World Development 168 (2023) 106268

Contents lists available at ScienceDirect

World Development

journal homepage: www.elsevier.com/locate/worlddev

Displacement and return in the internet Era: Social media for monitoring migration decisions in Northern Syria

Erin Walk [a,*], Kiran Garimella [b], Fotini Christia [c]

[a] MIT IDSS, United States
[b] Rutgers School of Communication and Information, United States
[c] MIT Political Science and MIT IDSS, United States

ARTICLE INFO

Article history:

Keywords:
Social media
Displacement
Refugee return
Syria
Migration
Civil War

ABSTRACT

According to UNHCR reporting there are over 27 million refugees globally, many of whom are hosted in neighboring countries which struggle with bureaucracy and service provision to support them. With the onset of Covid-19 in early 2020, gathering data on the location and conditions of these refugees has become increasingly difficult. Using Syria as a case study, where since 2011 80% of the population has been displaced in the civil war, this paper shows how the widespread use of social media could be used to monitor migration of refugees. Using social media text and image data from three popular platforms (Twitter, Telegram, and Facebook), and leveraging survey data as a source of ground truth on the presence of IDPs and returnees, it uses topic modeling and image analysis to find that areas without return have a higher prevalence of violence-related discourse and images while areas with return feature content related to services and the economy. Building on these findings, the paper uses mixed effects models to show that these results hold pre- and post-return as well as when migration is quantified as monthly population flows. Monitoring refugee return in war prone areas is a complex task and social media may provide researchers, aid groups, and policymakers with tools for assessing return in areas where survey or other data is unavailable or difficult to obtain.

1. Introduction

Following the onset of the civil war in 2011 Syria has seen unprecedented levels of displacement, making it the largest source of refugees worldwide from 2014 to present. As the war continues host nations, even those that had followed an open-border policy (such as Turkey, Jordan, and Lebanon), have seen their citizens become wary if not outright hostile toward refugees (Kirisci, 2014; Nielsen, 2016; Aktas et al., 2021; Getmansky et al., 2018; Yahya et al., 2018), renewing calls for return. However, the onset of COVID-19 and its spread in crowded refugee camps has only increased the difficulty of monitoring displacement and return. How can aid organizations and governments leverage novel data sources to understand these trends at smaller time scales and areas? In Syria, the first conflict in which "the lines between offline and online engagement have blurred," (Gohdes, 2020, 489) social media data reveals information about circumstances on the ground, such as ongoing violence and economic opportunity, that are linked to return considerations.

In exploring how social media data reflects return decisions, we expand upon a broad base of research analyzing attitudes around return with surveys (Ghosn et al., 2021; Alrababa'h et al., 2020; Krishnan et al., 2020; Balcilar & Nugent, 2019). Prior work emphasizes the role of security, the economy (Camarena and Hagerdal, 2020), family ties (Ghosn et al., 2021), and access to information (Alrababa'h et al., 2020) in motivating return decisions. This paper is, to our knowledge, the first exploration using observational social media text to monitor discussion of the factors which contribute to displacement and return. Drawing on research in the event data community, such as that on detecting violent events (Raleigh et al,. 2010) and protests (Zhang & Pan, 2019), as well as predicting displacement in other contexts (Abrishamkar et al., 2018; Singh et al., 2019), we use social media as an indicator of whether locations have experienced population displacement or return. The paper draws on original social media data from the three most popular social media platforms in Syria: Twitter, Telegram, and Facebook.

To identify discussions and conditions that reflect internal displacement and return we use unsupervised machine learning (ML) methods to cluster topics of interest from the text and image data. We also run 'seeded' topic models, models run on subsets of

* Corresponding author.
E-mail address: ewalk@mit.edu (E. Walk).

https://doi.org/10.1016/j.worlddev.2023.106268

the data filtered for keywords related to displacement, local governance, service provision, and return. Using insights from the topic models we run mixed effects models to identify changes in topic discussion post-return. We find that discussions of violence account for 15% more of overall discussion in areas with neither returnees nor IDPs. Such discussions include notifications about air strikes and troop movements. In areas with returnees and IDPs discussion of infrastructure and the economy is 2% more prevalent. Images posted in return groups are more likely to be goods for sale including cars, motorcycles, and other miscellaneous items (indicating active trade and commerce, 3.3% more prevalent in areas with only returnees than in areas with neither IDPs nor returnees). In non-return areas, images are more likely to be violence-related including militants and tanks (+1.4%). These results hold when accounting for correlation between topics and adding time and location effects, as shown using mixed effects models. Furthermore, using more granular data on return and displacement flows we show that increased return is positively associated with economic topics, while increased displacement is positively associated with violence topics.

In summary, our paper makes the following contributions: First, it shows that social media data, including images and messages, can be used to distill location-specific factors which are linked with displacement and return including discussion of the economy and ongoing violence. Second, we find that discussion of these factors differs at a statistically significant level in areas with and without returnees and IDPs, revealing the potential of social media as a monitoring tool for identifying returnee and IDP populations. Third, we combine text analysis with image analysis to see to what degree images dovetail with the discourse and what additional information we can extract from this medium.

The rest of the paper is organized as follows: First, we discuss the context for our project including relevant literature on displacement and attitudes around return, the situation in Syria, and social media use. Next, we outline our research design and data collection, including procedures for choosing groups as well as text and image analysis methods. We then present our results and discussion, outlining our findings on return and internal displacement from topic models, images, and mixed effects. We conclude with discussion and suggestions for future work.

2. Theories of displacement and return

Identifying effective ways to measure return practices is not straightforward as there are no systematic data sources on refugee return, and many current reports rely on surveys to test the underlying assumptions governing return decisions (Ghosn et al., 2021). Such surveys are essential for teasing out small differences in outcomes based on personal experience, yet difficult to use for consistent monitoring at small timescales. Due to the ubiquity of social media in the Syrian conflict (Lynch et al., 2014; Gohdes, 2020), online discussion is a ripe source of information on ongoing events which could impact displacement and return.

In order to pinpoint which social media conversations are linked with return it is vital to understand what factors lead to population movement. Though refugees may be displaced from their home country for many reasons including economic, political, and environmental factors (Arias et al., 2014; Lischer, 2006; Ruegger,2013; Davenport et al., 2003; Zolberg et al., 1989; Martin et al., 2019), violence is often considered the key driver of displacement (Davenport et al., 2003; Zolberg et al., 1989). In a 2020 CARE survey conducted through interviews with Syrian IDPs, 99% said they had been displaced due to violence or fighting (Hoffman & Makovsky, 2021). Other survey studies have found that extended violence can decrease the likelihood that individuals will return to their home country (Balcilar & Nugent, 2019). The impact violence has on migration is related to both the type of violence and individuals' experiences. Through interviews with Syrian refugees, Schon (2019) found that individuals who witness, but do not directly experience, violence delay fleeing their home country if they receive community support and experience post-traumatic growth. State sponsored violence and genocide are more likely to lead to refugees whereas civil wars are associated more highly with internal displacement (Steele, 2019; Moore & Shellman, 2006).

Though ongoing violence is often linked to displacement (Singh et al., 2019; Davenport et al., 2003) it is also an important deterrent to return. Willingness to confront danger on the path to return may be impacted by refugees' religious beliefs as well as experiences of violence prior to return. Strong religious beliefs increase both the likelihood that individuals stay in their home communities even in the presence of violence and the likelihood that they will encourage others to return (Kaplan, 2021). Similarly, experiences of violence may impact an individual's belief in their ability to adapt to circumstances on the ground (Ghosn et al., 2021). Refugees who spend more time in their home country before fleeing are thus more likely to return despite having experienced violence. Disparities in fear of and experiences around violence may alter the association between online discussion of violence and refugee return.

Aside from security considerations, return decisions are influenced by economic opportunities, services, and social networks (Aymerich & Zeyneloglu, 2019; Hoogeveen et al., 2019; Arias et al., 2014). In a survey study conducted on 3,003 displaced Syrian families in Lebanon, Alrababa'h et al. (2020) weigh the influence of "pull" factors, which draw individuals back to their former homes, and "push" factors which encourage them to leave their host communities. In doing so they find that conditions in a refugee's home country and access to information about those conditions are primary drivers in decision making (Alrababa'h et al., 2020). Overall desire to return is linked to strong ties with one's home country and hometown (Ghosn et al., 2021), and refugees who are single and male are more likely to return to Syria to reunite with their families (World Bank, 2019).

Though ties are important for resettlement, refugees with strong ties to their home country may still choose to return as regular visitors rather than permanently if their new residence offers more attractive economic opportunities (Camarena and Hagerdal, 2020). Emphasizing the importance of such opportunities, many young and educated Syrians in the capital city of Damascus indicate that they would choose to emigrate if given the opportunity (Jalabi, 2021). However, economic conditions in host countries are also challenging, with many refugees working informally and below minimum wage (Kumar et al., 2018). Access to services increases returns (World Bank, 2019; Elbadawi et al., 2019), a feature which may be linked with security and governance such as through UN peacekeeping missions (Bove et al., 2021). In Syria local governments and militias may provide similar services in certain regions, though in regions bordering Turkey it is often overseen by the governors of neighboring Turkish provinces (Hoffman & Makovsky, (2021; Al-Hilu, 2019). Conversely, good service provision in the host country may also decrease desire to return (Balcilar & Nugent, 2019). Regardless, infrastructure and economic opportunities are a vital factor in motivating return decisions.

Given the importance of information for decisions around return (Alrababa'h et al., 2020) and gauging living conditions in the home country, leveraging social media sources is of critical relevance because they offer an immediate assessment of local discussion as well as a way to monitor how refugees access and consume information pertinent to their return. Recent work in the event data community has emphasized the potential of social media data for detecting events such as protests (Zhang & Pan, 2019), acts of violence (Raleigh et al., 2010), and displacement (Singh et al., 2019; Abrishamkar et al., 2018). Singh et al. (2019) rely on Twitter data from Iraq to assess when violent events are taking place using sentiment analysis of tweets with the hashtag "ISIS". In combining this with traditional movement variables, they improved the accuracy of displacement predictions. To reveal how the relative importance of different displacement conditions varies depending on region and political climate, they expand this to a Spanish language model focusing on economic issues in Venezuela (Singh et al., 2020). Similarly, Abrishamkar et al. (2018) used information from news articles as an indicator of violence in order to predict displacement. Drawing on this work, we seek to apply these insights in the Syrian context to show the potential of social media for differentiating between areas with displacement and return by looking at conversations related to violence, infrastructure, and the economy.

3. Case: The Syrian civil war

Since the start of the Syrian civil war in March 2011, Syria has seen unprecedented levels of internal displacement, movement of refugees, and civilian casualties. The war began as a protest movement in 2011 to remove President Bashar Al-Assad from power, a movement that coincided with Arab spring protests that removed the Egyptian and Tunisian presidents. By 2012, the protests had evolved into militarized violence. As the war passes its 10th year over 400,000 Syrians have been killed and 13.2 million Syrians are refugees, asylum seekers, and internally displaced people, accounting for one sixth of the global total and 80% of the total Syrian population (UNHCR, 2020). Of these 13.2 million, 6.6 million are registered refugees, about two thirds of which reside in Turkey and five sixths of which reside in countries bordering Syria, the majority not in refugee camps (UNHCR, 2021). Following the defeat of the Islamic State of Iraq and Syria (ISIS) by the International Coalition forces in ArRaqqa in October 2017 the war in Syria seems to be nearing its end.[1] However, though 75% of displaced Syrians would like to one day return to their former homes (UNHCR, 2019), in 2020 only 467,000 Syrian refugees returned home while 1.8 million were newly displaced (Norwegian Refugee Council, 2021). Identifying where these refugees and returnees are located to provide services and aid is a difficult task.

Displacement within Syria has been non-uniform, with some governorates including Deir-ez-Zour, Ar-Raqqa, and Al-Hasekah losing a large share of their population (all over 25%) and others such as Idlib and Rural Damascus gaining inhabitants (World Bank, 2019). In the 2015–2016 Syrian Refugees and Host Communities Surveys (SRHCS) most refugees reported less than a week to prepare to leave, with the majority ending up in a neighboring country (Krishnan et al., 2020). Both the immediacy and unpredictability of displacement and return make Syria a practical case for this work. Many refugees also face lasting mental and physical health outcomes resulting from their experiences (Balcilar et al., 2022). Refugee and internally displaced populations face different difficulties. Though more Syrian men are able to work in Syria, refugees have better access to resources than IDPs and residents in Syrian governorates with a high level of conflict. Conversely, refugee children often have worse access to education. Housing conditions in Syria are poor, especially in Idlib, and Syrians living in all regions are subject to a lack of services including electricity and water. Many returnees lack documents (World Bank, 2019) and 60% of Syrians are food insecure (World Food Programme, 2021). Since 2020 the COVID-19 pandemic has deeply impacted refugee communities. Refugees in host countries were more likely to be living below the poverty line pre-pandemic and refugees were often more highly impacted due to reliance on wage work (World Bank, 2020).

Refugees have complex relationships with their host communities as well as with their home locations, leading to tensions that can make return more desirable. Host countries struggle with service provision and maintaining the requisite bureaucracy to determine refugee legal status, while also facing backlash and anti-refugee sentiment among their own citizens (Lazarev & Sharma, 2017; Alrababa'h et al., 2021; Ghosn et al., 2021; Braithwaite et al., 2019; Bradley, 2013). Turkey, the main host of Syrian refugees as well as refugees globally, has faced a lack of stability and domestic political strife over the influx of refugees since 2011, yet as the situation draws on many seem less likely to return. In 2017 the annual Syrians Barometer poll conducted by Murat Erdõgan indicated that 17% of Syrians would not return under any conditions, and in 2019 this had increased to 52% (Erdogan, 2020). By the most recent poll in 2020, 78% of Syrians indicated that they would not return under any conditions (Erdogan, 2021) despite the challenges of assimilation including lack of language skills and difficulties in having Syrian education and work accreditation recognized in Turkey (Kumar et al., 2018). In order to stem the flow of refugees, and in some instances force return, Turkey has taken strong offensive actions including resettling Syrians in areas which were formerly occupied by the Kurdish ethnic minority. This type of demographic engineering has limited international humanitarian aid through both stifling opportunity and igniting concerns over treatment of Kurds.[2]

3.1. Social media and information access

Beginning with the protests in 2011, social media has played an integral role in the Syrian civil war leading researchers to refer to it as the "most socially mediated civil conflict in history" (Lynch et al., 2014, 5). In the early stages of conflict platforms such as Twitter, YouTube, and Facebook were used by civilian activists to organize and share images of the protests (Freelon et al., 2015). As the protests evolved into war, many of the key players became involved on social media. Extremist groups have used Twitter to spread their ideology (Klausen, 2015; Wei et al., 2016), as well as sectarian hate speech (Siegel & Badaan, 2020; Abdo, 2015), and propaganda (Chatfield et al., 2015), while Syrian Opposition Forces have used Facebook to communicate their war narrative (Crilley (01 Jan. 2017)). The presence of perspectives from all sides of the conflict leads (Gohdes, 2020) to note that "the Syrian conflict is one of the first conflicts where lines between offline and on-line conflict engagement have become blurred" (Gohdes, 2020). Outside of the Syrian context, social media data is increasingly used alongside, or even in place of, public opinion surveys, often with highly similar results (Schober et al., 2016).

