better evaluation of the classification performance of the model and, thus, the model's reliability.

In seeking to further mine the negative tweets and detect those containing disinformation, we adopted a hybrid method that combines word-based analysis along with syntactic analysis such that word collocations and specific syntactic patterns are both considered. This is motivated by the fact that truth-tellers and liars have distinctive styles of writing. For instance, the authors of disinformation tweets tend to use fewer self-oriented pronouns than other-oriented pronouns, along with using more sensory-based words (Stahl, 2018). The used hybrid method combined the following steps:

- Spotting the most frequent word collocations/associations (pairs/triples) based on a scoring approach (TF-IDF[3]) (Mondal et al., 2022).
- Using the top 1,000 pairs/triples as search queries to filter the tweets that contain them.
- Identifying the Syrians-related disinformation content using Multi-Word Terms (MWTs) and Zemberek syntactic parser (Akın & Akın, 2007). MWTs are the meaning indicators of a sentence/document (Author, 2019a; Henry, Cuffy, & McInnes, 2018). In our case, they represent certain syntactic patterns (MWTs) that can refer to allegations (topics) about Syrians. In such syntactic patterns, proper nouns like "Suriyeli/Suriyeliler" *(Syrian/Syrians)* may collocate with nouns such as "Maas¸" *(salary)* or proper nouns such as "Esad" *(Bashar Assad)*, or verbs like "geziyorlar" *(traveling)*, and verbs + negations like "Mazlum değil" *(not oppressed)* such word collocations are inspired by the social/political context of Syrian refugees in Turkey.

To detect hate speech, we follow (Waseem et al., 2017), who argue that hate speech discourses are mainly composed of abusive language which makes hate speech detection a sub-task of abusive language detection (Waseem et al., 2017). Therefore, to detect the tweets containing hate speech, we explored all negative tweets that were found not to contain disinformation and filtered those that combine offensive/abusive terms and phrases as a preliminary stage before hate speech identification. This differentiates our research

[3] Term Frequency-Inverse Document Frequency (TF-IDF) is a statistical measure that indicates how many times a word/term appears in a piece of text.

from the previous studies, which either relied on the personal judgment of the author to detect hate speech tweets (Assimakopoulos et al., 2017) or adopted a limited list of hate speech keywords (Aslan et al., 2017). Thus, we first identified offensive/abusive tweets based on extended prepared lists of Turkish words used in cyberbullying[4] and offensive/abusive speech[5]. Then, to spot the hate speech tweets, we explored the offensive/abusive tweets looking for specific entities that represent potential hate speech targets. These entities were derived from a collection of words/phrases which are either related to the religion/ethnicity of Syrians, targeting them based on gender identity, or inspired by the slang words/phrases (nicknames) used by the Turkish community to describe Syrians, such as "Suriyeli/Suriyeliler" *(*Syrian/Syrians), "Arap" *(*Arab), "Bedevi" *(*Bedouin), "Çingene" *(*Gypsy), and "Bayanlar/Kadınlar/Kızlar" *(*Women/Girls).

Given that a tweet might include both disinformation and hate speech, we have all the negative tweets subjected to a 2-phase classification process in which the disinformation tweets are first filtered and then further mined for hate speech content. The rest of the negative sentiment tweets, what we call "oppositional", as well as the "supportive" or positive tweets were excluded from the analysis. Python[6] and Gephi[7] were utilized to support visual representations of the studied data.

As for the evaluation of the models adopted for disinformation detection and hate speech recognition, their performance was evaluated based on a subset of 2000 disinformation and 2000 hate speech tweets (control group) manually annotated by two Turkish native speakers. Like the sentiment classification task, for disinformation and hate speech classification tasks stratified sampling was applied during the training and evaluation of disinformation and hate speech classification models.

[4] https://github.com/elfswl95/turkish_cyberbullying_on_twitter
[5] https://github.com/ooguz/turkce-kufur-karaliste
[6] https://www.python.org
[7] https://gephi.org

Findings and Discussion

The Main Discourses About Syrian Refugees on Turkish Twitter

Out of the 33,100 subjective tweets in our sample, the sentiment analysis model Author (2018) recognized 29,595 (89%) tweets as being of negative opinion and 3,505 (11%) tweets of positive opinion achieving a classification accuracy of (94%). The percentage of negative tweets against the positive ones reflects the findings of existing literature that has pointed out the general negative sentiment against Syrian refugees on Turkish Twitter (e.g., (Aridici, 2022; Ozduzen, 2020)). It is also an illustration of the increasingly hostile environment against Syrians in Turkey (Farooq, 2021).

Although we used a limited number of trending hashtags as an initial basis for tweet collection (Table 1), the adopted 33,100 tweets contained 70,630 hashtags in total; out of these, 510 hashtags were unique. We calculated the frequency of hashtags that were used more than 1,500 times. The top frequent hashtags among them are listed in Table (3).

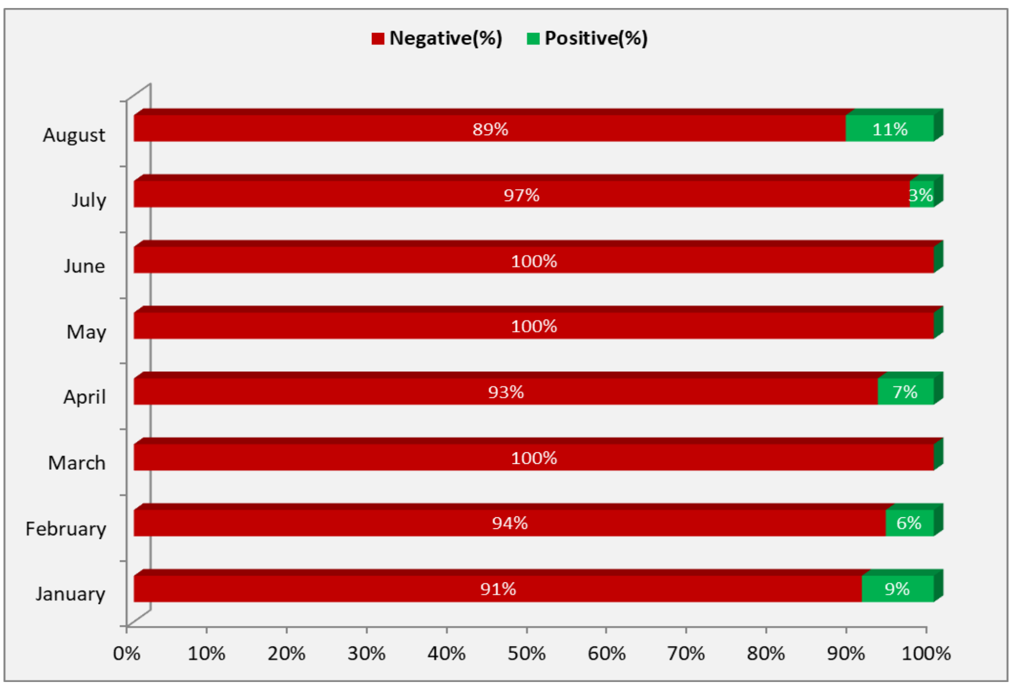
Table 3. Most frequent hashtags in the collected tweets.

Hashtag (Turkish)	Hashtag (English)
#SuriyelilerinVatanıSuriyedir	*#Syrian's Home is Syria*
#KardeşimeDokunma	*#Don't Hurt My Brother*
#VatanEldenGidiyor	*#Our Homeland is Gone*
#afganlariistemiyoruz	*#We don't Want Afghans*
#AfganlarıAlmayın	*#Don't Host the Afghans*
#Suriyelileriİstemiyoruz	*#We don't want Syrians*
#ÜlkemdeMülteciİstemiyorum	*#I don't Want Refugees in My Country*
#SığınmacılarSınırDışıEdilecek	*#Asylum Seekers Will Be Deported*
#SuriyelilerSuriyeye	*#Syrians! go to Syria*
#Müteciİstemiyoru	*#We don't Want Refugees*
#Altındağ	*#Altındağ (A Neighborhood in Ankara)*
#SuriyelileriAlmayın	*#Don't Host Syrians*
#YananHepBizOlduk	*#We're the ones who are burning*

Employing deictic words, such as "we", "my", and "our", the hashtags reproduce an exclusionary nationalist discourse, whereby Syrians are constructed as "other", as outsiders that do not belong in the national community. At the same time, Turkish people are constructed as victims, "the ones that are burning" and "whose homeland is gone". A few of the hashtags are explicitly associated with the August riots, such as #Altındağ, the area where hostilities against Syrians exploded, and #KardeşimeDokunma, used by pro-migrant groups during the riots.

We also explored how tweets of negative and positive opinions were distributed over time (see Figure 1).

Figure 1. Monthly Distribution (%) of positive/negative tweets.



Negative tweets were conspicuously dominant every month and represented the totality of tweets during March, May, and June. This can be attributed to the fact that these are the months leading to the Muslim Eid holidays when Syrian refugees are allowed by the Turkish government to visit their families in northern Syria, controlled by the Syrian opposition. These short-term holidays tend to fuel resentment among the Turkish community. The fact that Syrians can temporarily return to their homeland has been used as an argument to support the idea that Syria is a safe place for them al Mahmoud (2021). For example, Ümit Özdağ, Chairman of the Victory (Zafer) Party, addressed Syrians in a tweet that went viral: *"#ThisVisitHasLastedTooLong (#BuMisafirlikFazlaUzadı). This tweet is dedicated to the Syrian youth who went to Syria to celebrate Eid and get married. While they were going to the feast, the Turkish soldier who was either martyred in Syria and Mehmetçik or became disabled with his leg amputated, these soldiers were at the age of marriage... July 26, 2021"*.

The biggest percentage of positive tweets, albeit still small at 11%, was spotted posted in August, the month of the violent episodes against Syrians in Ankara. It seems that support was only mobilized as a reaction to an extreme moment of violence and has only been minimally expressed throughout the rest of the year. Given the politicization of the debate about Syrians, much of this support was expressed by government supporters.

Disinformation and Hate Speech Against Syrians

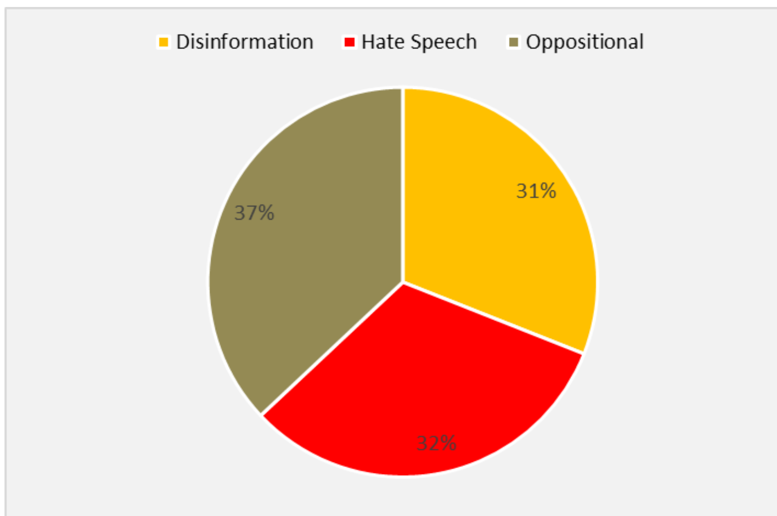
Having the 29,595 negative tweets subjected to specific NLP methods and lexical resources, as discussed in the methodology section, and with a classification accuracy of (91%) we identified 9,255 disinformation tweets while the hate speech detection model could recognize 9,374 hate speech tweets achieving an accuracy of (96%), while 1,077 tweets were found having both disinformation and hate speech content. As the latter group of tweets is significantly small compared to each of the disinformation and hate speech tweet collections, we opted to exclude it from the analysis and focus on the tweets that were identified either as disinformation or hate speech. Similarly, we did not include the "supportive" and "oppositional" tweets in further analysis. Table (4) reviews tweet examples of oppositional, disinformation, and hate speech categories while Figure 2 illustrates the distribution of these tweet categories across the negative tweet collection.

Table 4. Examples of oppositional, disinformation, and hate speech tweets.

Category	Tweet Example
Disinformation	Suriyeliler bombalandıkları için Türkiye'ye gelmiyorlar, Türkiye'ye gelmeleri için bombalanıyorlar. Emperyalizmin Orta Doğu'da oluşturmaya çalıştığı yeni yapay sınırlara Hayır demek için, #SuriyelilerSuriyeye! *Syrians do not come to Turkey because they are bombed, they are bombed to come to Turkey. Say No to the new artificial borders that imperialism is trying to create in the Middle East, #SyriansToSyria!*
	Benzine 55, motorine 67, LPG'ye 35 kuruş zam yapıldı. Türk halkı açlıktan çıldırıp intihar ederken, Suriyelilere 5 yıldızlı bahçeli ev yapılıyor. Türk halkı maruz kaldığı zamlarla Suriyelilerin yemesini, içmesini ve ev sahibi olmasını sağlıyor. Yeter artık! *Gasoline was raised by 55 cents, diesel by 67 cents, and LPG by 35 cents. While the Turkish people are going crazy from hunger and committing suicide, Syrians are being given 5-star houses with gardens. With the hikes Turkish people are subjected to, Syrians can eat, drink, and own a house. Enough is enough!*
Hate Speech	Araplar ayrı, bedeviler ayrıdır. Bizim ülkemizde arap sığınmacı yok,çöl bedevileri var. Bedeviler, dünyanın en pis ve barbar insanlarıdır Zamanında, küçük bir ücret karşılığında Peygamberi bile taşlamislardır. Esad,suriyeli bedevileri ülkemize pompalamıştır #KardeşimeDokunma *Arabs are one thing, Bedouins are another. There are no Arab refugees in our country, there are Bedouins from the desert. Bedouins are the most filthy and barbaric people in the world. In their time, they even stoned the Prophet for a small fee. Assad has pumped Syrian Bedouins into our country #KardeşimeDokunma*
	Ben ırkçı falan değilim! Nankör (arap) ve hain (arap) sevmiyorum! #SuriyelilerMutlu *I am not a racist! I don't like ungrateful(arab) and traitors (arab)! #SyriansAreHappy*

Oppositional	Ev içinde ev, devlet içinde devlet olmaz. Suriyelilerin milli varlığımız ve birliğimiz için tehdit oluşturmalarına izin vermeyeceğiz. Onun içindir ki 👇 #SuriyelilerSuriyeye #VatandaşlıkVermeYolVer *There cannot be a house within a house and a state within a state. We will not allow Syrians to pose a threat to our national existence and unity. For this reason, 👇 #SyriansGiveSyria #CitizenshipGrantingGiveWay*

Figure 2. Distribution of disinformation, hate speech, and oppositional tweets.



As seen in the previous Figure 2, the proportion of tweets containing disinformation and hate speech tweets within the negative tweet collection confirm the interrelation between the negative sentiment toward Syrian refugees and the dissemination of disinformation and hate speech. Indeed, about one-third of negative tweets against Syrians contain hate speech, and almost another third reproduces disinformation that aims at undermining them. Although platform restrictions about hateful conduct do not apply to the tweets expressing opposition to Syrian refugees as such - what we call here "oppositional" - the high volumes of hate speech and disinformation are disconcerting, given Twitter's efforts to battle both phenomena[8] and the real-life impact these can have, inciting hate crimes and violence. Although this can be partially justified by the fact that Twitter algorithms cannot capture low-resourced languages such as the Turkish language and slang Turkish variants for automatic toxic content removal Author (2019b), it is also revealing of the lack of resources the platform is willing to allocate to this specific threat.