In addition to Twitter and Facebook, the role of which have been more widely studied (Khamis et al., 2012; Metzger & Siegel, 2019), we extend analysis to include data from Telegram "an encrypted platform that is harder for governments to monitor" (Mitts, 2019). Much of the research on Telegram has focused on its use by ISIS (Prucha, 2016; McDowell-Smith et al., 2017; Yayla & Speckhard, 2017), though Telegram's privacy policies also support civilians and protest movements worldwide (Urman et al., 2020).

[1] ISIS, also known as ISIL (Islamic State of Iraq and the Levant) is a Sunni jihadist group that claims religious authority over all Muslims and has a particularly violent ideology.

[2] "German NGO scraps Syria project over claims it would aid Turkey's ethnic cleansing in Afrin," https://www.kurdistan24.net/en/story/23475-German-NGO-scraps-Syria-project-over-claims-it-would-aid-Turkey%27s-ethnic-cleansing-in-Afrin.

In addition to the role of social media in constructing violence narratives, it also plays a vital role in migration movement (Frouws et al., 2016; Miconi, 2020; Sanchez-Querubın & Rogers, 2018). Refugees share information about routes and conditions in potential host countries to ease the journey for future displaced peoples (Frouws et al. 2016). In interviews conducted with 44 young refugee or immigrant Syrian social media users, Miconi (2020) finds that these platforms are used for not only staying connected to war developments but also facilitating resettlement in the host country. A similar Facebook specific study notes that members of the diaspora turn to the platform to maintain social ties (Ramadan, 2017). Given the ubiquity of social media in the conflict, it follows that Syrians use social media not just for the war effort or to learn of migration routes but also to communicate local conditions that can be used to monitor displacement and return.

4. Research design and data collection

Our analysis relies on novel data collected from Telegram, Twitter, and Facebook in the latest stage of the war. The time frame for data collection begins after the recapture of Ar-Raqqa from ISIS on 17 October 2017 and ends on 1 December 2020. Group selection processes focus on entities which are located in Syria or primarily discussing issues related to Syria. The success of work on predicting displacement using social media conversations (Abrishamkar et al., 2018) as well as conclusions that social media can be an effective tool for measuring public opinion (Schober et al., 2016) indicate that users discuss salient events and opinions on social media and thus we anticipate differences in discussion in areas with and without return or IDPs. Analysis is limited to messages posted in Arabic, as messages in the native language are more predictive of displacement (Singh et al., 2020). We augment our collection of social media data with traditional indicators gathered from surveys and interviews by the REACH resource center (REACH Resource Center, 2018–2020) and the UN Office for the Coordination of Humanitarian Affairs (OCHA, UN OCHA, 2021).

The REACH resource center presents monthly data on a variety of indicators gathered from interviews with individuals living in communities across Syria starting in 2018 (REACH Resource Center, 2018–2020). Not all communities have data points in every report, but any community which is mentioned at least once is included in our data resulting in a total of 3,548 locations. All communities considered in the paper have community level postal codes per the Syrian census.[3] Our main indicators are whether or not the community has returnees and/or hosts internally displaced people (IDPs) as reported by the community contacts. We use this dataset as ground truth to build our findings from the social media data. However, such information remains limited as there is no indicator for the scale of return. For communities in which return first occurred during data collection we note that month as the 'return date'. There are no communities in the dataset which first hosted IDPs during the data collection period.

Return communities from the REACH data are broadly similar across metrics from the 2004 census (Central Bureau of Statistics, 2004) such as population, distance from the nearest large city in the sub-district, distance from the border, agricultural employment, and primary ethnicity as noted in Appendix Table A3 and A4. Areas with only IDPs have a slightly lower average population in Al-Hasakeh, Aleppo, and Idlib. Though census features are likely outdated after 11 years of war, there is not a more recent official data source for population comparison. The distances are calculated using border and city coordinates from the census and Syria shape files.

To supplement the REACH data we use data from the UN Office for the Coordination of Humanitarian Affairs (OCHA) on population flows, which is aggregated data collected by several humanitarian aid partners (UN OCHA 2021). In total this dataset contains information on 2,882 different locations, of which 1,550 have at least one month of return information and 1,983 have at least one month of displacement information. A map of these locations is available in Appendix Figure A6. Out of 24 available months between January 2019 and the end of our data collection in December 2020, the average location appears three times in the returnee data and six times in the displacement data.

4.1. Group selection on social media

To examine refugee displacement and return on these platforms, we created parallel processes to identify, collect data from, and analyze accounts, channels, and groups posting public messages about the Syrian conflict on Twitter, Telegram, and Facebook. While this process aimed to limit selection effects to the extent possible, the different natures of the platforms ultimately mean that our samples comprise different populations, and the differences between search functionalities meant there was no single starting point. Underscoring the replicability of our process — and the process's resemblance to other selection processes common in the use of social media data — we present source specific models in Appendix A.2.2 to show the selection effects inherent in the study of social media users on any singular platform. We limited our analysis to Arabic language, the predominant language of use among all actors in the conflict.[4] Our final dataset contains messages in Arabic from 657 public channels and groups on Telegram, 2,106 public Twitter accounts, and 2,124 public Facebook groups and pages.

4.2. Identifying Location specific messages

The first step of our process was attributing message sets to certain locations. Twitter, Telegram and Facebook's functionalities make it difficult to verify the location that users post from in all but a few cases. To understand location specific discussion we searched all of the collected messages for 2004 Syrian census locations using string matching on the names in Arabic. We augmented this dataset with locations from the GeoNames API,[5] which includes common misspellings and dialectic spellings of locations.[6]

The Location Mentions dataset, comprised of the location mentioning messages associated with the REACH data, includes messages from 770 unique locations, 54 of which are mentioned more than 10,000 times, 229 of which have returnees, and 129 of which have IDPs. Maps with locations mentioned by returnee status and year of return are in Appendix Figure A6. The majority of locations mentioned in the Location Mentions dataset are in northeastern Syria in Idlib, Aleppo and northern Ar-Raqqa, an outcome of REACH data emphasis, as can be seen in Figure 1. For 218 of

[3] The areas which appear in the dataset are available here: https://docs.google.com/spreadsheets/d/12eFw5iS24saWNcQGDVJYYsKtJX-cZ0O-bZhATEhGEMs/edit?usp=sharing

[4] To truncate the dataset to Arabic, we first obtained the language of a message using *langid* (Lui & Baldwin, 2012), an off the shelf tool for detecting language from text. The tool is based on pre-trained machine learning models and can detect over 90 languages. We performed basic cleaning on the text data using Nielsen's stemmer (Nielsen, 2017), while preventing the stemming of proper nouns like key locations and political figures. Next, we removed stopwords using a base list of Arabic stop words and some words in Syrian dialect (my brother, where, why, how, etc.), and the words 'channel', 'subscribe', 'Telegram' and 'Twitter'. The base list of Arabic stop words is from Mohatahar Arabic Stopword, *Github*, https://github.com/mohataher/arabic-stop-words/blob/master/list.txt

[5] http://www.geonames.org/

[6] We removed any locations which had the same names as governorates to improve granularity. A few top locations which were disproportionately represented due to false positive matches, 'mil', 'san', and 'ada', were removed.

these locations return first occurred during the time of data collection, and this set of messages was used to create the Return Date dataset. Of the full set of locations, 119 have both returnees and IDPs. In the overall REACH data it is also true that most locations with IDPs also have returnees (26% of locations have IDPs and no return). Table 1 shows the size of the datasets and the percentage of messages within that dataset that are from areas with returnees and IDPs. For example, in the main Location Mentions dataset there are 2,360,559 messages of which 1,264,871 are from Telegram. Of the total messages, 16% (370,345 messages) are from areas with only return, 5% are from areas with only IDPs, 62% are from areas with neither and 17% are from areas with both. Of the messages from locations where return occurred during the time of data collection, the Return Date dataset, 53% (411,694 messages) are from pre-return and 47% are from areas post-return.

We also compiled a set of groups which mention location information in the name of the group or group description (the Researcher Coded dataset). Such groups include @National.Defense.in.maharda on Facebook,[7] @newsmanbij on Twitter[8] and @saraqib2017n on Telegram.[9] We use this smaller set of 173,906 messages as a robustness check for our results since it includes information we may be missing by using messages with location mentions. For example, in sales messages individuals might not include their location if they are already posting in a group where the location is apparent, and thus such messages would not appear in the Location Mentions and Return Date datasets but would in the Researcher Coded dataset. For the image analysis we used all images posted in the Researcher Coded dataset groups, a final image set of 24,304 images. Of these images, 921 are from areas with both IDPs and returnees, 9,001 are from areas with only returnees, and 14,382 are from areas with neither. None of the Researcher Coded groups are from areas with only IDPs.

To better understand displacement and return flows rather than solely IDP settlement we incorporate monthly data from OCHA which is compiled from a variety of humanitarian actors following methodology agreed to by an inter-agency IDP task force (UN OCHA, 2021). Though this dataset, as discussed above, is somewhat smaller than the REACH dataset– especially in terms of return– it has increased granularity regarding the size of return and displacement movements. We associated the data with location by postal code for the seeded return and displacement datasets resulting in a displacement percentage dataset of 41,434 messages from 316 locations and a return percentage dataset of 62,434 messages from 184 locations.

4.3. Methods

Our primary results rely on unsupervised tools for text and image analysis. For text analysis, we use structural topic modeling (Roberts et al., 2014) and seeded topic models using words from word2vec (Mikolov et al., 2013). For image analysis, we use feature extraction and clustering to obtain 'visual topics.' Structural topic models identify topics in text data, assigning a probability of belonging in a topic to each word. We run two unseeded models: the Location Mentions model with a four-way variable identifying locations as 'IDP only,' 'Returnee Only,' 'Both,' and 'Neither,' and the Return Date model with a dichotomous return variable for pre- and post-return. We allow topic content and proportion to vary with the given variable.

The results below display our analysis on models with 20 topics, chosen to maximize topic coherence both mathematically and through observation of models with different topic numbers (Mimno et al., 2011). We further combine these 20 topics into four larger topic areas: 'Violence', 'Infrastructure, economy', 'News, politics', and 'Other.' Topics within these areas can be seen in Table 2. The prevalence images, such as Figure 2, display regression results where the independent variable is return status of the location and the dependent variable is the topic proportion with week fixed effects. For all structural topic model analysis, expert annotators labelled each topic by looking at the 30 most salient keywords in that topic both in terms of frequency "F" and frequency and exclusivity "FREX". Throughout our results, we present the English translations of the topic labels. For the image analysis we use ResNet-50 (He et al., 2016), a pre-trained convolutional neural network model trained on the ImageNet dataset (Deng et al., 2009) to extract image features and then cluster them using k-means clustering. Our final analysis uses 30 clusters.

The mixed effects models include results from 'seeded' topic models obtained by using a subset of messages filtered for keywords about services, local governance, or displacement. The goal of these seeded models is to pull out additional insights which may be obscured in the Location Mentions and Return Date models. For seeded analysis, we worked with a Syrian researcher who provided an initial list of words focused on services, local governance, return, and displacement available in Appendix Figure A.14. We then used word2vec (Mikolov et al., 2013) with Continuous Bag of Words (CBOW) embeddings to find the 20 most contextually similar terms for each keyword.[10] The resulting list was manually filtered for relevance and the expanded keyword list was used to collect all messages that included any of the words. The seeded return model includes 672 locations with 477,581 messages from areas without return and 419,090 messages from areas with return, and the seeded displacement model includes 583 locations with 55,737 messages from areas with IDPs and 125,158 messages from areas without. Both models were run with 15 topics.

4.4. Mixed effects modeling

For the Return Date model as well as the seeded return model we run a mixed effects model to identify changes in topic prevalence pre- and post- return while both enabling correlation between topics and accounting for location characteristics and time shocks which occur across all messages. Similarly, for the seeded displacement and seeded return models we run mixed effects models to identify the association between population displacement or return and percentage of topic discussion. We model the scalar outcome percent Y_{ijk} as a function of return in the previous month, with varying intercepts based on subdistrict of location i, month j, and topic k. We model this as:

$$Y_{ijk} = X_{ijk}^{T}\beta + Z_{ijk}^{T}\delta k + \epsilon_{ijk} \qquad (1)$$

In our specification X_{ijk} is a vector containing the global intercept, control variables, and return values. The return values are denoted one of two ways, either as a dichotomous variable indicating whether there is return in location i in month j or a continuous variable indicating the percent of the 2004 population displaced or returned to location i in month j. Z_{ijk} is a vector including the global intercept and corresponding with the component of X_{ijk} which we allow to vary across topics, in this case return values. Finally, we have $\epsilon_{ijk} \sim N(0, 1)$ and $\delta_k \sim N(0, \Omega)$.[11]

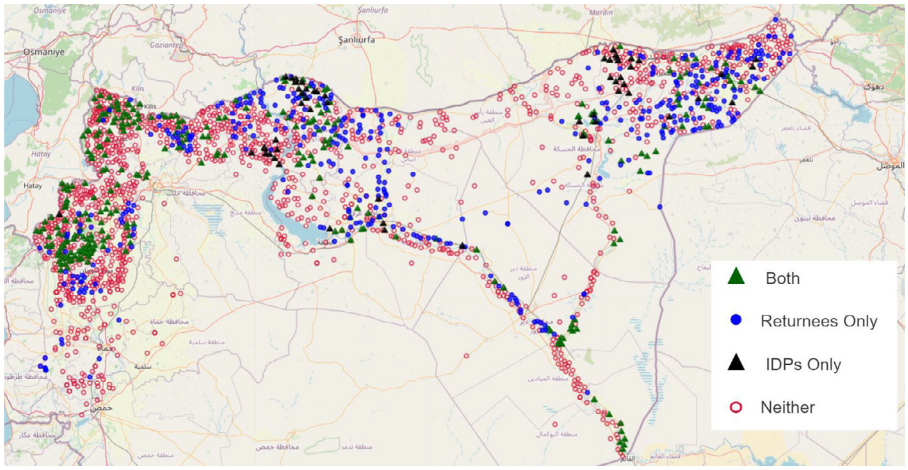
[7] https://www.facebook.com/National.Defense.in.maharda

[8] https://twitter.com/newsmanbij

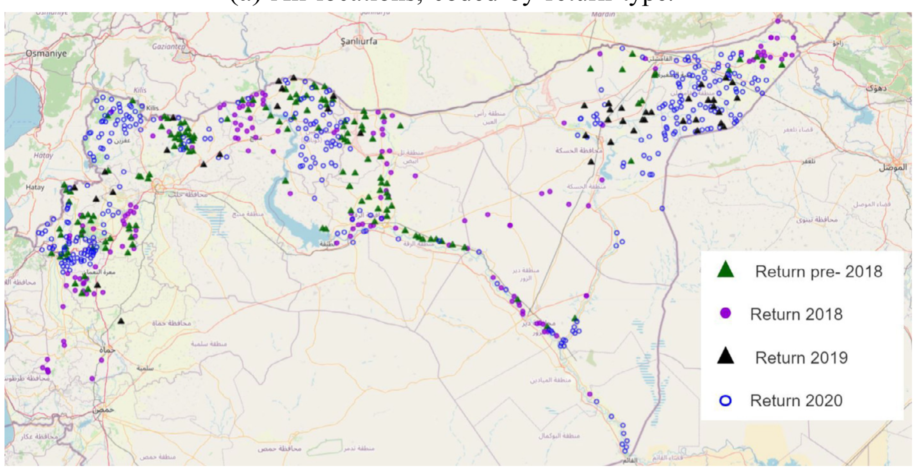
[9] https://t.me/saraqib2017n

[10] A note on CBOW word embeddings: Two words with a high cosine similarity in the word embedding space are either frequently co-located or used interchangeably. For the final models we used tokenized un-stemmed text with bigrams and kept stop words for context.