When it comes to disinformation, the most common claims made in our data sample can be seen in Table 4 below.

Table 4. Common claims extracted from disinformation tweets.

Top Frequent Claims
Suriyeliler Devleti kuruluyor *Syrians establish their country in Turkey*
Mülteciler değiller *They are not refugees*
Stratejik göç yapıyorlar *They perform a systematic migration*
Su bedava *Water service is free (i.e., Syrians don't pay water bills)*
Vergi yok *No taxes for Syrians*

Some of these disinformation claims reproduced the myth of financial benefits enjoyed by Syrians(Aydınlı, 2020), constructing them as a financial burden to Turkey, as well as implicitly comparing them to Turkish citizens that do not have access to such benefits. Others denied their status as refugees, claiming that their move to Turkey was deliberate and calculated rather than a desperate attempt to escape the war.

[8] https://help.twitter.com/en/rules-and-policies/manipulated-media

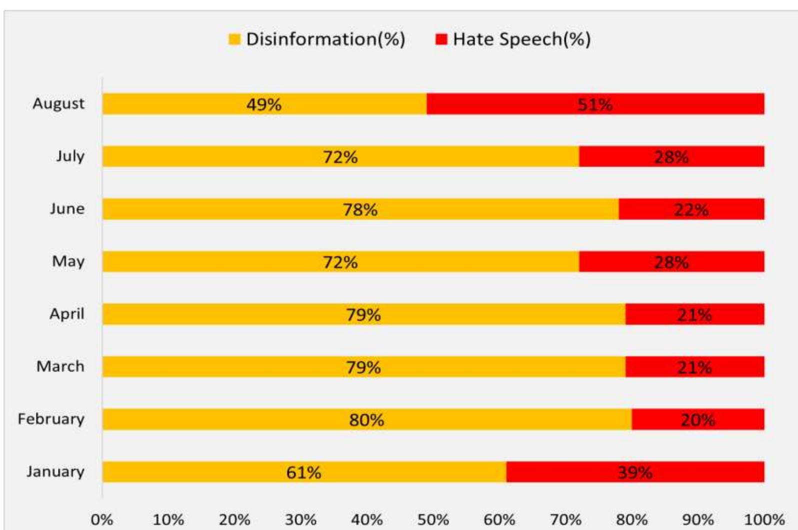
Similarly, when exploring the tweets recognized as hate speech, we identified the most frequent words/terms listed in Table 5. Most of these words are used in the context of threatening and dehumanizing Syrians, mocking their culture, religion, and ethnicity as Arabs, while claiming the superiority of Turkish people over Syrians.

Table 5. Top frequent terms in hate speech tweets.

Top frequent terms	Top frequent terms
Türküm (I am Turkish)	Irkçıyım (I am racist)
Arap (Arabs)	Bedevi (Bedouin)
Türkiye Türklerindir (Turkey is for Turks)	Piçler, şerefsizler (Bastards)
Pislik (Dirt)	Saldırısına (To attack)

We explored how disinformation and hate speech tweets were distributed over time, during the period studied (see Figure 3).

Figure 3. Distribution of disinformation and hate speech tweets (Jan - Aug 2021).



As seen in Figure 3, throughout the studied period, the percentages of disinformation tweets have been consistently high, ranging between 49% and 80%. These tweets can be seen as part of disinformation campaigns against Syrians, often politically motivated by some opposition parties that adopt an anti-refugee stance such as CHP and Zafer Party (Karabat, 2018), as well as the broader disinformation that seems to shape the information ecosystem of an increasingly polarized Turkey (Erdoğan, 2021). On the other hand, hate speech tweets constituted nearly 40% of the tweets in January. A couple of Syrian-related news stories preoccupied the public agenda that month. First, the Ministry of Interior published the numbers of Syrians expected to voluntarily return to their country by the end of 2021. At the same time, an attack against a Syrian family took place in Izmir in mid-January. Although hate tweets did not exceed 20% between February and July 2021, they represent more than half (51%) of negative tweets against Syrians during August, when the Ankara attacks against Syrians took place. This can be seen as an illustration of the close relationship between hate speech and violence, given that the former seems to have been instigated and likely further fueled the mobs against Syrian refugees in the August Altındağ riots.

Who Tweets Against Syrians on Turkish Twitter?

Out of the 12,177 tweets that included geolocation information in our sample, 10,701 tweets were associated with real locations, such as country or city. Figure 4 shows the distribution of the tweets across the international source locations where we can see that the tweets posted from Turkey formed 90% of the investigated tweets. This is, of course, to be expected, given many Turkish Twitter users reside in Turkey. The rest of the tweets were posted from countries in Europe, Canada, and the US, where Turkish and Syrian ex-pats reside.

Figure 4. Tweet Distribution among international locations.

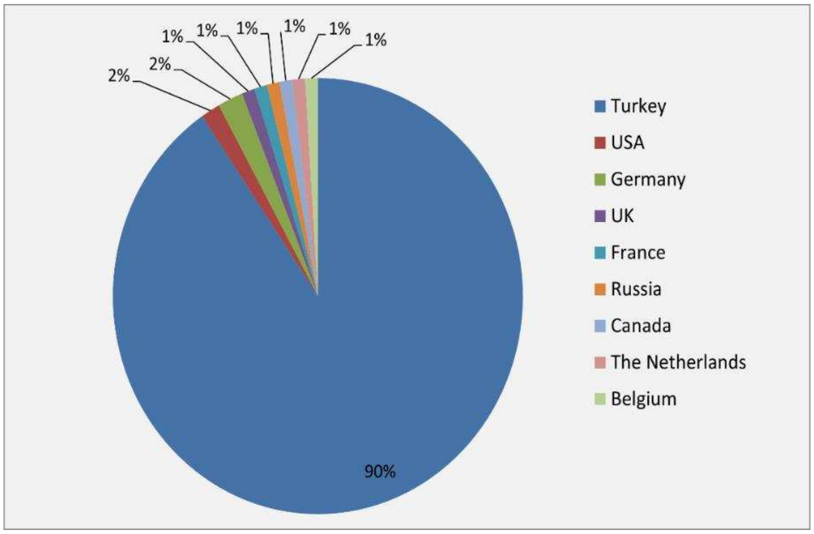
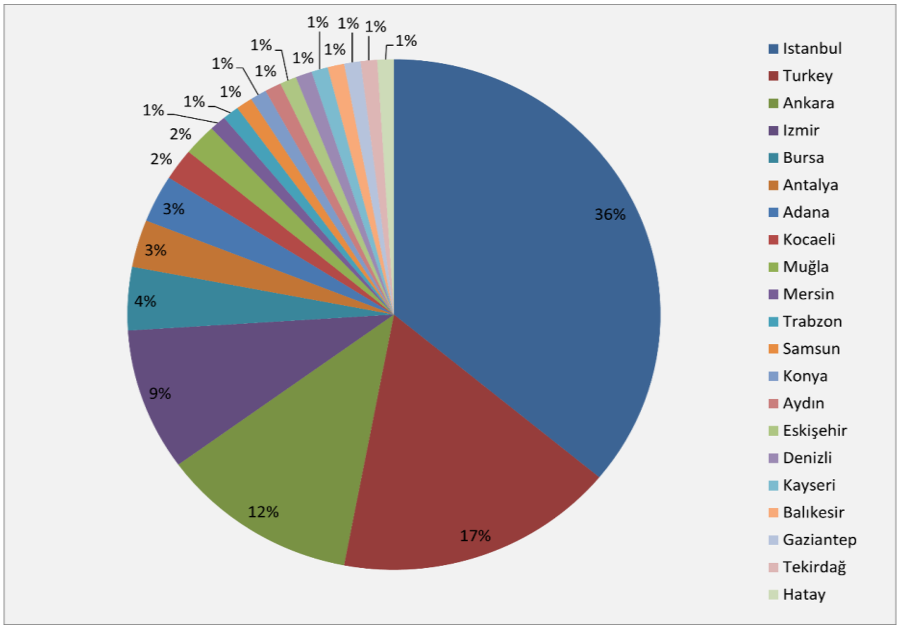