[11] We chose this model based upon the hierarchical structure of our data, which naturally had time and location groups. However, the time and location effects did not end up affecting results, either alone or combined, in any of the cases, indicating that the primary variation is between topic groups as well as the relationship between topic groups and return.



(a) All locations, coded by return type.



(b) Locations where return has occurred, coded by return year.

Figure 1. Maps of locations in northern Syria for which there is both REACH and census data. Each point corresponds to the coordinates of a community level postal code.

Table 1
The fraction of messages from return and IDP areas by source for each dataset.

	Location	Mentions				Return	Date
Platform	Total Messages	% IDP	% Returnee	% Neither	% Both	Total	%Pre
Total	2,360,559	5%	16%	62%	17%	769,745	53%
Telegram	1,264,871	5%	16%	64%	15%	391,356	57%
Facebook	676,580	4%	15%	60%	21%	241,964	46%
Twitter	419,108	8%	15%	58%	19%	136,425	56%

5. Results

In this section, we present our findings on differences in content sharing in return and non-return areas and areas with and without IDPs. In Section 5.1, we assess differences in the prevalence of text topics and image clusters.[12] We add a time element by running a mixed effects model on topic prevalence pre- and post-return to determine if return coincides with changes in discussion in Section 5.2. In this section we also consider quantity of displacement from and return to a region and how this is associated with discourse. The results indicate substantive differences which could help agencies and researchers monitor return.

Overall, we see that areas with return discuss infrastructure, the economy, news and politics more and differences are statistically significant at the 99% level accounting for Bonferroni correction for multiple hypothesis testing.[13] In the Location Mentions dataset in Figure 2, the ‘Infrastructure, economy’ topic makes up 2% more of the overall discussion in areas with both returnees and IDPs than it does in those with neither.[14] Discussion of the ‘News, politics’ topic area is also more prevalent, increasing from 34% to over 41%. Violence discussion, related to military action and air strikes, is less prevalent in return areas (-15%). Images also support these conclusions, with images shared in accounts from areas with only

[12] Researcher Coded analysis is in Appendix A.2.3.

[13] The 99% significance level applies for all results discussed in the remainder of this section using Bonferroni correction for multiple hypothesis testing unless otherwise noted.

[14] Percentage changes are calculated by taking the difference between the higher and lower percentages. Thus, they represent the increase in percent discussion rather than the percent increase in discussion, which we would calculate by then dividing this difference by the percentage we are comparing against.

Table 2
Location Mentions Topics in each combined topic from model with four way covariate. The words in each topic can be found in Appendix A.2.1.

Topic	# Messages	Composite Topics
Violence	889,942	'Regime military action, air strike', 'anti-ISIS campaign', 'Air strike warning', 'Regime military actions', 'War news, war reporting', 'Air strikes, damage, civilians', 'Liberation army, regime military'
Infrastructure, economy	511,418	'Idlib, roads, governance', 'Economy, weather', 'Goods, coronavirus, numbers'
News, politics	889,942	'Aleppo news', 'News reports', 'Afrin, Turkish-Kurdish politics', 'Foreign intervention', 'ME politics, Assad, Lebanon', 'Crimes, investigations'
Other	434,241	'Description, names', 'Horoscopes, description', 'Religion, texts', 'Children, civilians'

returnees including 3.3% more goods for sale such as cars, motorcycles and other items and images from areas with neither returnees nor IDPs including more violence related images such as tanks and groups of militants (+1.4%) as seen in Figure 3. Interestingly, though politics is more discussed in areas with both IDPs and returnees, images of politicians and protests are more common in areas without returnees or IDPs (+1.5%).

Adding additional granularity by considering changes pre- and post-return, we use mixed effects models and find that the only topic area which is more represented pre-return at a statistically significant level is 'Infrastructure, economy', with an increase of 3.5% as seen in Figure 4a. Discussion of violence related topics is not substantively different pre- and post-return, Figure 4, but is positively associated with increased displacement, Figure 5b. Increased return, on the other hand, is associated with infrastructure and economy topics (+17.1%), Figure 5a. In general, the most consistent signal of return from social media is on economic issues, suggesting that social media reports on aspects of the economy and service provision that are important to returnees.

5.1. Content in return and non-Return areas

We compare the prevalence of different types of content using the Location Mentions model run with a 'four way' covariate where the levels correspond to the location hosting returnees, IDPs, both, or neither. For a wholistic analysis, we consider prevalence of image clusters given the same 'four way' covariate. Images are from the Researcher Coded groups, in which no locations with only IDPs are represented. The Location Mentions STM model is dominated by violence related topics, motivating the use of the seeded return and displacement models discussed later in this section to attain additional insights. For analysis and figures we sort the topics into four main topic groups: 'Violence', 'Infrastructure, Economy', 'News, Politics', and 'Other.' The topics which make up each larger discussion area are in Table 2, and the English and Arabic words in each topic are available in Appendix A.2.1. For the image analysis, sample clusters can be seen in the top of Figure 3.

As shown in Figure 2, discussion around violence, which includes primarily topics related to military violence and air strikes as seen in Table 2, is more represented in areas with neither returnees nor IDPs. Violence topics make up close to 37% of discussion in these areas, and less than 22% of discussion in areas with both returnees and IDPs. This result supports research indicating that individuals are unlikely to want to return to areas with ongoing violence and military campaigns, especially if they were exposed to violence prior to leaving (Ghosn et al., 2021). Image sharing aligns with the message outcomes, as can be seen in Figure 3, with images related to militants and tanks more represented in areas without returnees or IDPs. Just over 3% of the total images in areas with both returnees and IDPs correspond with this topic, while it rises to 4.5% in areas with neither. Of note, the vast majority of images are news logos which were eliminated for this analysis.

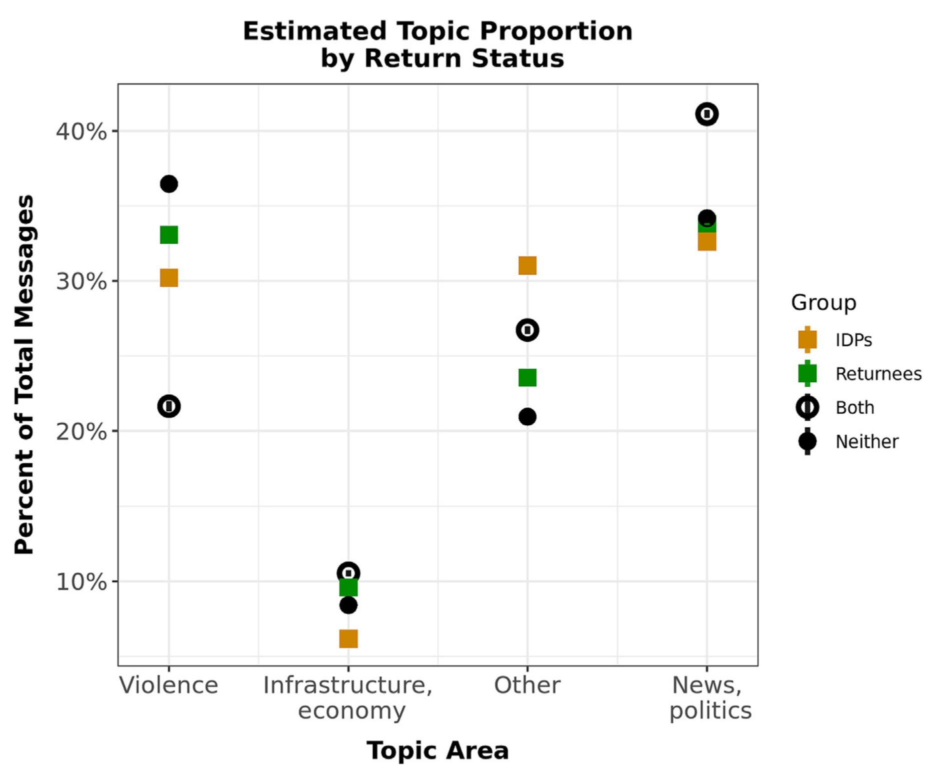


Figure 2. Variation in topic proportion by return status from linear regression run on Location Mentions STM topic outcomes. Topic proportion represented as percentage of total discussion.

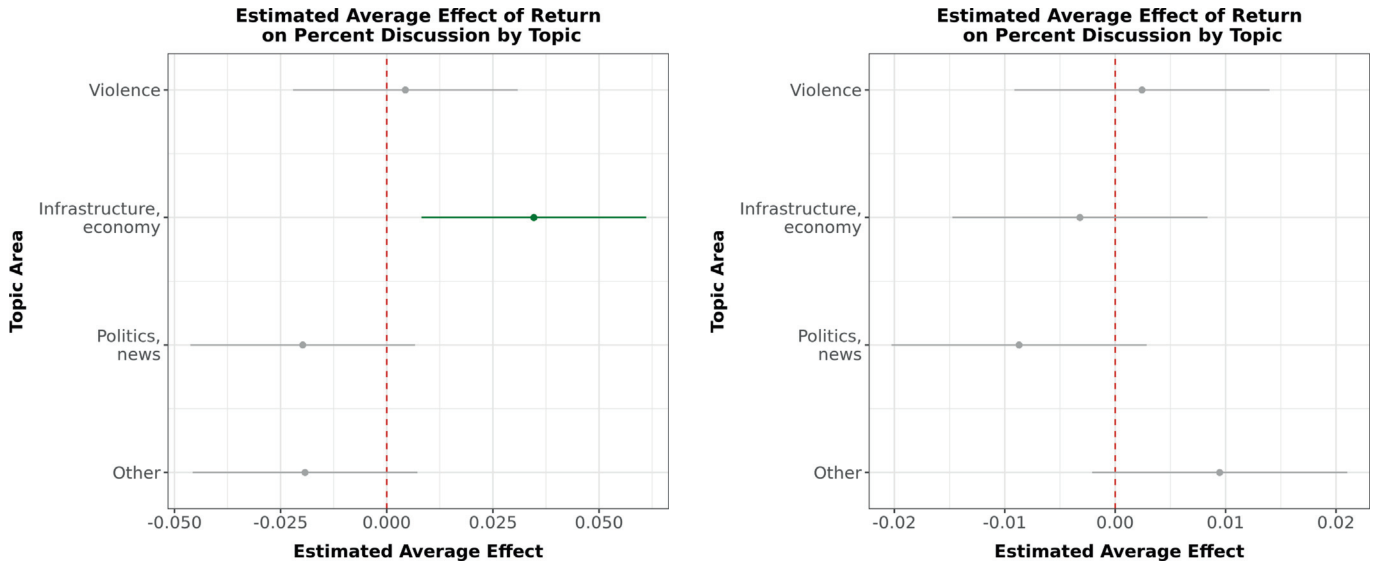
(a) Militants, Violence topic (b) Gatherings & Protests, News & politics topic

Figure 3. Sample image topic clusters, top, and prevalence of images by topic in return and non-return areas, bottom. Topic proportion measures are computed by dividing the number of messages in each image topic and group by the total number of images in that group. The prevalence comparison does not include error bars because all images were used.

Topics in the Location Mentions model which are more prevalent in areas with both returnees and IDPs (Figure 2) fall into the 'News, politics', and 'Infrastructure, economy' topic areas. In areas with returnees and IDPs, news and politics makes up 41% of discussion and infrastructure and the economy 10.5%. Conversely, in areas with neither these are 34% and 8.4% respectively. Discussion of topics in the 'Other' category, which are primarily related to religion and description, is also more prevalent in areas with return, suggesting that results on the impact of religion on return from Bove et al. (2021) are also upheld in Syria. For infrastructure and economy related topics, unlike the violence topics, areas with only returnees are more similar to areas with both returnees and IDPs than areas with only IDPs are. The prevalence of economic discussion in return areas supports research from Camarena and Hagerdal (2020) on the salience of economic considerations in driving longer term return. In addition to text discussion, over 5% of images in areas with only returnees are images of items, close to three times as many of these images as in areas with neither returnees nor IDPs as seen in Figure 3. Items images appear to be of goods for sale, cars, and motorcycles. Singh et al. (2020) found these same factors to be salient in Iraq, where the key drivers of displacement were politics, insecurities and infrastructure. The higher prevalence of content related to local governance and goods provision in return and IDP areas, contrasted with the higher prevalence of violent content in non-return and non– IDP areas, indicates that these are important indicators for communities which individuals feel safe returning to. For robustness, we also note that these results are partially reflected in the Researcher Coded models in Appendix Figure A13.