Figure (5) illustrates how tweets in our sample were distributed among Turkish cities. In this chart, it is observed that Istanbul, Ankara, and Izmir were the most prominent in our sample as geopolitical locations. This is also to be expected, as these are the three major cities, as well as important decision-making centers. On the other hand, although most of the Syrian refugees reside in these three cities[9] an important part of the tweets with Turkish locations (26%) were posted from different cities across the country. This indicates that the debates about Syrians within the Turkish community are not necessarily related to the distribution of Syrian refugees across Turkish cities but is an issue of general concern.

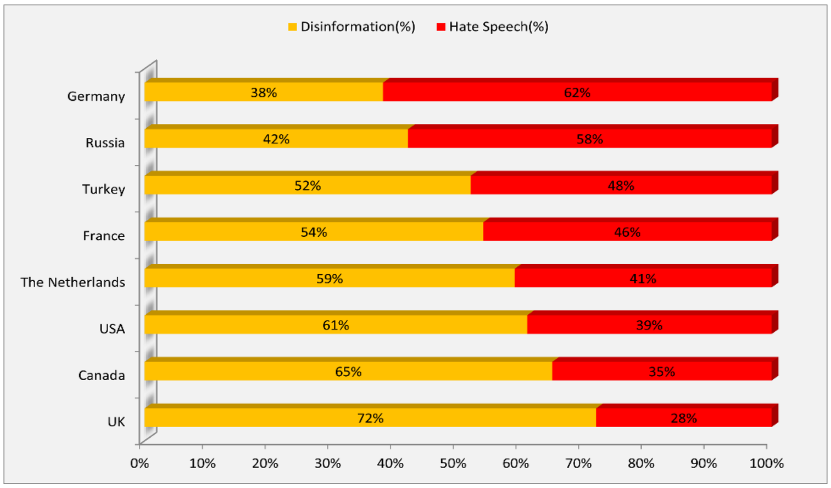
[9] https://www.goc.gov.tr/gecici-koruma5638

Figure 5. Tweet Distribution among Turkish locations.



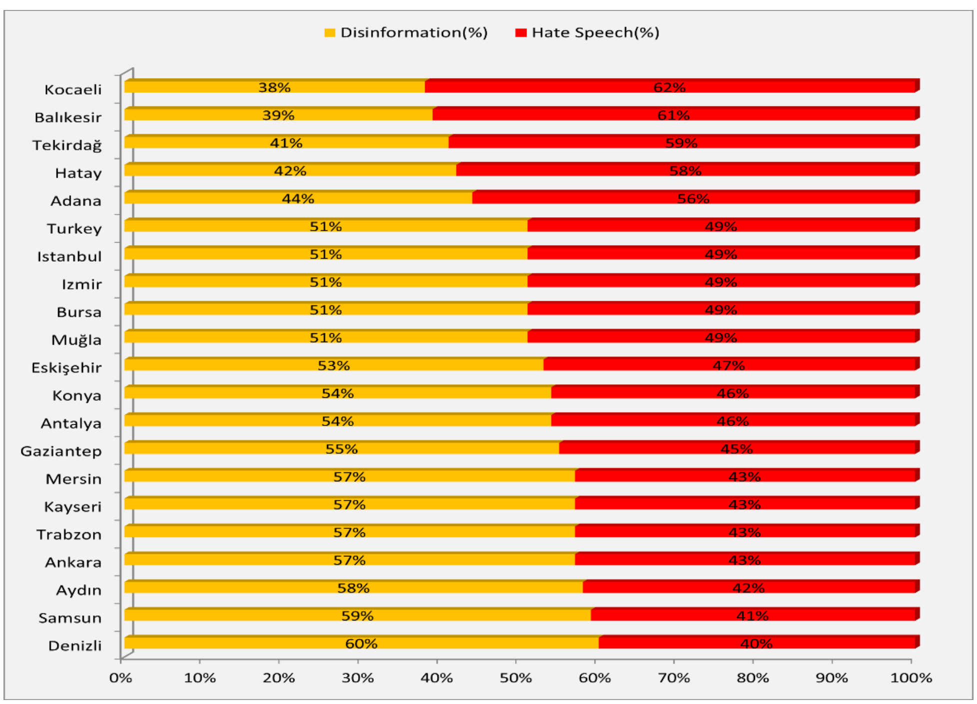
We further identified the source geolocations of disinformation and hate speech tweets. As seen in Figure 6, the distribution of disinformation and hate speech among international geolocations indicates that there are two countries, where hate speech against Syrians seems to be higher than in Turkey, namely Germany and Russia, which can be explained on geopolitical grounds. Turkish people make up Germany's largest minority at 3 million, importing a lot of national political tensions. At the same time, Russia's support for the Assad regime is expressed against Syrian refugees.

Figure 6. Disinformation/hate speech Distribution for international locations.



On the other hand, Figure 7 shows the distribution of disinformation and hate speech tweets across the source geolocations in Turkey.

Figure 7. Disinformation/hate speech Tweet Distribution across Turkish locations.

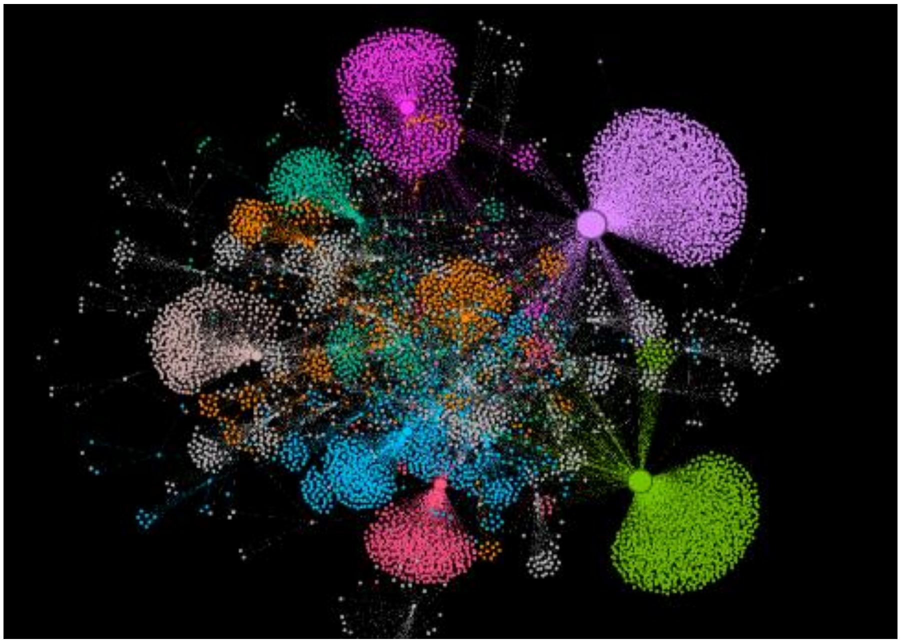


It seems that hate speech was more prominent in the areas where Syrians were discussed the least, as seen in Figure 5.