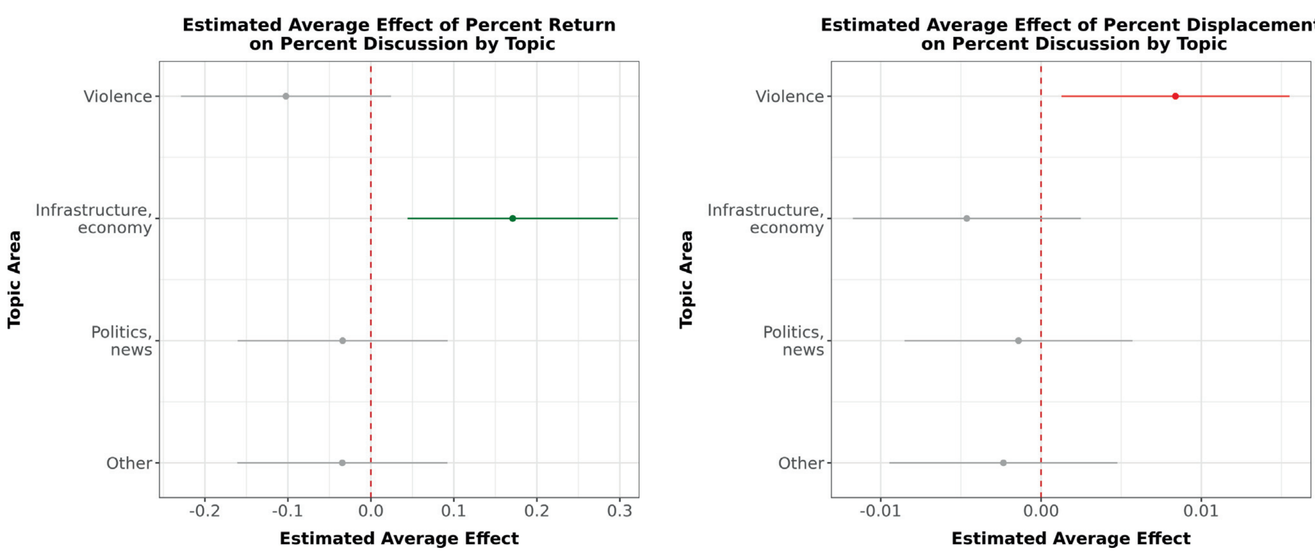
The prevalence of violence-related discussion overall in the Location Mentions models, accounting for 7 of the 20 topics and thus a high percentage of the overall discourse, outlines the persistent nature of violence in the daily lives of Syrians. Social media, especially public groups and channels, serve the population through announcements documenting daily violence and military movements which are ultimately useful for both those currently living in Syria and those choosing whether or not to return. Though images are often used along-side, or in the place of, text to convey a message, explicitly violence related topics are less represented in the image analysis than in the text. This may be the result of filtering choices by the different platforms to remove violent or disturbing imagery.[15] For violence related topics, discussion in areas with only IDPs is more similar to discussion in areas with both returnees and IDPs, whereas discussion in areas with just returnees is more similar to discussion in areas with neither returnees nor IDPs. Individuals fleeing violence often have little choice in their destination (World Bank, 2019), but

[15] Twitter transparency site: https://transparency.twitter.com/ and Facebook Community Standards: https://transparency.fb.com/policies/community-standards/



(a) Return Date (b) Seeded Return

Figure 4. Mixed effects outcomes for topic changes pre- and post-return as indicated by whether a message was posted before or after the month in which REACH indicates return began. Positive effects indicate a post-return increase.



(a) Seeded Return, Return Flow (b) Seeded Displacement, Displacement

Figure 5. Mixed effects outcomes for topic changes based on percentage of population return (left) or displacement (right) from UN OCHA data. Positive effects indicate an increase in discussion with an increase in return or displacement respectively.

if fleeing violence are unlikely to remain in areas where it is prevalent. Returnees, on the other hand, are returning in part because of ties to their home town (Ghosn et al., 2021) and thus may be willing to endure more uncertainty to be there. In fact, a UNHCR monitoring survey of voluntary returnees found that 54% returned to Syria to reunite with their families (Hoffman & Makovsky, 2021). The central role that violence plays in discussion also underscores the importance of using multiple platforms.

Different social media platforms service disparate populations, and thus have a higher prevalence of certain types of content. As shown in Appendix A.2.2, Telegram has more of the items, motorcycles, and cars images whereas tanks and construction vehicles and militants are more represented on Facebook. Twitter includes more people, gatherings and protests. In terms of text, violence related content is over represented on Telegram, while 'Infrastructure, economy' is over represented on Facebook. Heterogeneity by source may be linked to uses of the platforms by different types of groups, as explored in Walk et al. (2021). Twitter is used frequently by journalists and activists hoping to reach a foreign audience, thus images of protests may be used in an attempt to incite action beyond Syria. Facebook and Telegram are used more for day to day activity and discussion, and Telegram has a large number of groups for buying and selling goods. By using a variety of data sources we are getting a richer image of the situation on the ground.

5.2. Time and population Granularity: Mixed effects

In addition to the above models, we consider a model run exclusively on locations in which return occurred within the time frame under consideration, the Return Date model. For this model we examine the impact of return on discussion proportion including random intercepts for topic, location, and time, and random slopes which allow the impact of return to vary by topic. This model enables correlation between topics, which is likely given connections between topics, and thus removes multiple hypothesis testing issues. We run the same mixed effects model on the return seeded messages. Outcomes in these cases indicate the change in topic area discussion in areas with return after the return occurred. Finally, we run mixed effects models on the seeded return and displacement topics considering the impact of return percentage and displacement percentage respectively. These percentages are found by taking the OCHA population flow data and dividing by the population of each community per the 2004 census. These outcomes indicate the change in topic area discussion per percent increase in return or displacement.

First we consider the pre- and post-return models, Figure 4. Similar to the results in the previous section, the topics more represented post-return are infrastructure and economy related topics, as can be seen in Figure 4a, with an increase of 3.5% post return. Change in discussion of violence is not significant, and neither is politics, which differs from the previous models. In the seeded return model none of the pre- and post-return effects are significant, as can be seen in Figure 4b.

Associating the seeded return model with information from UN OCHA on return flows, we find that an increase in return again corresponds with increased discussion of 'Infrastructure, economy' topics, just over 17%. Similar to the above mixed effects models, none of the topic conglomerates is significantly more represented when return is lower. Though an increase of 17% per 1% increase in return percentage is large, it is worth noting that return flows in this dataset tend to be very small, only up to a few hundred individuals even in large cities, and thus a very small population percentage. The disparity between this model and the above model based upon a binary return date, specifically the emergence of a significant effect, shows the importance of obtaining more granular information about population flows.

In our analysis of the seeded displacement model associated with UN OCHA data on displacement flows we find that increased displacement is positively associated with violence related topics, with an increase of 0.8% per 1% increase in displacement percentage. Despite not being a clear pre- and post-return signal, online discussion of violence remains a reliable indicator of displacement as noted by Singh et al. (2019). Though the pre- and post-return mixed effects models, population flow mixed effects models, and Location Mentions model in the previous section all lead to similar conclusions, the slight differences between effects indicate the importance of considering a variety of datasets on return and the difficulties in measuring return in a systematic manner, especially for prediction of return.

6. Discussion and conclusion

This paper has showcased the types of information available in social media data and used it to understand how online conversations reflect the migration decisions of refugees and IDPs in Syria. Our conclusions support work emphasizing the importance of economic opportunities (Camarena and Hagerdal, 2020) and access to information (Alrababa'h et al., 2020) in the return decision making process, as well as the detrimental effects of ongoing violence (Davenport et al., 2003; Zolberg et al., 1989). Namely, we find that discussion of violence, especially regime related violence and coordinated campaigns, is associated with displacement and locations with neither returnees or IDPs, as are violence related images. Conversely, discussions of infrastructure, the economy, and groups for buying and selling goods, as noted through images of goods for sale, are associated with areas which have both returnees and IDPs, as well as higher levels of return. Given the ubiquity of social media, and the difficulty of surveying refugee populations, these findings are important in revealing how displacement and return choices within Syria may be documented within non-traditional data sources, aiding humanitarian groups in identifying areas with returnees or areas where future returnees may choose to settle. Specifically, groups interested in locating returnee communities or assessing return to a region would want to get a pulse on discussions of the economy and infrastructure within communities over time and compare across months and regions.

Methodologically, we paired novel data from Twitter, Telegram, and Facebook with data collected using news sources and surveys of on the ground contacts. Within the social media data we leveraged information from both text and images to obtain a more holistic picture of the discourse. Many social media sources rely heavily on images, yet the work done combining both text and image data remains sparse. By using a wide array of data sources we are able to show that social media data can be a source of information for refugees making return decisions, and many factors of interest are reflected in online discussions. Furthermore, we showcase the granularity at which social media data can provide insights, down to specific postal codes and subdistricts.

Throughout the paper we focus on presenting the promise of social media data, exploring trends and discussion at a macro level which is still valuable for policymakers (Munger et al., 2021). In future studies, research may focus on expanding these insights to a predictive setting. The presence of indicators for refugee return across various types of analysis give us hope that such trends can be used to predict whether future return will occur in a location based on changes in social media discussion. Additionally, though we focus on the Syrian case, specialists in other areas may consider how these methods would be best adapted to their setting.

Given the longitudinal nature of the data, we have the opportunity to clearly identify and quantify changes in tone of the conversation, particularly around the time of return in various locations. Many sentiment analysis tools for Arabic rely on messages from platforms such as Yelp and thus are unlikely to capture the nuances of opinion in a conflict zone. For this project we ran exploratory analysis with polyglot, which categorizes each word as positive or negative, but averaging the results of such word counts can lead to in-cohesive results since the sentiment of a sentence may not be captured by the positive or negative nature of its individual words as seen in Appendix Figure A15 (Chen & Skiena, 2014). Future research may train such a tool on conflict related social media messages using expert annotators to verify tone. This could be used to study potential conflicts or differences in opinions between returnees and the population already living in an area in order to understand the assimilation of returnees or IDPs in a location. Similarly, researchers may wish to explore sentiment towards refugees in neighboring countries such as Turkey and Lebanon and how social media discussion responds to increased migration. Such work could be augmented with satellite data and other mobility related data from google maps to identify differences in locations where returnees settle and where settlements already exist.

Finally, our work showcases the importance of access to information for refugees seeking to make decisions about return. Social media is one such source of information, and others such as Alrababa'h et al. (2020) have explored the role of access to information from trusted contacts. Future researchers may further explore the prevalence of online information about certain cities and its connection to return both within and outside of the Syrian context, or seek to understand what online sources potential returnees find reliable.

Declaration of Competing Interest

The authors declare that they have no known competing financial interests or personal relationships that could have appeared to influence the work reported in this paper.

Acknowledgements:

This paper was commissioned by the World Bank Social Sustainability and Inclusion Global Practice as part of the activity "Preventing Social Conflict and Promoting Social Cohesion in Forced Displacement Contexts." The activity is task managed by Audrey Sacks and Susan Wong with assistance from Stephen Winkler. This work is part of the program "Building the Evidence on Protracted Forced Displacement: A Multi-Stakeholder Partnership". The program is funded by UK aid from the United Kingdom's Foreign, Commonwealth and Development Office (FCDO), it is managed by the World Bank Group (WBG) and was established in partnership with the United Nations High Commissioner for Refugees (UN-HCR). The scope of the program is to expand the global knowledge on forced displacement by funding quality research and disseminating results for the use of practitioners and policy makers. This work does not necessarily reflect the views of FCDO, the WBG or UNHCR.

For helpful comments, the authors thank participants at the 2021 TADA conference, the 2022 MPSA conference, the 2022 APSA conference, and the MIT Global Diversity Lab. The authors are also thankful for constructive feedback from World Bank reviewers and discussants David Mimno, Daniel Masterson, and Kara Ross Camarena.

Walk is funded by the NSF Graduate Research Fellowship grant ID 2021311265. Garimella completed part of this work while at MIT and was supported by a Michael Hammer postdoctoral fellowship.

Appendix

A.1. Information about locations represented in the data

A.1.1. Location mention distributions by return status

A.1.2. Census statistics on return location by governorate

A.2. Figures and tables on model outputs, including by source

A.2.1. Top Words in Location Mentions and Return Date Topic Models, English and Arabic.

A.2.2. Results on topics by source

A.2.3. Topic Prevalence: Researcher Coded Model

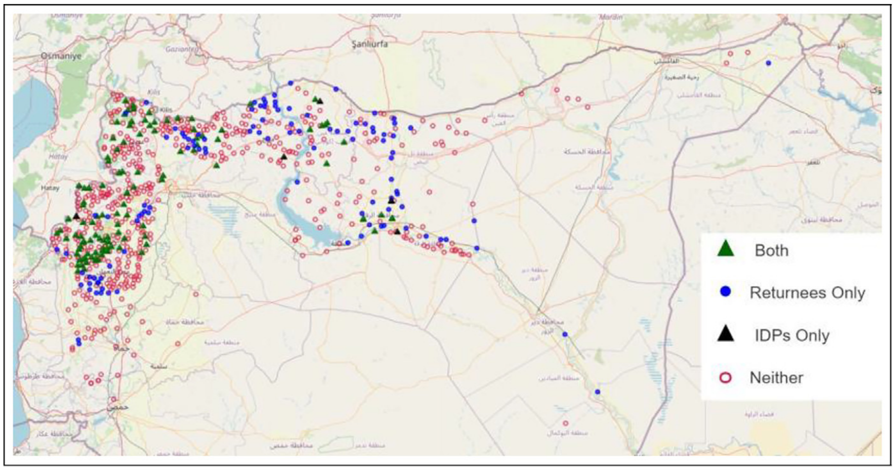
A.3. User sampling processes for Twitter, Telegram and Facebook

A.3.1. Account sampling process

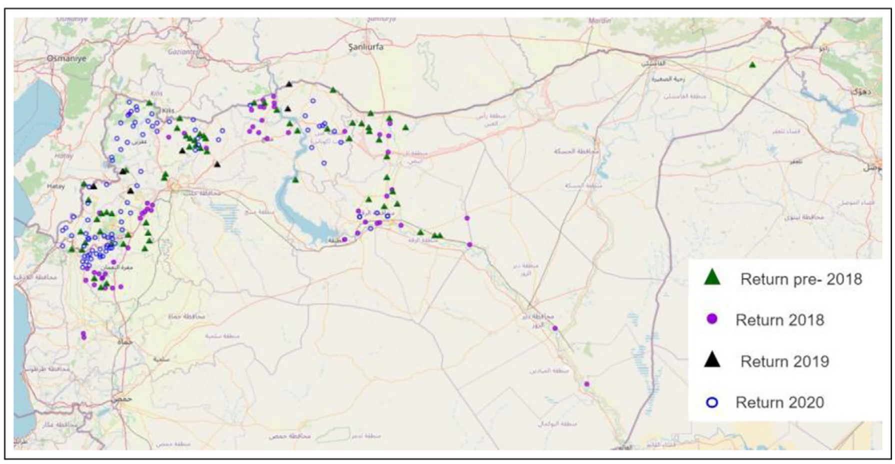
Twitter data. We used a multi-step process to collect data from Twitter accounts focused on Syria. First, we compiled a list of 304 well-known accounts focused on Syria, using a set of keywords as a guide. The complete list of words we used is provided in Appendix A.3.2. From there, we searched through these accounts to see the Twitter "lists" they belonged to (Barbera, 2015), and from there took twelve large and comprehensive lists of users focused on Syria, often curated by journalists and other Syria watchers. We collected information on the 5,000 most followed accounts from the sub-sample of users generated through this process, and again performed the same filtration. Finally, we manually combed through the user list to ensure that foreign media or political accounts were excluded from the analysis, and that all included accounts explicitly focused on Syria. This process produced 4,061 accounts of which 2,106 were actively posting after October 17th, 2017, were public facing, and had posted in Arabic.