Most importantly, given that the retweet rate achieved for all the tweets on a user's timeline is considered among the factors that identify how influential this user is, we tracked the tweeting and retweeting activity among the users in the investigated disinformation and hate speech tweets, to spot the influential users whose tweets were retweeted the most. In the graphs below, the nodes denote the accounts involved in tweeting and retweeting, while the edges indicate that two users are related to each other by the retweet activity (i.e., one retweeted a tweet of the other). The colored clusters represent the communities of influencers and their retweeting users. These communities were created using the modularity-based clustering algorithm. In each community, we focused on the out-degree centrality that indicates the retweet rate of a user in this community such that nodes of higher out-degree appear as bigger-sized nodes and denote the influential users while the surrounding nodes connected to it represent the retweeting users. This was applied to disinformation and hate speech tweet collections to highlight the influencers and their communities of users who share the mentality/affiliation within which propaganda and polarized information spread especially well.

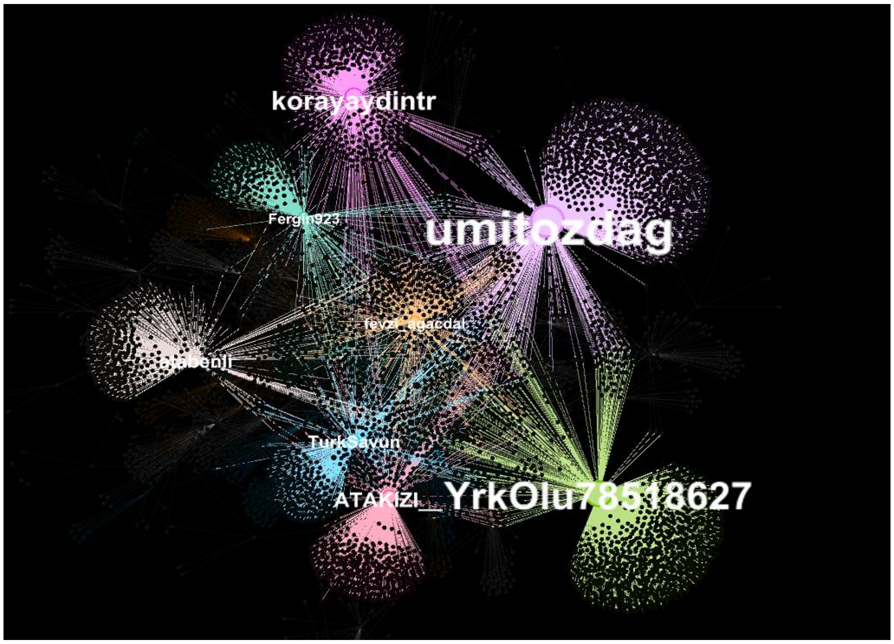
Figure 8 shows the network of disinformation tweets with 6,385 accounts 7,713 interactions.

Figure 8. Disinformation tweeting/Retweeting network.



What is evident from the network analysis, is that disinformation circulates among a network of a few influential users, which are discerned here as big-sized nodes. As seen in Figure 8, influential users have their own community of retweeters, who amplify disinformation against Syrians by retweeting their tweets. What is notable, however, is that these influential users do not share followers or retweeters, which is due to their different political affiliations within the spectrum of the oppositional parties. Indeed, when exploring the accounts of the influential users involved in spreading disinformation as shown in Figure 9, we can see that the list of the top eight most retweeted users included politicians Social media activists, and journalists either affiliated with the Victory (Zafer) Party, Good (İyi) Party and the Republican People's Party (CHP) or represented nongovernmental organizations and news agencies that adopt the Turkish nationalist line such as the Yeni Çağ newspaper. These actors are representative of different parts of the political spectrum and are therefore followed and retweeted by different people.

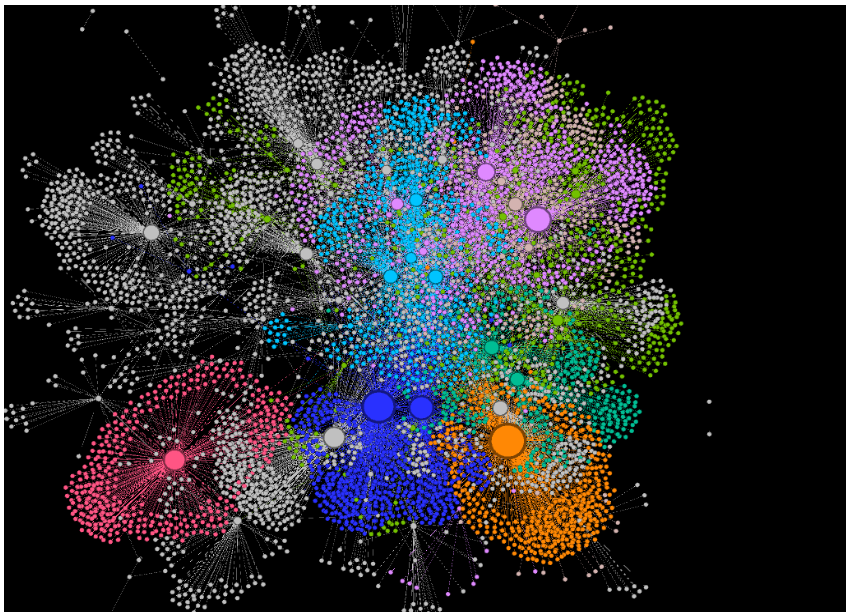
Figure 9. Most influential users in disinformation tweets.



At the same time, while 62% of the influential users had verified accounts and millions of followers, the rest of the accounts (38%) were not verified and had a moderate or small number of followers. Hence, when it comes to spreading disinformation about Syrian refugees, both the author's profile/affiliation and the content of the tweet, irrespective of the verification of the author's identity, play an instrumental role in how viral this tweet might become through likes and retweets.

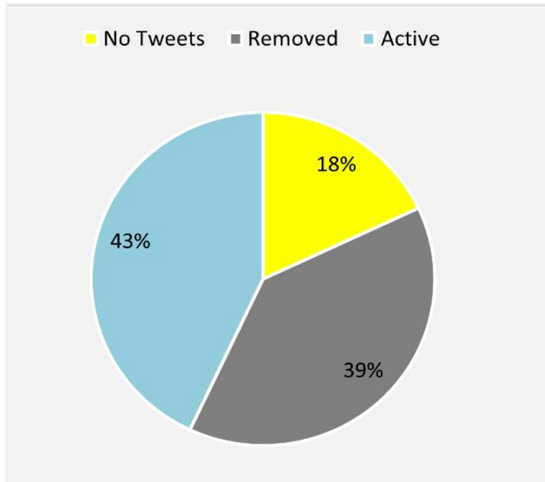
On the other hand, what is remarkable in the hate speech network shown in Figure 10 is its high density and the overlapping nodes and edges. This reflects the fact that, unlike the disinformation network, the influential users here have many retweeting users in common. It seems that almost the same group of users have retweeted and amplified the hateful tweets posted by influential users of this network. This can be, therefore, indicative of an orchestrated tweeting and retweeting activity.

Figure 10. Hate speech tweeting/Retweeting network.



Further supporting this conclusion is the fact that, when exploring the accounts of the influential users after August of 2021, we discovered that more than half the accounts involved in the hate speech network were either deactivated by their owners, removed by Twitter or had erased all their tweets (See Figure 11). If removal from Twitter can be seen as part of the platform's effort to tackle hate speech, the other two practices can be an indication of an orchestrated effort to instigate violence against Syrians. These tweets and related accounts seemingly had no reason to exist after the August 2021 attacks.