Telegram data. On Telegram, we collected data only from public channels (one-to-many conversation) and groups (many-to-many conversation). In this paper we do not make a distinction between groups and channels. The majority of the groups included were groups acting as marketplaces for buying and selling goods. Unlike Twitter, there are no publicly aggregated lists of Telegram users, and there is also no existing strategy in literature to collect Telegram data at scale. We devised a two step strategy that similarly built on an initial, manual-compiled list followed by network connections. First, we searched the same 118 keywords for publicly available channels. These are crowd sourced collections of public channels like: https://lyzem.com, https://tgstat.com/, and https://tele.me and are generally not exhaustive. Next, we obtained all posts made in these 269 groups and channels using the Telethon APwefor Telegram at https://docs.telethon.dev/en/latest/. Within these posts, we collected all mentions or links to additional groups or channels which gave us 1,530 accounts. Finally, we manually filtered all these accounts to include only those identified as purportedly Syrian-run, focused on Syria, and which post within our timeframe about northern Syria in Arabic. This produced a dataset of 657 groups and channels. For each post, we gathered the date of posting along with the number of views the post received.

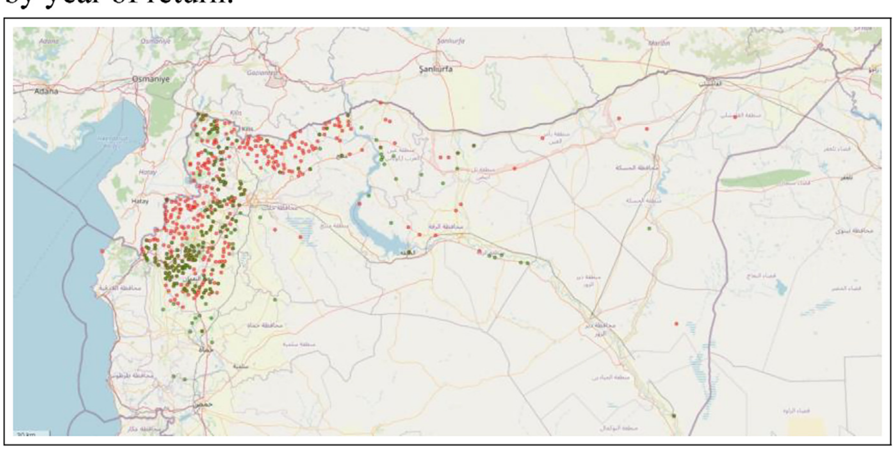
Facebook data. Finally, to collect data from Facebook, we used the CrowdTangle API at https://github.com/CrowdTangle/API/wiki CrowdTangle is officially a part of Facebook and indexes a large fraction of popular, publicly available groups and pages on Facebook. Using the API, we first searched through the CrowdTangle database for posts containing the same set of keywords as for Telegram. Then, we sorted the resulting data according to frequency of keyword mentions by various accounts. A Syrian researcher examined the resulting top 4,000 accounts and determined that 2,124 of them were Syria-focused, Syrian run, and concentrated either on any of the seven governorates in our study or on Syria in its entirety. For these 2,124 accounts, we used the APweto get all



(a) Locations mentioned in the Location Mentions dataset by IDP and returnee status.



(b) Return locations mentioned in Location Mentions dataset by year of return.



(c) Locations mentioned in Location Mentions and in the OCHA dataset, green is areas with return and red is those with displacement.

Figure A6. Maps of locations in Location Mentions dataset.

Table A3
Return type for REACH data also in the 2004 census, aggregated by governorate.

four way	governorate	# locations	average population	ethnicity prim	agriculture	distance border	distance city
Both	Al-Hasakeh	46	2118	sunnitr	0.778043	23.113008	15.825619
Both	Aleppo	117	1412	kurdish	0.502222	15.810422	11.310419
Both	Ar-Raqqa	13	3054	sunnitr	1.296154	79.466114	16.943789
Both	Deir-ez-Zor	24	6759	sunnitr	0.967917	57.303180	15.555733
Both	Hama	3	3057	sunnifam	0.273333	25.016009	2.148169
Both	Idlib	88	3025	sunnifam	1.406364	21.087119	6.768372
IDP Only	Al-Hasakeh	26	298	kurdish	0.528462	23.389464	18.939988
IDP Only	Aleppo	25	694	sunnitr	0.116400	20.622087	11.302980
IDP Only	Ar-Raqqa	5	45552	sunnitr	8.126000	84.653970	19.458138
IDP Only	Idlib	1	752	alawi	2.610000	0.560642	2.401188
Neither	Al-Hasakeh	333	1588	kurdish	0.413273	17.430056	16.013479
Neither	Aleppo	500	1388	sunnitr	0.483860	18.767451	10.409180
Neither	Ar-Raqqa	174	1672	sunnitr	0.827701	61.080247	21.508317
Neither	Dar'a	95	5710	sunnifam	1.621474	27.380196	8.402183
Neither	Deir-ez-Zor	95	8441	sunnitr	1.996316	70.134879	15.369757
Neither	Hama	91	4050	sunnifam	1.766593	48.037720	8.946009
Neither	Homs	19	10122	sunnifam	1.799474	30.391614	4.971255
Neither	Idlib	290	2853	sunnifam	0.997207	28.069215	8.170862
Neither	Rural Damascus	28	23745	sunnifam	3.254643	27.706400	2.912928
Return Only	Al-Hasakeh	145	1207	sunnitr	0.371655	19.016981	16.457821
Return Only	Aleppo	106	2119	sunnitr	0.770849	17.654791	14.355594
Return Only	Ar-Raqqa	74	2142	sunnitr	1.042568	55.250437	18.310287
Return Only	Dar'a	5	15159	sunnifam	2.752000	16.774116	5.463571
Return Only	Deir-ez-Zor	23	4904	sunnitr	2.830000	89.920148	22.121496
Return Only	Hama	12	3287	alawi	1.505000	40.083442	7.163808
Return Only	Idlib	31	5852	sunnifam	1.250000	36.161500	7.304762

Table A4
Return year for REACH data also in the 2004 census, aggregated by governorate.

year	governorate	# locations	average population	ethnicity prim	agriculture	distance border	distance city
2018	Al-Hasakeh	28	4008	kurdish	0.739286	21.322755	14.893916
2018	Aleppo	42	4140	sunnitr	1.499286	18.556167	12.029480
2018	Ar-Raqqa	29	2566	sunnitr	1.743448	58.448747	19.715763
2018	Dar'a	5	15159	sunnifam	2.752000	16.774116	5.463571
2018	Deir-ez-Zor	20	4883	sunnitr	2.724000	90.729891	22.330015
2018	Hama	11	3544	alawi	1.641818	36.359788	6.791851
2018	Idlib	22	6643	sunnifam	1.590909	37.477727	8.070008
2019	Al-Hasakeh	13	824	kurdish	0.368462	17.564086	12.040481
2019	Aleppo	58	1223	sunnifam	0.233793	15.215984	12.379789
2019	Ar-Raqqa	44	1894	sunnitr	0.598636	54.026432	17.361503
2019	Deir-ez-Zor	3	5042	sunnitr	3.536667	84.521861	20.731374
2019	Idlib	31	2130	sunnifam	1.156452	24.306679	7.917267
2020	Al-Hasakeh	122	1097	sunnitr	0.424016	18.179761	16.403270
2020	Aleppo	112	1079	kurdish	0.478125	17.187616	13.235720
2020	Ar-Raqqa	13	3054	sunnitr	1.296154	79.466114	16.943789
2020	Deir-ez-Zor	24	6759	sunnitr	0.967917	57.303180	15.555733
2020	Hama	3	3057	sunnifam	0.273333	25.016009	2.148169
2020	Idlib	62	3709	sunnifam	1.478387	21.469372	6.019038
pre-2018	Al-Hasakeh	28	558	sunnitr	0.445000	27.762843	19.271699
pre-2018	Aleppo	11	2199	kurdish	0.944545	12.211644	12.667772
pre-2018	Ar-Raqqa	1	790	sunnitr	0.250000	16.355649	19.297956
pre-2018	Hama	1	463	sunnifam	0.000000	81.043640	11.255345
pre-2018	Idlib	4	1373	sunnifam	0.000000	16.888717	6.477131

the posts made during the period Oct 2017 to Dec 2020. For each message we collected account information, post time, interactions, and engagement.

For all three platforms, we first filtered accounts in these lists according to whether their accounts' self-reported locations or descriptions mentioned Syria or any Syrian governorates in English or Arabic, and then performed a manual check to exclude foreign accounts. This process prevented the inclusion of foreign media or political accounts not exclusively focused on Syria. Our dataset thus includes self-described Syrian users posting in Arabic from inside and outside of Syria. However, there is no reason to believe that accounts on any of the platforms – from identified Syrian users posting about Syria – are more or less likely to be inside or outside of the country. Qualitatively, many opposition news sites feature the works of internal correspondents working with editors outside of Syria.

A.3.2. Keywords for account collection

Keywords for data collection across platforms are as follows. For each key term without an obvious Syrian signifier, a Syrian researcher appended "Syria" or one of the seven governorates in our study. Keywords focus on key locations, conflict trends, and local governance and economic conditions. All search terms translated from Arabic:

- "Syrian War, Syrian Civil War, Bashar al-Assad, Idlib, Aleppo, Raqqa, Qamishli, Deir Al Zour, Hasaka, Hama, Arab Spring, Azaz, Olive Branch, Euphrates Shield, Syrian Civil War, Syrian news,

1.	Regime military action, Air strike	237,592 messages	F: rural, Idlib, side, cannon, rural_Idlib, Rikh, shelling FREX: bombardment, countryside_Idlib, Idlib, artillery shelling, shot_war, hit_cannon, shot_rikh
2.	Description, names	41,649 messages	F: Muhammad, Hamad, Sheikh, Hassan, Ahmed, good, girl FREX: I know, Master, I saw, Thank God, I said, praised_praised, Hadal
3.	Idlib, roads, governance	74,131 messages	F: City, camp, cities, road, Idlib, Idlib, bridge FREX: Boil, Kfar Takh, Kfar Takh_Rima, Mahal_City, Idlib_City, City_Sarmed, place_religion
4.	anti-ISIS campaign	230,185 messages	F: Monastery, Sri, East, Dimkar, SDF, Sri_Dimkar, Raqqa FREX: Brif_Deir, Ain_Issa, rural_Raqqa, rural_Hasakah, Lazar_East, Raqqa_Sri, Demqar_arrested
5.	Air strike warning	120,025 messages	F: Accurate, reach, possible, accurate, possible_reach, flying, war FREX: Accurate, accurate, possible_reach, flying, accurate_accurate, war_flying, sanctity
6.	Aleppo news	24,223 messages	F: Aleppo, countryside, west, rural Aleppo, leave, countryside of Aleppo, points FREX: Rural_Aleppo, Aleppo_west, city_of_Aleppo, Aleppo_side, friend_friend, Aleppo_road, Tarnab
7.	Liberation army, Regime military	40,165 messages	F: Army, Mujahid, Liberation, Erh, Idlib, walking, armed FREX: Mahar_Brief, Telegram, shutter_telegram, telegram_bridge, army_controlling, walk_er, er_complex
8.	News reports	145,970 messages	F: Confidential, sketch, corridor, class, confidential, across report, NA FREX: attachment_report, inherit, inherit_press, rent, shutter_telegram, ether, sight_attachment
9.	War news, war reporting	61,719 messages	F: War, east, military, Damascus, army, Erh, side FREX: Eastern, war_centered, eastern_Aust, eastern_Marj, east_bombing, east_rear, conscience_askar
10.	Economy, weather	269,595 messages	F: Stair, price, stoop, lower, east, side, west FREX: rate, precipitation, light, gasoline, partial, rise, wet
11.	Afrin, Turkish-Kurdish politics	87,141 messages	F: Afrin, occupy, hero, mercenary, debtor, witness, Reikh FREX: mercenaries, hero_honor, occupied_mercenaries, Rech_martyrdom, Team_hamza, mercenary_occupation, martyrdom_hero
12.	Foreign intervention	128,413 messages	F: undercover, America, anchor, confidential, let, united, saddle FREX: please, Washington, Jin_constitution, ok_please, Tayyib_Ardagh, Rais_Rajab, United_Rick
13.	ME politics, Assad, Lebanon	87,026 messages	F: people, chief, company, Arabs, Lebanon, party, council FREX: Mr. President, People's Assembly, Palestinian, Ba'ath_Arab, Netni, of the Ba'ath Party, Arab_Shter
14.	Horoscopes, description	93,816 messages	F: Pregnancy, new, good, pressure, partner, horoscope, cancer FREX: cancer, pregnancy_neese, your time, neez_pressure, pressure_carry, can be, NA
15.	Air strikes, damage, civilians	125,279 messages	F: Bird, war, Idlib, bombing, bird_war, explosive, fun FREX: explosive, explosive, merry_plane, barrel, wounded man, Idlib summary, barrel bomb
16.	Goods, coronavirus, numbers	167,692 messages	F: Corn, numeral, manager, company, servant, new, reward FREX: press_cren, new_log, send, new_press, send_cren, kern_rising, sterilization
17.	Regime military actions	74,977 messages	F: Assad, countryside, window, ower, cannon, do shield, NA FREX: to fall, Fee_Shield, Pearl_Tiger, Fall_Of, Bomb_Mine, Exhaust_Trade, Land_Mine
18.	Religion, texts	219,306 messages	F: same, much, all, debt, less, earth, wealth FREX: I want, make, you turn, you were, dict, let down, location
19.	Children, civilians	79,470 messages	F: child, civilian, bombardment, circle, hospital, witness, city FREX: Ratif_Haseel, child_Mohammed, patient_hospital, city_Arab, transfer_hospital, Kafr Batn, hospital_hospital
20.	Crimes, investigations	45,349 messages	F: condition, kill, person, carry, poison, investigation, prison FREX: betrayed, crime_murder, confess, grimm, thousand_breaths, disappear, rape

Figure A7. Location Mentions Topics with four way covariate. ''F'' indicates words that are most frequent in each topic. ''FREX'' indicates words that are both frequent in and exclusive to each topic. Message numbers are the total number of messages from all sources combined for that topic. All results are translated from Arabic.