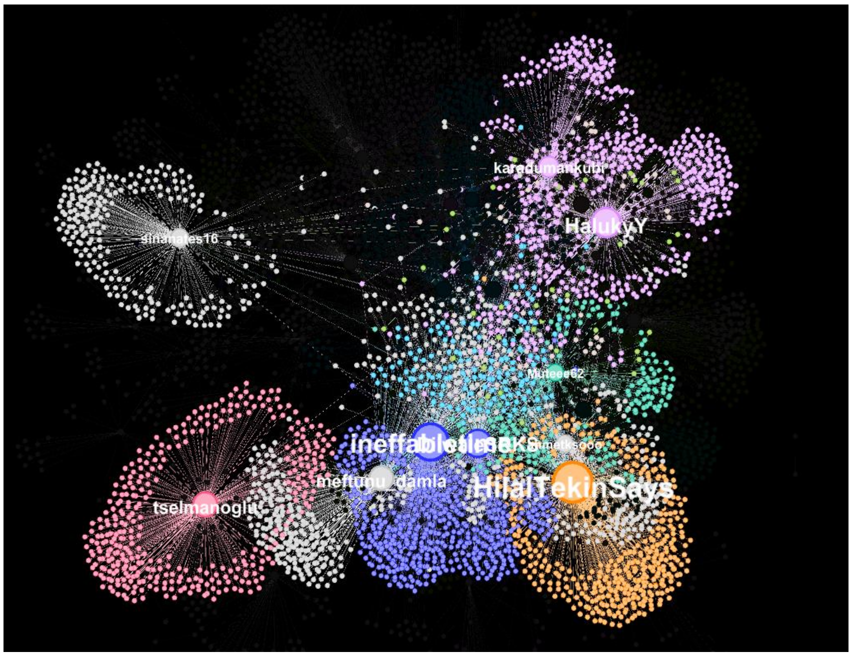
Figure 11. Activity status of hate speech influential accounts.



On the other hand, surprisingly, few influential users in the hate speech network shown in Figure 12 belong to politicians/former politicians affiliated with the opposition parties such as the Victory (Zafer) Party or the Republican People's Party (CHP), while the rest of the influential accounts seem to represent either ordinary Twitter users not affiliated with any political party or social media influencers (one of them was found to be supporting the Justice and Development Party (AKP)) besides journalists/activists who represent private media agencies and civil society associations such as Ülkü Ocakları (Grey Wolves) organization affiliated with Milliyetçi Hareket Party (MHP), People's Municipalities (Halkın Belediyeleri) news agency that echoes the ideas of both the Republican People's Party (CHP) party and Good (İyi) Party, and The Silent Occupation (Sessiz İşgal) Youth Association and Voices from Home (Yurtten Sesleri) news agency which is affiliated with Victory (Zafer) Party and the Republican People's Party (CHP) party, respectively. The tweets/retweets made by most of these accounts have espoused the discourse of the opposition toward Syrians while spreading hate speech against them.

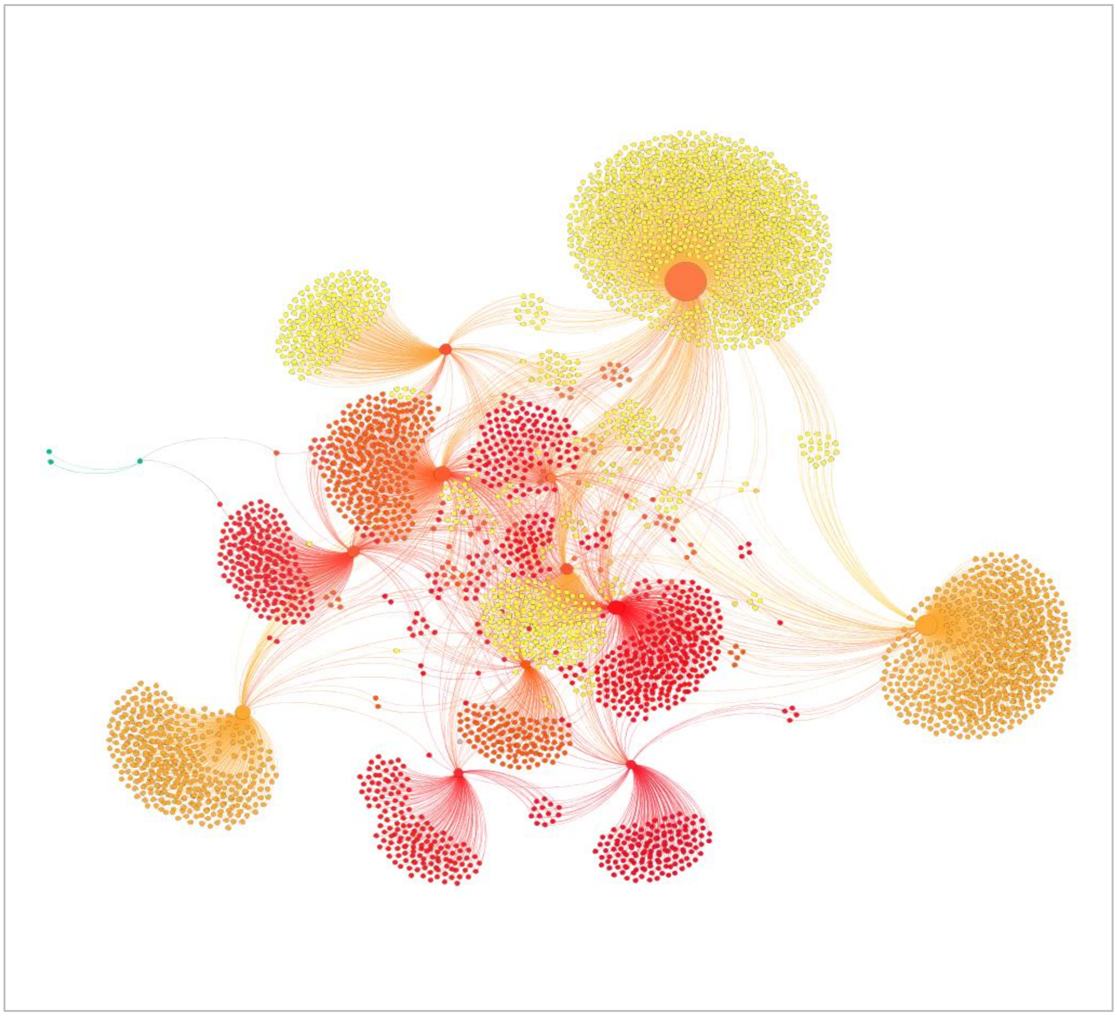
Moreover, when exploring their following/follower lists, we found that 66% of hate speech influential accounts are followers or friends of the disinformation influential accounts identified in Figure 9 while 32% of hate speech influential users retweeted the tweets posted by disinformation influential users.

Figure 12. Most influential users in Hate Speech tweets.



The interaction between influential users in the hate speech network and disinformation network was further confirmed when merging the two networks as shown in Figure 13, where the red clusters refer to the hate speech influential user communities, the yellow clusters indicate the disinformation influential user communities while orange clusters/nodes denote the communities of users who were influential/retweeters in both disinformation and hate speech networks.

Figure 13. Disinformation-Hate speech merged tweeting/retweeting network.



A closer look at the influential users within the merged hate speech and disinformation networks allows us to draw further conclusions. As seen in Figure 13, the influential users who were involved in hate speech are also influential users in the disinformation network and/or retweeters of disinformation influential users. This confirms the interplay between disinformation and hate speech, illustrating that investigating hate speech actors and their propagation strategies cannot be conducted separately from the actors and amplification mechanisms identified in disinformation tweets. Although it is important to think about hate speech and disinformation as different discursive phenomena attacking Syrians, it is also crucial to see their interplay, which is crucial in reproducing online toxicity against refugees, ultimately inciting violent acts against them in consistency with specific political interests and agendas.