1.	**Regime military action, Air strike**	**237,592 messages**	F: بريف, إدلب, جنب, مدفع, بريف_إدلب, ريخ, قصف FREX: تقصف, بريف_دلب, دلب_جنب, تقصف_مدفع, حرب_يسدف, يسدف_ريخ, مدفع_يسدف
2.	**Description, names**	**41,649 messages**	F: محمد, حمد, شيخ, حسن, أحمد, خير, بنت FREX: بعرف, سيدن, شفت, حمدلل, قلتل, حمد_حمد, هدل
3.	**Idlib, roads, governance**	**74,131 messages**	F: مدين, مخيم, مدن, طريق, إدلب, دلب, جسر FREX: سلق, كفرتخ, كفرتخ_ريم, محل_مدين, دلب_مدين, مدين_سرمد, محل_دين
4.	**anti-ISIS campaign**	**230,185 messages**	F: دير, سري, شرق, ديمقر, قسد, سري_ديمقر, لرقة FREX: بريف_دير, عين_عيسى, بريف_لرقة, بريف_لحسكة, لزر_شرق, لرقة_سري, ديمقر_تعتقل
5.	**Air strike warning**	**120,025 messages**	F: دقيق, تصل, ممكن, دقيقت, ممكن_تصل, تحلق, حرب FREX: دقيق, دقيقت, ممكن_تصل, تحلق, دقيق_دقيق, حرب_تحلق, حرمة
6.	**Aleppo news**	**24,223 messages**	F: حلب, ريف, غرب, بريف_حلب, ترك, ريف_حلب, نقط FREX: ريف_حلب, حلب_غرب, مدين_حلب, حلب_جنب, صديق_صديق, طريق_حلب, ترنب
7.	**Liberation army, Regime military**	**40,165 messages**	F: جيش, محر, تحرير, إره, إدلب, مشي, مسلح FREX: محر_بريف, بتلغر, شتر_بتلغر, بتلغر_جسر, جيش_يسيطر, مشي_إره, مجمع_إره
8.	**News reports**	**145,970 messages**	F: سري, رسم, صيل, أخب, سرية, عبر, تقرير FREX: تقرير_مرفق, نرث, نرث_برس, رينت, شتر_تلغر, أثير, مرفق_بصر
9.	**War news, war reporting**	**61,719 messages**	F: حرب, شرق, عسكر, دمشق, جيش, إره, جنب FREX: شرقية, حرب_مركز, شرقية_أسط, شرقية_مرج, قصف_شرق, شرق_مؤخر, ضمير_عسكر
10.	**Economy, weather**	**269,595 messages**	F: درج, سعر, رتف, خفض, شرق, جنب, غرب FREX: سعر, هطل, نخف, بنز, جزئي, يرتفع, رطب
11.	**Afrin, Turkish-Kurdish politics**	**87,141 messages**	F: عفرين, حتل, بطل, مرتزق, مدين, شهد, ريخ FREX: مرتزقت, بطل_شرف, حتل_مرتزقت, ريخ_ستشه, فرق_حمز, مرتزق_حتل, ستشه_بطل
12.	**Foreign intervention**	**128,413 messages**	F: سري, ريك, رسي, سرية, ترك, متحد, أردغ FREX: أردغ, شنطن, لجن_دستر, رجب_طيب, طيب_أردغ, رئيس_رجب, متحد_ريك
13.	**ME politics, Assad, Lebanon**	**87,026 messages**	F: شعب, رئيس, سرية, عرب, لبن, حزب, مجلس FREX: سيد_رئيس, مجلس_شعب, فلسطيني, بعث_عرب, نتني, لحزب_بعث, عرب_شتر
14.	**Horoscopes, description**	**93,816 messages**	F: حمل, جديد, جيد, ضغط, شريك, برج, سرط FREX: سرط, حمل_نيز, قتك, نيز_ضغط, ضغط_حمل, يمك, فكن
15.	**Air strikes, damage, civilians**	**125,279 messages**	F: طير, حرب, إدلب, قصف, طير_حرب, متفجر, مرح FREX: متفجر, ميل_متفجر, طير_مرح, ببرميل, أصيب_رجل, ملخص_إدلب, برميل_متفجر
16.	**Goods, coronavirus, numbers**	**167,692 messages**	F: كرن, عدد, مدير, شرك, خدم, جديد, إجر FREX: برس_كرن, تسجيل_جديد, يرس, جديد_برس, يرس_كرن, كرن_رتفع, تعقيم
17.	**Regime military actions**	**74,977 messages**	F: أسد, ريف, نفج, مدين, مدفع, أدى, درع FREX: لسقط, رسم_درر, درر_تغر, أدى_لسقط, نفج_لغم, نفج_مفخخ, لغم_أرض
18.	**Religion, texts**	**219,306 messages**	F: نفس, كثير, جميع, دين, يقل, أرض, ثرة FREX: أريد, يصنع, ثرتن, كنتم, ديكت, خذل, مرقع
19.	**Children, civilians**	**79,470 messages**	F: طفل, مدني, قصف, أطف, مشفى, شهد, مدين FREX: رتف_حصيل, طفل_محمد, مريض_مشفى, مدين_عرب, نقل_مشفى, كفربطن, مشفى_مشفى
20.	**Crimes, investigations**	**45,349 messages**	F: شرط, قتل, شخص, تقل, تسم, تحقيق, سجن FREX: مغدر, جريم_قتل, عترف, جريم, ألف_نسم, ختف, غتص

Figure A8. Words for the topics in the Location Mentions model with four way covariate, stemmed Arabic.

Syrian Democratic Forces, SDF, Syrian National Army, Kurdistan Workers Party, Islamic State of Iraq and the Levant, ISIS, YPG, People's Protection Units, Mujahideen, Ahrar al-Sham, Tahrir al-Sham, Free Syrian Army, Russia in Syria, Regime, Iran in Syria, smuggling, Syrian Lira, Syrian currency, (Province) markets, Azaz markets, Afrin markets, Situation of displaced persons in areas outside of government control, Haramain camp, Atmh camp, Mahmodla camp, Tall Abyad Camp, IDP camps in (Province), Syrian returnees, Local council in (Location), Cases of return to (Location), Autonomous Administration, People's municipality in (Location), Civil council in (Location), Military council in (Location), Eid in Syria, Ramadan in Syria, Ramadan in (Province), Holiday Sweets in (Province), Education in (Province), Schools in (Province) Azaz, Tabqa, Baghouz, Albukamal, Al Hasakah, Al Shadadi, Al Mayadeen, Al Houl, Semalka, Semalka Border Crossing, Bab Al-Hawa, Manbij, Jarablus, Al Dana, Bab Al Howa, Afrin, Saraqib, Ma'aret al Nauman, Kobani, Ein al Arab, Atareb, Al bab, Jarablus, Azaz, Jarablos, Aldana, Al Raee".

1.	Insurgency, anti-ISIS campaign	27,374 messages	F: east, monastery, QSD, countryside, Raqqa, walking, tunnel FREX: Free_control, Sayed_Quneitra, staircase, Hassi_Qurayt, frost_rift, send_welcome, Hasakah_paralyzed
2.	Air strike, civilian witness	50,657 messages	F: plane, war, plane_war, arch, bombing, city, merry FREX: airplane_merry, war_slashing, barrel_blowning, explosive_tendency, sdf_airplane, war_clopping, barrel
3.	Shops, economy	61,613 messages	F: company, price, electric, expand, oil, cut, race FREX: origin_monitor, benz, name_rise, rise_origin, observe_name, price, exchange_rate
4.	Names, martyrs	102,790 messages	F: Muhammad, Hamad, Hassan, Sheikh, Ahmed, hero, Mustafa FREX: Muhammad_Muhammad, Hamad_Mohammed, Hamad_Hamad, Muhammad_Hamad, Aleppo_Hamad, Aleppo_Mohammed, Muhammad_Sahb
5.	Aleppo news	65,426 messages	F: Aleppo, countryside, rural_Aleppo, countryside_Aleppo, friend, west, old man FREX: Kafr_Hamar, Aleppo_Janb, Aleppo_West, Rural_Aleppo, City_Aleppo, Aleppo_Aleppo, Aleppo_Sri
6.	Idlib, hospitals, roads	28,301 messages	F: city, Idlib, cities, hospital, road, center, team FREX: Syrian_Idlilb,hit_hit, hit_eye, city_team, hospital_transport, clothed, receive, NA
7.	Crimes, children	10,227 messages	F: carry, condition, Afrin, meteor, capture, shop, military FREX: carry_neez, neez_press, press_carry, military condition, arrest_to, draw_to, section_condition
8.	Regime military action	54,338 messages	F: Assad, countryside, cannon, walk, side, drive, east FREX: Assad_engine,led_to_fall, engine_deek, wounded_arrayed, engine_feteer, killed_injured, managed_to_destroyed, NA
9.	Air strike warning	30,049 messages	F: accurate, possible, reach, fly, possible_reach, minute, war FREX: Take_Kafr, Idlib_city, Hamra, Kafr_Hamra, Sarm_war, flying_east, flying
10.	Office, government	53,397 messages	F: chief, company, council, director, master, minister, GM FREX: Mr. President, Ba'ath_Arab, Baath_Party, Arab_Shooter, Decree_legislation, Legislation, Decree
11.	Medicine, health	5,887 messages	F: heart, you, stranger, girl, infant, breath, man FREX: I know, I saw, I said, I went, it became, you know, a muscle
12.	ME politics, Turkey	37,221 messages	F: brigade, America, Russia, brigade, leave, united, Erdogan FREX: Erdogan, President_Recep, Recep_Tayyip, Washington, Jeff, rick_sri, Recep_Erdogan
13.	Religion	22,350 messages	F: arab, religion, much, soul, hurricane, earth, say FREX: Quran, serial, less, Jesus, Holy Qur'an, poem, tenth century
14.	Camps, displacement, violence	26,890 messages	F: armor, damascus, poison, Assad, kill, fight, go out FREX: Countryside_Daraa, Karak_east, last_newer, Daraa_east, Rint, rural_Homs, Daraa_west
15.	Days, weather	23,915 messages	F: stairs, west, sea, lower, side, east, flat FREX: Push,stair_stair, extend_lower, west_tall, fall_directed, extended_affected, inclined_stair, NA
16.	Military control	14,237 messages	F: army, military, erh, liberation, armed, leave, Arab FREX: army_army, army_dominated, army_dominated, armored, ere_army, entered_army, organized_victory
17.	Politics news	23,767 messages	F: Kurd, Lebanon, Israel, captive, company, Israel, new FREX: Israel, corn_rose, new_press, Israel, party_master, stinky, capturing
18.	Air strike Idlib	42,930 messages	F: Idlib, countryside, side, shelling, countryside_Idlib, Idlib_side, artillery FREX: Telegram_Bridge, Telegram, Shutter_Telegram, Idlib_Syrian, Buy_Telegram, Sarm_East, shotgun_Confirmed
19.	Horoscope, description	47,480 messages	F: good, new, person, maybe, chances, adverb, analyze FREX: Imper_pregnancy, your condition, your surroundings, friend's_sign, good_day, your time, your situation
20.	Turkish-Kurdish politics	38,263 messages	F: people, secret, wealth, Kurdish, Assad, demagogue, occupy FREX: Kurdistan, abundance, our people, Reich_Satsheh, Thartan, Kurdish_Council, Dict

Figure A9. Words for the topics in the Return Date model with a pre- and post-return covariate, English translation.

1.	**Insurgency, anti-ISIS campaign**	**27,374 messages**	F: شرق, دير, قسد, ريف, لرقة, مشي, نفج FREX: لحر_سيطر, سيد_قنيطر, درج_بحر, حسي_قريت, صقيع_رتف, مرسل_رحب, لحسكة_مشل
2.	**Air strike, civilian witness**	**50,657 messages**	F: طير, حرب, طير_حرب, ريخ, قصف, مدين, مرح FREX: طير_مرح, حرب_يسدف, يسدف_ريخ, ميل_متفجر, سدف_طير, حرب_تسدف, بيرميل
3.	**Shops, economy**	**61,613 messages**	F: شرك, سعر, كهرب, أسع, نفط, قتص, ريس FREX: أصل_رصد, بنز, مسم_يرتفع, يرتفع_أصل, رصد_مسم, سعر, سعر_صرف
4.	**Names, martyrs**	**102,790 messages**	F: محمد, حمد, حسن, شيخ, أحمد, بطل, مصطفى FREX: محمد_محمد, حمد_محمد, حمد_حمد, محمد_حمد, حلب_حمد, حلب_محمد, محمد_صحب
5.	**Aleppo news**	**65,426 messages**	F: حلب, ريف, بريف_حلب, ريف_حلب, صديق, غرب, شيخ FREX: كفر_حمر, حلب_جنب, حلب_غرب, بريف_حلب, مدين_حلب, حلب_حلب, حلب_سري
6.	**Idlib, hospitals, roads**	**28,301 messages**	F: مدين, دلب, مدن, مشفى, طريق, مركز, فريق FREX: سري_دلب, حدث_إخب, إخب_عين, فريق_مدن, نقل_مشفى, ملبس, لتلق
7.	**Crimes, children**	**10,227 messages**	F: حمل, شرط, عفرين, نيز, قبض, محل, عسكر FREX: حمل_نيز, نيز_ضغط, ضغط_حمل, شرط_عسكر, إلق_قبض, برسم_برسم, قسم_شرط
8.	**Regime military action**	**54,338 messages**	F: أسد, ريف, مدفع, مشي, جنب, محر, شرق FREX: أسد_محر, أدى_لسقط, محر_ديخ, جرحى_صفف, محر_فطير, قتل_جرح, تمكنت_تدمير
9.	**Air strike warning**	**30,049 messages**	F: دقيق, ممكن, تصل, تحلق, ممكن_تصل, دقيقت, حرب FREX: تصل_كفر, إدلب_مدينة, حمرة, كفر_حمرة, سرم_حرب, تحلق_شرق, تحلق
10.	**Office, government**	**53,397 messages**	F: رئيس, سرية, مجلس, مدير, سيد, زير, جتم FREX: سيد_رئيس, بعث_عرب, لحزب_بعث, عرب_شتر, مرسم_تشريع, تشريع, مرسم
11.	**Medicine, health**	**5,887 messages**	F: قلب, كنت, غريب, بنت, طفل, نفس, رجل FREX: بعرف, شفت, قلتل, رحت, يصير, بتعرف, عضل
12.	**ME politics, Turkey**	**37,221 messages**	F: سري, ريك, شنطن, سرية, رسي, متحد, أردغ FREX: أردغ, رئيس_رجب, رجب_طيب, شنطن, جيفر, ريك_سري, طيب_أردغ
13.	**Religion**	**22,350 messages**	F: عرب, دين, كثير, نفس, إسل, أرض, يقل FREX: قرآن, مسلسل, أقل, مسيح, قرآن_كريم, قصيد, قرن_عشر
14.	**Camps, displacement, violence**	**26,890 messages**	F: درع, دمشق, تسم, أسد, قتل, صيل, خرج FREX: بريف_درع, كرك_شرق, أخر_مستجد, درع_شرق, رينت, بريف_حمص, درع_غرب
15.	**Days, weather**	**23,915 messages**	F: درج, غرب, بحر, خفض, جنب, شرق, متسط FREX: أرض, درج_درج, متد_خفض, غرب_تدل, تقعت_مدير, تأثر_متد, تميل_درج
16.	**Military control**	**14,237 messages**	F: جيش, عسكر, إره, تحرير, مسلح, ترك, عرب FREX: جيش_جيش, سيطر_جيش, جيش_يسيطر, مدرع, إره_جيش, دخل_جيش, تنظيم_نصر
17.	**Politics news**	**23,767 messages**	F: كرن, لبن, ئيل, إسر, سرية, إسر_ئيل, جديد FREX: إسر_ئيل, كرن_رتفع, جديد_برس, ئيل, لحزب_سيد, نتني, إسر
18.	**Air strike Idlib**	**42,930 messages**	F: إدلب, بريف, جنب, قصف, بريف_إدلب, إدلب_جنب, مدفع FREX: بتلغر_جسر, بتلغر, شتر_بتلغر, إدلب_سري, إشتر_تلغر, سرم_شرق, مسدف_تأكدت
19.	**Horoscope, description**	**47,480 messages**	F: جيد, جديد, شخص, ريم, فرص, ظرف, حلل FREX: أبر_حمل, يمك, محيطك, برج_صديق, يمك_جيد, قتك, ضعك
20.	**Turkish-Kurdish politics**	**38,263 messages**	F: شعب, سري, ثرة, كرد, أسد, ديمقر, حتل FREX: كردست, ثرة, شعبن, ريخ_ستشه, ثرتن, مجلس_كرد, ديكت

Figure A10. Words for the topics in the Return Date model with a pre- and post-return covariate, stemmed Arabic.