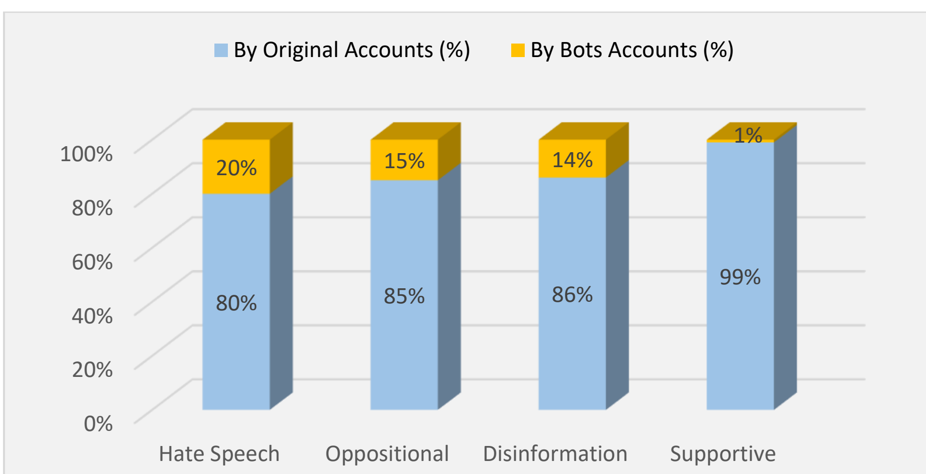
Finally, to further investigate the amplification of disinformation and hate speech within our sample, we performed a deep analysis of the accounts involved in retweeting the disinformation and hate speech content. This was done to check whether the propagation of disinformation and hate speech was spontaneous or further organized and propagated by swarms of fake accounts, the so-called bots. To this end, we performed the following steps. First, we spotted the users whose retweets hit great values in each of the disinformation and hate speech tweets. Out of the most retweeting users, we selected the users whose account properties met those that describe bots, as bot accounts usually have few followers, their owners have recently joined Twitter, have no information in the bio section, or have no avatar or cover photo (Yang, Ferrara, & Menczer, 2022). Later, we fed the accounts identified from the previous steps into a machine-learning model trained to detect bot accounts.

As a result, we spotted 970 accounts as potential bots. All these users had recently joined Twitter, particularly right after the unrest between Turks and Syrians in Altındağ Ankara. Also, most of these accounts (78%) had very few or no followers at all. Finally, all these users followed the same accounts that were identified as influential users in either disinformation or hate speech tweets. Consequently, these accounts represent ideal bots. To confirm that, we subjected each of the potential bot accounts to an off-the-shelf application developed by (Yang et al., 2022) where we found that out of the 970 studied accounts, 809 accounts were identified as bots or fake accounts. Hence, it could be deduced that the anti-Syrian hashtags that were trending as part of disinformation and hate speech campaigns on Twitter were not just spontaneous expressions of public resentment. On the contrary, they were further supported by what seemed to be anti-Syrian propaganda, consisting of influential users and their followers from the bot account swarms. Figure 14 illustrates an example of the report produced for a bot account.

Figure 14. A sample result report for a bot account.

Also, we investigated the quantitative impact of bots on the studied tweet dataset where we found that 15% of the tweets were posted by bot accounts. However, the contribution of these accounts for disinformation, hate speech, and oppositional tweet categories is shown in Figure 15.

Figure 15. Percentages of tweets posted by bot accounts for each tweet category.



Conclusion

In this paper, we investigated Syrian-related narratives on Turkish Twittersphere. We went beyond classifying the attitudes toward Syrians to empirically show the distinction but also the interplay between online hate speech and disinformation and how they are both encapsulated under the concept of online toxicity. This was practically conducted by introducing a novel framework that utilizes a machine learning-based sentiment analysis model along with Natural Language Processing (NLP) techniques to collect and investigate more than 30K Turkish tweets and recognize both hate speech and disinformation tweets.

Our empirical findings construct a bleak image of the overall role of Twitter in the portrayal of Syrians on Turkish Twittersphere. Despite the potential of social media as an alternative space of representation, which can afford a voice to migrants and challenge hostile mainstream discourses, our research confirms earlier scholarship that illustrated social media as echoing rather than subverting negative portrayals of migrants and refugees. Employing the concept of online toxicity as a broader conceptual framework, we argued that it encompasses disinformation and hate speech as distinct but also interacting phenomena. We set out to study such negative discourses through a big data study of a sample of more than 30,000 tweets about Syrians in the Turkish language.

Our analysis proved the overwhelming presence of online toxicity against Syrians on Turkish Twitter. A bit more than 30% of these negative tweets were examples of hate speech, attacking Syrians and constructing them as inferior to the host population. A similar number of 31% among the negative tweets were reproducing disinformation against Syrians, accusing them of unnecessarily leaving their homeland and ripping off benefits provided in Turkey. The analysis indicated that there was an increase in hate tweets against Syrians during specific time frames, namely while the refugees were visiting Syria during the holidays and, most importantly, during the violent attacks against them in Ankara in August 2021. On the other hand, we believe that, if we had enough tweets that contain both disinformation and hate speech, the interplay between hate speech and disinformation would have been investigated more deeply at the linguistic level in terms of how certain claims about Syrians can incite hatred against them.

At the same time, we also identified the Twitter networks circulating and amplifying these hate speech and disinformation tweets. We found that disinformation was most often instigated by political actors associated with the opposition, amplified by their followers among which there was little overlap. Most of the hate speech actors in the network, on the other hand, had either deleted their accounts, whether forcibly or voluntarily, or their relevant tweets attacking Syrians. There was considerable overlap between the remaining hate speech accounts and those (re)producing disinformation, indicating the interplay between the two forms of online toxicity. Finally, we identified that a considerable number (809) accounts in our sample, responsible for (re)tweeting and spreading hate speech and disinformation, were actually bots. This, we argue, is another illustration of the orchestrated attempts to undermine the presence of Syrians in Turkey.

Partially, these empirical findings are particular to Turkey, where the debate about Syrians has been highly politicized and has become a clear point of contention between the government and opposition parties. The intensity of these debates is such that politicians from the opposition are often the ones instigating disinformation against Syrians on Twitter. The background of this intense political polarization is a deepening economic crisis in Turkey, which has rendered Syrians evident scapegoats for populist politicians and the host population. The consequences of these populist arguments are experienced by Syrians, who are faced with increasing hostility in the country.

These findings are also complementary to extant research in other countries that have illustrated social media as spaces that amplify and normalize negative sentiment against refugees and migrants (Aslan et al., 2017; Georgiou, 2018). The methodological design of our big data analysis, however, allowed us to move beyond critical analysis of media discourses and questions of representation. We illustrated how Turkish Twitter not only reflects and reproduces symbolic hierarchies of belonging that are ostensible in mainstream media and within Turkish society (Güney, 2021), but it does so in ways that are orchestrated and operationalized by influential social media actors. We, therefore, argue that to fully understand the dimensions and consequences of online toxicity and hate speech, we need to see them concerning disinformation and the ways they are politicized in specific sociopolitical contexts.

References

Akın, A. A., & Akın, M. D. (2007). Zemberek, an open source NLP framework for Turkic languages. *Structure*, *10*(2007), 1–5.

Alkiviadou, N. (2019). Hate speech on social media networks: towards a regulatory framework? *Information & Communications Technology Law*, *28*(1), 19–35.

Alp, H. (2018). Suriyeli sğnmaclara yönelik ayrmc ve ötekileştirici söylemin yerel medyada yeniden üretilmesi. *Karadeniz Teknik Üniversitesi \Iletişim Fakültesi Elektronik Dergisi*, *5*(15), 22–37.

Aridici, N. (2022). Constructing the ‘National Ideal’: The ‘Inclusive’and ‘Exclusive’Representations of Syrian Refugees in Turkish Print Media. *Journal of Balkan and Near Eastern Studies*, 1–19.