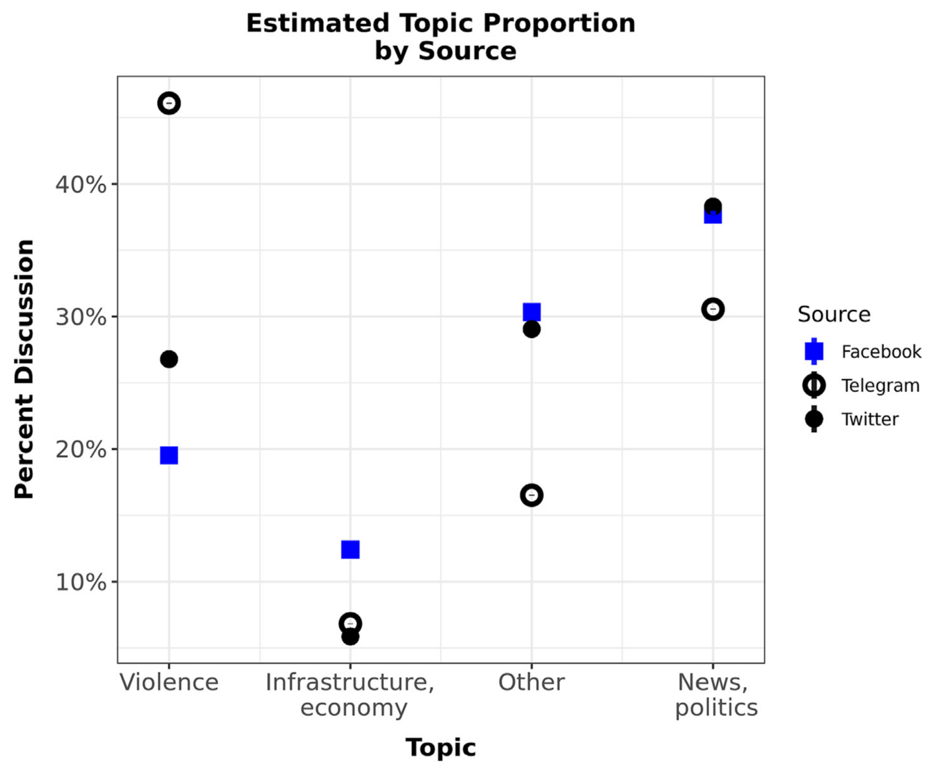


Figure A11. Variation in topic proportion by data source from linear regression run on Location Mentions STM topic outcomes. Topic proportion represented as percentage of total discussion.

Figure A12. Prevalence of images by topic for each data source. Topic proportion measures are computed by dividing the number of messages in each image topic and group by the total number of images in that group. The prevalence comparison does not include error bars because all images were used.

A.4. Seed words for seeded models

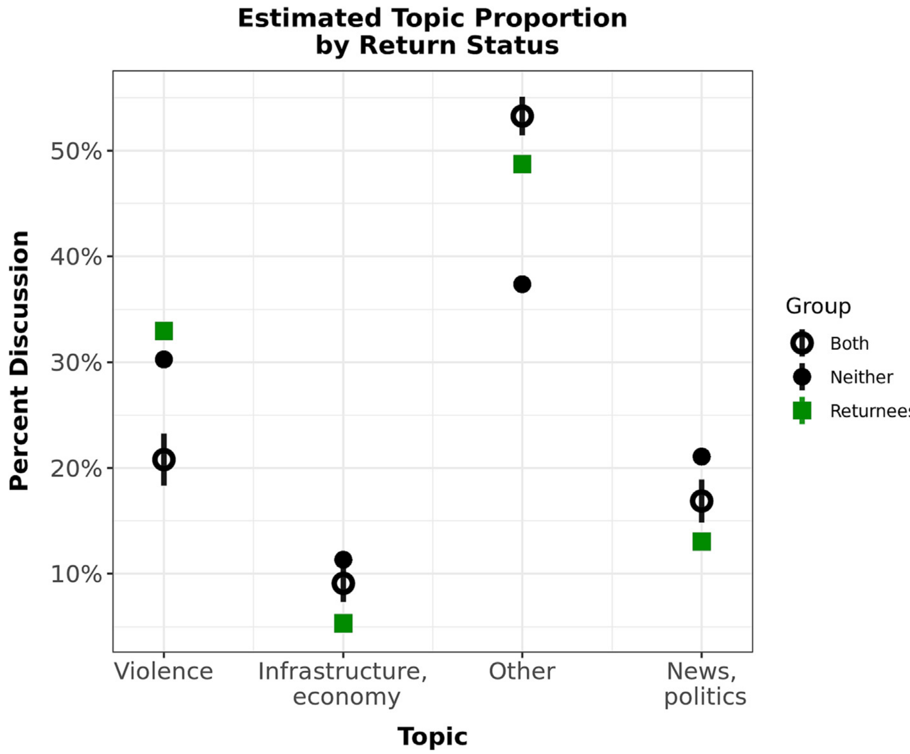


Figure A13. Variation in topic proportion by return status from linear regression run on Researcher Coded STM topic outcomes. Topic proportion represented as percentage of total discussion. Results on infrastructure are not significant.

services	الكهرباء , السعر , الأسعار , الكيلوغرام , كيلو, الصرف_الصحي , السباكة , البناء , الإعمار ,اعادة_الاعمار, الوظائف , التوظيف , المصانع, افتتاح, مستشفى, مشفيات, المواد_الغذائية, الزراع, عامل, بقايا, ترميم, الفقر, جوع, بيع_وشراء,مركز, اسواق, سوق, محل, بلدية, مجلس_مدينة, مجلس_محلي, التعليم, النظافة, ية_ازالة, ازالة, القمامة, الخدمات_الفنية , الاتصالات , مواصلات, طرق, جسور, القيود_المدنية, السجلات_المدنية, الضرائب, الرسوم
displacement	هرب, فتح_حدود, قافلة_الأمل, نزوح_مدنيين, مدنيين, النزوح, النازحين
return	العودة, العائدين, تصريح, حواجز, حاجز, باسبورت, جواز_السفر, تجنيد, العودة الى_بيوتهم, العودة_الى_وطنهم, العودة_الى_مدينتهم"

Figure A14. Seeding keywords for return/services and displacement seeded models. These lists were expanded with word2vec and then used to create message sets.

A.5. Sentiment Analysis of Messages

Message Arabic	Message English	Polarity	Sentiment
#مرصد_اخبار_الثورة_السورية عاجل| قوات سوريا الديمقراطية تعتقل أقرباء للجيش الحر قدموا لقضاء إجازة عيد الأضحى من تركيا إلى مدينة منبج. #منبج | #سوريا mirsadakhbaralthawratalsuwria@	#Syrian_Revolution_News_Observatory Urgent| SDF arrests relatives of the Free Army who came to spend the Eid al-Adha holiday from Turkey to the city of Manbij. #Manbij | #Syrian Arab Republic @mirsadakhbaralthawratalsuwria	0	Neutral
رسالة الآن من أهل كناكر إلى ثوار الشمال يذكر إن عصابات الأسد تبدأ بإقتحام بلدة كناكر بريف دمشق الجنوبي بعد ثمانية أيام على محاصرة البلدة مطالبة عصابات بتسليمها مطلوبين لها و كمية من السلاح. اللّهم كن عوناً لإخواننا و ثبتهم حسبنا اللّه و نعم الوكيل يلعن روحك يا حافظ https://t.co/6prdccutmg	A message now from the people of Kanaker to the rebels of the North It is reported that Assad's gangs begin storming the town of Kanaker in the southern countryside of Damascus, eight days after the besieging of the town, demanding that the gangs hand over wanted men and a quantity of weapons. Oh God, be of help to our brothers and make them steadfast.	0	Neutral
rt @ramashh0: جرائم عبدالحليم لا تنتهي بدءا من تعاونه مع حافظ لتسليم القنيطرة مروراً بمجازر حماة وحلب ودفنه للنفايات النووية في الصحراء السورية وصو لـ اً لتعديله للدستور ليتمكن بشار من تسلم الحكم، وكل ما فعله	rt @ramashh0: Abdel Halim's crimes do not end, starting with his cooperation with Hafez to hand over Quneitra, through the massacres of Hama and Aleppo, and his burial of nuclear waste in the Syrian desert, to his amendment of the constitution so that Bashar can take over the rule, and all that Khaddam did after his defection was to enjoy the money he stole. it now!! https://t.co/c0i5hybwvw	0	Neutral
#عاجل || نظام الأسد يعتقل بعض العائدين إلى الغوطة الشرقية و الذين تم تهجيرهم سابقاً للشمال السوري في شهر آذار الماضي .	#urgent || The Assad regime arrests some of the returnees to Eastern Ghouta, who were previously displaced to northern Syria last March.	0	Neutral
حمص : غارات جوية من طيران الأجرام الأسدي و طيران الغزاة الروس استهدفت قرى "سليم والحمرات وعزالدين وديرفول" بالريف الشمالي ... لمتابعة مزيد من الاخبار والفيديو والصور الملحقة و حتى يصلك كل جديد , اشترك بشبكة #ديري_نيوز الإخبارية على تلغرام من خلال الرابط التالي : https:// telegram.me/derynews	chickpeas : Air raids from the warplanes of the Assad regime and the planes of the Russian invaders targeted the villages of "Salim, Hamrat, Ezzedine and Derful" in the northern countryside... :// telegram.me/derynews	0.0125	Positive

Figure A15. Examples of sentiment miscoding for messages. A human annotator would code all of these as negative.

References

Abdo, G. (2015). Salafists and sectarianism: Twitter and communal conflict in the middle east. Center for Middle East Policy at Brookings. Technical Report. Retrieved from https://www.brookings.edu/research/salafists-and-sectarianism-twitter-and-communal-conflict-in-the-middle-east/.

Abrishamkar, S., Khonsari, F., An, A., Huang, J. X., & McGrath, S. (2018). Mining large-scale news articles for predicting forced migration. In *Anchorage '19: Acm sigkdd conference on knowledge discovery and data mining, august 04–08, 2019, anchorage, ak*.

Aktas, V., Tepe, Y. K., & Persson, R. S. (2021). Investigating turkish university students' attitudes towards refugees in a time of civil war in neighboring syria. *Current Psychology, 40*(2), 553–562. https://doi.org/10.1007/s12144-018-9971-y.

Al-Hilu, K. (2019). Afrin under turkish control : political, economic and social transformations. Middle East Directions. Technical Report. Retrieved from https://cadmus.eui.eu/handle/1814/63745

Alrababa'h, A., Dillon, A., Williamson, S., Hainmueller, J., Hangartner, D., & Weinstein, J. (2021). Attitudes toward migrants in a highly impacted economy: Evidence from the syrian refugee crisis in jordan. *Comparative Political Studies, 54*(1), 33–76. https://doi.org/10.1177/0010414020919.

Alrababah, A., Masterson, D., Casalis, M., Hangartner, D., & Weinstein, J. (2020). The dynamics of refugee return: Syrian refugees and their migration intentions. *SocArXiv.*. https://doi.org/10.31235/osf.io/7t2wd.

Arias, M. A., Ibanez, A. M., & Querubin, P. (2014). The desire to return during civil war: Evidence for internally displaced populations in colombia. *Peace Economics, Peace Science and Public Policy, 20*(1), 209–233. https://doi.org/10.1515/peps-2013-0054.

Aymerich, O., & Zeyneloglu, S. (2019). House damage revisited: How type of damage and perpetrating actor affect intentions and actions of idps in iraq. *International Migration, 57*(2), 65–79. https://doi.org/10.1111/imig.12497.

Balcilar, M., & Nugent, J. B. (2019). The migration of fear: An analysis of migration choices of syrian refugees Special Issue on the Economies of Middle East and North Africa in an Era of Political Turbulence. *The Quarterly Review of Economics and Finance, 73*, 95–110. https://doi.org/10.1016/j.qref.2018.09.007.

Balcilar, M., Nugent, J. B., & Xu, J. (2022). Adversities in syria and their relation to their physical and mental health conditions as syrian refugees in turkey. *Scottish Journal of Political Economy, 69*(1), 37–59. https://doi.org/10.1111/sjpe.12295.

Barbera, P. (2015). Birds of the same feather tweet together: Bayesian ideal point estimation using twitter data. *Political analysis, 23*(1), 76–91. https://doi.org/10.1093/pan/mpu011.

Bove, V., Di Salvatore, J., & Elia, L. (2021). "What it takes to return: UN peacekeeping and the safe return of displaced people", *Unpublished Working paper*. Commissioned as part of the "Preventing Social Conflict and Promoting Social Cohesion in Forced Displacement Contexts" Series. Washington, DC: World Bank Group.

Bradley, M. (2013). *Refugee repatriation: Justice, responsibility and redress*. Cambridge, UK: Cambridge University Press.

Braithwaite, A., Chu, T. S., Curtis, J., & Ghosn, F. (2019). Violence and the perception of risk associated with hosting refugees. *Public Choice, 178*(3), 473–492. https://doi.org/10.1007/s11127-018-0599-0.

Camarena, K. R., & Hagerdal, N. (2020). When do displaced persons return? postwar migration among christians in mount lebanon. *American Journal of Political Science, 64*(2), 223–239. https://doi.org/10.1111/ajps.12500.

Central Bureau of Statistics. (2004). *Population and housing census 2004*. Central Bureau of Statistics, Syrian Arab Republic. Retrieved from https://catalog.ihsn.org/index.php/catalog/4085.

Chatfield, A. T., Reddick, C. G., & Brajawidagda, U. (2015). Tweeting propaganda, radicalization and recruitment: Islamic state supporters multi-sided twitter networks. In *Proceedings of the 16th annual international conference on digital government research* (p. 239–249). New York, NY, USA: Association for Computing Machinery. 10.1145/2757401.2757408.

Chen, Y., & Skiena, S. (2014). Building sentiment lexicons for all major languages. In *Proceedings of the 52nd annual meeting of the association for computational linguistics (short papers)* (pp. 383–389). 10.3115/v1/P14-2063.

Crilley, R. (2017). Seeing syria: The visual politics of the national coalition of syrian revolution and opposition forces on facebook. *Middle East Journal of Culture and Communication, 10*(2–3), 133–158. https://doi.org/10.1163/18739865-01002004.