Aslan, A., & others. (2017). Online hate discourse: A study on hatred speech directed against Syrian refugees on YouTube. *Journal of Media Critiques*, *3*(12), 227–256.

Assimakopoulos, S., Baider, F. H., & Millar, S. (2017). *Online Hate Speech in the European Union:A Discourse-Analytic Perspective*.

Aydınlı, S. (2020). HATE AND DISINFORMATION NARRATIVES IN THE ONLINE MEDIASCAPE IN TURKEY. *RESILIENCE : For Media Free of Hate and Disinformation*, 1–23.

Az, İ., Gelişli, N., Barak, R., & Arslan, Z. (2017). *media watch on hate speech report: January-April 2015. Istanbul: Hrant Dink Foundation*.

BBC-News. (2021). *Turkish capital reels from violent protests against Syrians* (bbc.com, Ed.).

Bozdağ, Ç. (2020). Bottom-up nationalism and discrimination on social media: An analysis of the citizenship debate about refugees in Turkey. *European Journal of Cultural Studies*, *23*(5), 712–730.

Chouliaraki, L., & Stolic, T. (2017). Rethinking media responsibility in the refugee 'crisis': A visual typology of European news. *Media, Culture & Society*, *39*(8), 1162–1177.

Erdoğan, E. (2021). *How Syrian refugees became victims on Turkey's "fake news" front line* (T. G. M. F. of the United States, Ed.).

Erdogan-Ozturk, Y., & Isik-Guler, H. (2020). Discourses of exclusion on Twitter in the Turkish Context:# ülkemdesuriyeliistemiyorum (# idontwantsyriansinmycountry). *Discourse, Context & Media*, *36*, 100400.

Farooq, U. (2021). *Uncertainty for Syrians in Turkey as opposition warms to Assad* (aljazeera.com, Ed.).

Georgiou, M. (2018). Does the subaltern speak? Migrant voices in digital Europe. *Popular Communication*, *16*(1), 45–57.

Güney, Ü. (2021). Syrian Refugees between Turkish Nationalism and Citizenship. *Journal of Immigrant & Refugee Studies*, 1–14.

Hameleers, M., van der Meer, T., & Vliegenthart, R. (2021). Civilized truths, hateful lies? Incivility and hate speech in false information–evidence from fact-checked statements in the US. *Information, Communication & Society*, 1–18.

Henry, S., Cuffy, C., & McInnes, B. T. (2018). Vector representations of multi-word terms for semantic relatedness. *Journal of Biomedical Informatics*, *77*, 111–119.

Hübl, F., Cvetojevic, S., Hochmair, H., & Paulus, G. (2017). Analyzing Refugee Migration Patterns Using Geo-tagged Tweets. *ISPRS International Journal of Geo-Information*, *6*. https://doi.org/10.3390/ijgi6100302

Karabat, A. (2018). *How Syrian refugees became victims on Turkey's "fake news" front line* (M. E. Eye, Ed.).

Karakas, B. (2021). *Refugees in Istanbul: Is anti-migrant sentiment growing in Turkey* (dw.com, Ed.).

Kim, J. W., Guess, A., Nyhan, B., & Reifler, J. (2021). The distorting prism of social media: How self-selection and exposure to incivility fuel online comment toxicity. *Journal of Communication*, *71*(6), 922–946.

Kreis, R. (2017a). # refugeesnotwelcome: Anti-refugee discourse on Twitter. *Discourse & Communication*, *11*(5), 498–514.

Kreis, R. (2017b). # refugeesnotwelcome: Anti-refugee discourse on Twitter. *Discourse & Communication*, *11*(5), 498–514.

Marwick, A., & Lewis, R. (2017). Media manipulation and disinformation online. *New York: Data & Society Research Institute*, 7–19.

Masullo Chen, G., Muddiman, A., Wilner, T., Pariser, E., & Stroud, N. J. (2019). We should not get rid of incivility online. *Social Media+ Society*, *5*(3), 2056305119862641.

Mondal, S. K., Sahoo, J. P., Wang, J., Mondal, K., Rahman, M., & others. (2022). Fake News Detection Exploiting TF-IDF Vectorization with Ensemble Learning Models.

In *Advances in Distributed Computing and Machine Learning* (pp. 261–270). Springer.

Nockleby, J. T. (2000). *Hate Speech* (Vol. 1). Encyclopedia of the American Constitution (2nd ed., edited by Leonard W. Levy, Kenneth L. Karst et al. New York: Macmillan.

Ozduzen, O. (2020). 'We are not Arabs and Taksim is ours' YouTubed political agency, place-making and social exclusion of Syrians. *City*, *24*(5–6), 741–758.

Özerim, M. G., & Tolay, J. (2021). Discussing the populist features of anti-refugee discourses on social media: an anti-Syrian hashtag in Turkish Twitter. *Journal of Refugee Studies*, *34*(1), 204–218.

Secen, S., Al, S., & Arslan, B. (2023). *Electoral Dynamics, Nationalist Ideology, and Party Positions on Syrian Refugees in Turkey*.

Şen, F. (2017). Bir öteki olarak mülteciler: suriyeli mültecilerin anaakim ve alternatif medyada temsili. *Atatürk \.Iletişim Dergisi*, *12*, 27–42.

Sheth, A., Shalin, V. L., & Kursuncu, U. (2021). Defining and detecting toxicity on social media: context and knowledge are key. *Neurocomputing*.

Siapera, E. (2019). Refugee solidarity in Europe: Shifting the discourse. *European Journal of Cultural Studies*, *22*(2), 245–266.

Siapera, E., Boudourides, M., Lenis, S., & Suiter, J. (2018). Refugees and network publics on Twitter: Networked framing, affect, and capture. *Social Media+ Society*, *4*(1), 2056305118764437.

Stahl, K. (2018). Fake news detection in social media. *California State University Stanislaus*, *6*, 4–15.

Udwan, G., Leurs, K., & Alencar, A. (2020). Digital resilience tactics of Syrian refugees in the Netherlands: Social media for social support, health, and identity. *Social Media+ Society*, *6*(2), 2056305120915587.

UNHCR. (2021). *UNHCR Turkey - Fact Sheet September 2021* (U. N. H. C. for Refugees, Ed.).

Vidgen, B., & Yasseri, T. (2020). Detecting weak and strong Islamophobic hate speech on social media. *Journal of Information Technology & Politics*, *17*(1), 66–78.

Wardle, C., Derakhshan, H., & others. (2018). Thinking about 'information disorder': formats of misinformation, disinformation, and mal-information. *Ireton, Cherilyn; Posetti, Julie. Journalism,'Fake News'& Disinformation. Paris: Unesco*, 43–54.

Waseem, Z., Davidson, T., Warmsley, D., & Weber, I. (2017). Understanding abuse: A typology of abusive language detection subtasks. *ArXiv Preprint ArXiv:1705.09899*.

Winter, A. (2019). Online hate: from the far-right to the 'alt-right'and from the margins to the mainstream. In *Online othering* (pp. 39–63). Springer.

Yang, K.-C., Ferrara, E., & Menczer, F. (2022). Botometer 101: Social bot practicum for computational social scientists. *ArXiv Preprint ArXiv:2201.01608*.

Yücel, A. (2021). Symbolic annihilation of Syrian refugees by Turkish news media during the COVID-19 pandemic. *International Journal for Equity in Health*, *20*(1), 1–11.

Zhao He and Chen, X. and N. T. and H. J. Z. and W. G. and C. H. (2016). Stratified Over-Sampling Bagging Method for Random Forests on Imbalanced Data. In G. A. and C. H. Chau Michael and Wang (Ed.), *Intelligence and Security Informatics* (pp. 63–72). Springer International Publishing.