Davenport, C., Moore, W., & Poe, S. (2003). Sometimes you just have to leave: Domestic threats and forced migration, 1964–1989. *International Interactions, 29*(1), 27–55. https://doi.org/10.1080/03050620304597.

Deng, J., Dong, W., Socher, R., Li, L.-J., Li, K., & Fei-Fei, L. (2009). Imagenet: A large scale hierarchical image database. In *2009 ieee conference on computer vision and pattern recognition* (pp. 248–255). 10.1109/CVPR.2009.5206848.

Elbadawi, I., Fallah, B., Louis, M., Makdisi, S., Youssef, J., Albinyana, R., & Tumen, S. (2019). Repatriation of refugees from arab conflicts: Conditions, costs and scenarios for reconstruction. FEMISE. Technical Report. Retrieved from https://www.femise.org/wp-content/uploads/2019/09/FEMISE EuroMed4-FINAL-small-upd.pdf.

Erdogan, M. (2020). *Syrians barometer 2019*. Ankara, TR: UNHCR.

Erdogan, M. (2021). *Syrians barometer 2020*. Ankara, TR: UNHCR.

Freelon, D., Lynch, M., & Aday, S. (2015). Online fragmentation in wartime: A longitudinal analysis of tweets about syria, 2011–2013. *The ANNALS of the American Academy of Political and Social Science, 659*(1), 166–179. https://doi.org/10.1177/0002716214563.

Frouws, B., Phillips, M., Hassan, A., & Twigt, M. (2016). Getting to europe the whatsapp way: The use of ict in contemporary mixed migration flows to europe. *SSRN Electronic Journal*. https://doi.org/10.2139/ssrn.2862592.

Getmansky, A., Sınmazdemir, T., & Zeitzoff, T. (2018). Refugees, xenophobia, and domestic conflict: Evidence from a survey experiment in turkey. *Journal of Peace Research, 55*(4), 491–507. https://doi.org/10.1177/0022343317748719.

Ghosn, F., Chu, T. S., Simon, M., Braithwaite, A., & Jandali, M. F. J. (2021). The journey home: Violence, anchoring, and refugee decisions to return. *American Political Science Review, 115*(3), 982–998. https://doi.org/10.1017/S0003055421000344.

Gohdes, A. R. (2020). Repression technology: Internet accessibility and state violence. *American Journal of Political Science, 64*(3), 488–503. https://doi.org/10.1111/ajps.12509.

He, K., Zhang, X., Ren, S., & Sun, J. (2016). Deep residual learning for image recognition. In *Proceedings of the ieee conference on computer vision and pattern recognition* (pp. 770– 778). 10.48550/arXiv.1512.03385.

Hoffman, M., & Makovsky, A. (2021). Northern syria security dynamics and the refugee crisis. Center for American Progress. Technical Report. Retrieved from https://www.americanprogress.org/article/northern-syria-security-dynamics-refugee-crisis/.

Hoogeveen, J. G., Rossi, M., & Sansone, D. (2019). Leaving, staying or coming back? migration decisions during the northern mali conflict. *The Journal of Development Studies, 55*(10), 2089–2105. https://doi.org/10.1080/00220388.2018.1510119.

Jalabi, S. (2021). Attitudes toward emigration in the syrian capital of damascus: A survey in three neighborhoods. Operations and Policy Center. Technical Report. Retrieved from https://opc.center/attitudes-toward-emigration-in-the-syrian-capital-of-damascus-a-survey-in-three-neighborhoods/.

Kaplan, O. (2021). *"Superstitions and civilian displacement: Evidence from the colombian conflict", Unpublished Working paper. Commissioned as part of the "Preventing Social Conflict and Promoting Social Cohesion in Forced Displacement Contexts" Series. Washington, DC: World Bank Group.*

Khamis, S., Gold, P. B., & Vaughn, K. (2012). Beyond egypt's "facebook revolution" and syria's "youtube uprising": Comparing political contexts, actors and communication strategies. *Arab Media & Society, 15*(spring), 1–30.

Kirisci, K. (2014). Syrian refugees and turkey's challenges: Going beyond hospitality. Brookings Washington, DC. Technical Report. Retrieved from https://www.brookings.edu/research/syrian-refugees-and-turkeys-challengesgoing-beyond-hospitality/.

Klausen, J. (2015). Tweeting the jihad: Social media networks of western foreign fighters in syria and iraq. *Studies in Conflict & Terrorism*, *38* (1), 1–22. 10.1080/1057610X.2014.974948.

Krishnan, N., Russo Riva, F., Sharma, D., & Vishwanath, T. (2020, July). The lives and livelihoods of syrian refugees in the middle east: Evidence from the 2015-16 surveys of syrian refugees and host communities in jordan, lebanon and kurdistan, iraq. Policy Research Working Paper; No. 9327. World Bank, Washington, DC. Retrieved from http://hdl.handle.net/10986/34173.

Kumar, K. B., Culbertson, S., Constant, L., Nataraj, S., Unlu, F., Bouskill, K. E., . . . Afashe, F. (2018). *Opportunities for all: Mutually beneficial opportunities for syrians and host countries in middle eastern labor markets*. Santa Monica, CA: RAND Corporation. 10.7249/RR2653.

Lazarev, E., & Sharma, K. (2017). Brother or burden: An experiment on reducing prejudice toward syrian refugees in turkey. *Political Science Research and Methods, 5*(2), 201. https://doi.org/10.1017/psrm.2015.57.

Lischer, S. K. (2006). *Dangerous sanctuaries: Refugee camps, civil war, and the dilemmas of humanitarian aid*. Ithaca, NY: Cornell University Press.

Lui, M., & Baldwin, T. (2012). langid.py: An off-the-shelf language identification tool. In *Proceedings of the ACL 2012 system demonstrations* (pp. 25–30). Jeju Island, Korea: Association for Computational Linguistics. Retrieved from https://aclanthology.org/P12-3005.

Lynch, M., Freelon, D., & Aday, S. (2014). Syria's socially mediated civil war. United States Institute of Peace. Technical Report. Retrieved from https://www.usip.org/publications/2014/01/syrias-socially-mediated-civil-war.

Martin, S. F., Davis, R., Benton, G., & Waliany, Z. (2019). International responsibility-sharing for refugees: Perspectives from the mena region. *Geopolitics, History and International Relations, 11*(1), 59–91. https://doi.org/10.22381/GHIR11120193.

McDowell-Smith, A., Speckhard, A., & Yayla, A. S. (2017). Beating isis in the digital space: Focus testing isis defector counter-narrative videos with american college students. *Journal for Deradicalization, 10*, 50–76.

Metzger, M. M., & Siegel, A. A. (2019). When state-sponsored media goes viral: Russia's use of rt to shape global discourse on syria. *Unpublished working paper* .

Miconi, A. (2020). News from the levant: A qualitative research on the role of social media in syrian diaspora. *Social Media+ Society*, *6* (1), 2056305119900337. 10.1177/2056305119900337.

Mikolov, T., Sutskever, I., Chen, K., Corrado, G. S., & Dean, J. (2013). Distributed representations of words and phrases and their compositionality. In *Advances in neural information processing systems* (pp. 3111–3119). 10.48550/arXiv.1310.4546.

Mimno, D., Wallach, H., Talley, E., Leenders, M., & McCallum, A. (2011). Optimizing semantic coherence in topic models. In *Proceedings of the 2011 conference on empirical methods in natural language processing* (pp. 262–272).

Mitts, T. (2019). From isolation to radicalization: Anti-muslim hostility and support for isis in the west. *American Political Science Review, 113*(1), 173–194.

Moore, W. H., & Shellman, S. M. (2006). Refugee or internally displaced person? to where should one flee? *Comparative Political Studies, 39*(5), 599–622. https://doi.org/10.1177/0010414005276457.

Munger, K., Guess, A. M., & Hargittai, E. (2021). Quantitative description of digital media: A modest proposal to disrupt academic publishing. *Journal of Quantitative Description: Digital Media*, *1* . 10.51685/jqd.2021.000.

Nielsen, R. A. (2017). *Deadly clerics: Blocked ambition and the paths to jihad*. Cambridge, UK: Cambridge University Press. 10.1017/9781108241700.

Nielsen, S. Y. (2016). Perceptions between syrian refugees and their host community. *Turkish Policy Quarterly, 15*(3), 99–106.

Norwegian Refugee Council. (2021). Syria: Another decade of crisis on the horizon expected to displace millions more. Norwegian Refugee Council. Technical Report. Retrieved from https://www.nrc.no/news/2021/march/syria-another-decade-of-crisis-on-the-horizon-expected-to-displace-millions-more/.

Prucha, N. (2016). Is and the jihadist information highway–projecting influence and religious identity via telegram. *Perspectives on Terrorism, 10*(6), 48–58.

Raleigh, C., Linke, A., Hegre, H., & Karlsen, J. (2010). Introducing acled: An armed conflict location and event dataset: Special data feature. *Journal of Peace Research, 47*(5), 651–660. https://doi.org/10.1177/0022343310378914.

Ramadan, R. (2017). Questioning the role of facebook in maintaining syrian social capital during the syrian crisis. *Heliyon, 3*(12), e00483.

REACH Resource Center. (2018-2020). Humanitarian situation overview of syria (hsos). REACH Resource Center. Dataset. Retrieved from https://www.reach-initiative.org/where-we-work/syria/.

Roberts, M. E., Stewart, B. M., Tingley, D., Lucas, C., Leder-Luis, J., Gadarian, S. K., ... Rand, D. G. (2014). Structural topic models for open-ended survey responses. *American Journal of Political Science, 58*(4), 1064–1082. https://doi.org/10.1111/ajps.12103.

Ruegger, S. (2013). Refugee flows, transnational ethnic linkages and conflict diffusion: Evidence from the kosovo refugee crisis. *Presentation at the rrpp annual conference*.

Schober, M. F., Pasek, J., Guggenheim, L., Lampe, C., & Conrad, F. G. (2016). Social media analyses for social measurement. *Public Opinion Quarterly, 80*(1), 180–211. https://doi.org/10.1093/poq/nfv048.

Schon, J. (2019). Motivation and opportunity for conflict-induced migration: An analysis of syrian migration timing. *Journal of Peace Research, 56*(1), 12–27. https://doi.org/10.1177/0022343318806044.

Siegel, A. A., & Badaan, V. (2020). # no2sectarianism: Experimental approaches to reducing sectarian hate speech online. *American Political Science Review, 114*(3), 837–855.

Singh, L., Donato, K., Arab, A., Belon, T. A., Fraifeld, A., Fulmer, S., . . . Wang, Y. (2020). Identifying meaningful indirect indicators of migration for different conflicts. arXiv. 10.48550/arXiv.2007.06116.

Singh, L., Wahedi, L., Wang, Y., Wei, Y., Kirov, C., Martin, S., ... Kawintiranon, K. (2019). Blending noisy social media signals with traditional movement variables to predict forced migration. In *Proceedings of the 25th acm sigkdd international conference on knowledge discovery & data mining* (pp. 1975–1983). https://doi.org/10.1145/3292500.

Steele, A. (2019). Civilian resettlement patterns in civil war. *Journal of peace research, 56*(1), 28–41. https://doi.org/10.1177/0022343318820576.

Śanchez-Querubın, N., & Rogers, R. (2018). Connected routes: Migration studies with digital devices and platforms. *Social Media + Society*, *4* (1), 2056305118764427. 10.1177/2056305118764427.

UN OCHA. (2021). Syrian arab republic: Idp movements and idp spontaneous return movements data. United Nations Office for the Coordination of Humanitarian Affairs. Dataset. Retrieved from https://data.humdata.org/dataset/syrian-arab-republic-idp-movements-and-idp-spontaneous-return-movements-data.

UNHCR. (2019). Fifth Regional Survey on Syrian Refugees' Perceptions and Intentions on Return to Syria. UN Refugee Agency. Technical Report. Retrieved from https://reliefweb.int/sites/reliefweb.int/files/resources/68443.pdf.

UNHCR. (2020). 1 per cent of humanity displaced: UNHCR global trends report. UN Refugee Agency. Technical Report. Retrieved from https://www.unhcr.org/en-us/news/press/2020/6/5ee9db2e4/1-cent-humanity-displaced-unhcr-global-trends-report.html.

UNHCR. (2021). Syria Regional Refugee Response. UN Refugee Agency. Technical Report. https://data.unhcr.org/en/situations/syria#ga=2.202775441.1694891000.1621886375-1110293497.1621297771.

Urman, A., Ho, J.-C.-T., & Katz, S. (2020). "No Central Stage": Telegram-based activity during the 2019 protests in hong kong. *SocArXiv.*. https://doi.org/10.31235/osf.io/ueds4.

Walk, E., Parker-Magyar, E., Garimella, K., Akbiyik, A., & Christia, F. (2021). *Social media narratives on conflict from northern syria. SSRN MIT Political Science Department Research Paper No. 2022-2.* dx.doi.org/10.2139/ssrn.4075120.

Wei, Y., Singh, L., & Martin, S. (2016). Identification of extremism on twitter. In *2016 ieee/acm international conference on advances in social networks analysis and mining (asonam)* (pp. 1251–1255). https://doi.org/10.1109/ASONAM.2016.7752398.

World Bank. (2019). The mobility of displaced syrians: An economic and social analysis. Washington, DC: World Bank Group. Technical Report. Retrieved from https://www.worldbank.org/en/country/syria/publication/the-mobility-of-displaced-syrians-an-economic-and-social-analysis.

World Bank. (2020). Compounding misfortunes: Changes in poverty since the onset of covid-19 on syrian refugees and host communities in jordan, the kurdistan region of iraq and lebanon. Washington, DC: World Bank Group. Technical Report. Retrieved from https://www.worldbank.org/en/region/mena/publication/compounding-misfortunes-changes-in-poverty-since-the-onset-of-covid-19-on-syrian-refugees.

World Food Programme. (2021). Twelve million syrians now in the grip of hunger, worn down by conflict and soaring food prices. UN World Food Programme. News Release. Retrieved from https://www.wfp.org/news/twelve-million-syrians-now-grip-hunger-worn-down-conflict-and-soaring-food-prices.

Yahya, M., Kassir, J., & El-Hariri, K. (2018, April 16). Unheard voices: What syrian refugees need to return home. Carnegie Endowment for International Peace. Technical Report. Retrieved from https://carnegieendowment.org/2018/04/16/unheard-voices-what-syrian-refugees-need-to-return-home-pub-76050.

Yayla, A. S., & Speckhard, A. (2017). Telegram: The mighty application that isis loves. International Center for the Study of Violent Extremism. Retrieved from https://www.icsve.org/telegram-the-mighty-application-that-isis-loves/.

Zhang, H., & Pan, J. (2019). Casm: A deep-learning approach for identifying collective action events with text and image data from social media. *Sociological Methodology, 49*(1), 1–57. https://doi.org/10.1177/0081175019860244.

Zolberg, A. R., Suhrke, A., & Aguayo, S. (1989). *Escape from violence: Conflict and the refugee crisis in the developing world*. Oxford, UK: Oxford University Press.