Civilian Behavior on Social Media During Civil War

Anita R. Gohdes*[1] and Zachary C. Steinert-Threlkeld†[2]

[1]Hertie School, Berlin

[2]University of California, Los Angeles

July 4, 2023

Abstract

Recent research emphasizes social media's potential for citizens to express shared grievances. In active conflict, however, social media posts indicating political loyalties can pose severe risks to civilians. We develop a theory that explains how civilians modify their online behavior as part of efforts to improve their security during conflict. After major changes in territorial control, civilians should be more likely to post positive content, and more content that supports the winning side. We study social media behavior during and after the siege of Aleppo in November 2016. We match Aleppo-based Twitter users with users from other parts of Syria and use large language models to analyze changes in online behavior after the regime's retaking of the city. Results show that users in Aleppo post more positive and pro-Assad content, but only when self-disclosing their location. The findings have important implications for our understanding of digital communication in civil conflict.

*gohdes@hertie-school.org

†zst@luskin.ucla.edu

Introduction

Social media are a central companion to contentious political processes. From isolated protests to country-wide uprisings to organized armed conflict, civilians and government actors routinely use social media to document and participate in such events. A growing body of research studies how social media reduces the cost of coordination (Little, 2015), provides passive polling to conflict actors (Zeitzoff, 2017), is used by non-state groups as a tool for direct diplomacy (Jones and Mattiacci, 2019), and by state actors for spreading propaganda during domestic upheaval (Stukal et al., 2022). Yet in the context of armed conflict, little is currently known about how civilians who are caught in the midst of fighting use social media (an exception is Walk et al., 2022).

This paper develops and tests a theory of civilians' use of social media during civil conflict. Civilians use social media to strategically signal both tacit and overt support for armed actors, in particular in conflicts where at least one side actively monitors social media for intelligence. Existing work on the agency and role of civilians in conflict analyzes how individuals adapt their behavior in response to changing conflict dynamics and highlights the intentionality with which civilians seek security (Lyall, Blair and Imai, 2013, Popkin, 1979, Schubiger, 2021). Just as a wide variety of offline behaviors are replicated online (Bisbee and Larson, 2017), civilians use social media to seek security the same way they do offline. This paper draws on three previously disconnected literatures, civilian behavior during civil war, online behavior, and digital authoritarianism to develop and test theoretical expectations about the strategic use of social media by civilians caught in conflict.

In the context of armed conflict, major changes in territorial control produce situations of information scarcity that increase the incentives of civilians to publicly show support for the victorious side. An armed group that has (re)taken territory is keenly interested in collecting information about civilians' loyalties, and social media monitoring supports this process. In the aftermath of shifts in territorial dynamics, civilians should therefore modify their online behavior by publishing more content in support of the local rulers, more content that is positive, and users who support the victorious side will increase their activity. Moreover, users can geolocate posts to specific locations, facilitating the work of monitors who seek to map users to their location. Strategic online behavior should therefore be strongest in content where individuals disclose their location information.

Posting on social media is costly, especially during conflict. Across the world, social media has become a vital if imperfect means of receiving and exchanging real-time information on conflict dynamics (Schon, 2021). At the same time, governments monitor social media to better understand who supports them, and many security services use it to better predict unrest and more efficiently target those who oppose the political status quo (Feldstein, 2021, Qin, Strömberg and Wu, 2017). In 2018, for example, the Egyptian government arrested Amal Fathy for posting a critical video to her Facebook page (Cullinane, 2018). Bahrain's Ministry of the Interior routinely arrests individuals for "misuse of social media" (IFEX, 2018), and China has arrested citizens who post on Twitter, a social media service most people in Chinese cannot access (Wong, 2021). Posting political content online can thus have severe consequences for the offline security of individuals (Pan and Siegel, 2020).

This theory of online civilian behavior during civil war is tested using social media data from the Syrian conflict, focusing specifically on dynamics around the end of the siege of Aleppo, Syria's most populous city before the civil war. The siege ended on December 15, 2016 with a victory for the Assad regime. Accordingly, civilians based in Aleppo should post more pro-Assad (i.e. pro-regime) and more positive content once the siege ends, in particular in posts tagged with location information. Moreover, the Syrian regime has long placed special importance on public signals of loyalty combined with the widespread use of social media during the conflict (Lynch, Freelon and Aday, 2014, Wedeen, 1999). Twitter is used to measure behaviors since it was used intensively use by actors supportive, antagonistic and in opposition to the Assad regime (Freelon, Lynch and Aday, 2015).

Twitter users in Aleppo are matched to users in other parts of Syria to understand how the Syrian regime's retaking of the city impacted Aleppo-based users. Pro- and anti-Assad content is detected via fine-tuning ARBERT, a leading large language model designed for Arabic text (Abdul-Mageed, Elmadany and Bill, 2021), and sentiment is measured via a model trained on Levantine Arabic tweets (Baly et al., 2019). Accounts are also labeled manually in order to understand if they overtly supported or opposed the Assad regime before the end of the siege.

The results suggest a range of changes in behavior compatible with theoretical expectations of strategic signalling. In the aftermath of the siege users based in Aleppo are more likely to share pro-Assad and positive content, but only in content where they self-disclosed their location. Content that does not include location information is more likely to be *anti*-Assad in Aleppo once the siege ends. There are no changes post-siege elsewhere in Syria. After the siege's end, there is an increase in the likelihood of tweets being sent from pro-Assad accounts, suggesting that support for the regime is also signalled through increased activity by its supporters. These results are robust to additional tests that take into account a fuzzy cut-off for the end of the siege, placebo dates for the siege's end, and a placebo treatment group. The results do not change: civilians use social media strategically during civil conflict.

Signalling support in conflict

Who do civilians support in civil war?

At the heart of much research on the dynamics of violence in civil war as well as the outcomes of conflicts lies the question of whom civilians choose to support during wartime (Kalyvas and Kocher, 2007). Only a fraction of civilians end up becoming combatants, but support can manifest via other channels. Local support can be pivotal for conflict actors as their ability to operate in a territory is facilitated by the engagement of the local population. A core way in which civilian support is therefore frequently measured is through their willingness to cooperate with an armed actor. Kalyvas (2006) contends that incentives to collaborate are principally driven by who controls a given territory, and that civilians will be more willing to collaborate with armed groups where they need not fear reprisals. Overall, he argues that

> [i]rrespective of their preference (and everything else being equal), most people prefer to collaborate with the political actor that best guarantees their survival. However, collaboration is much more uncertain in areas of fragmented sovereignty where control is incomplete. (Kalyvas, 2008, 406)

In conflict settings, collaboration can take on many forms. In addition to providing food, supplies, and shelter, non-material support in the form of public displays of loyalty, solidarity, and positive morale is a key method of signalling support for armed actors. Outside of conflict, behaviors such as displaying pro-government slogans in windows, whether at home or one's business, is a common signal (Havel and Wilson, 1986). Singing opposition songs or flying their flags is a more identifiable, and therefore riskier, signal (Pfaff, 1996).

Because civilian support is so crucial to winning wars and maintaining control over territory, conflict actors have a strong incentive to monitor whether and how civilians display loyalties towards the conflict parties. Governments that are confronted with insurgent groups will be particularly intent on undercutting local support for insurgents (Valentino, Huth and Balch-Lindsay, 2004) and fostering public displays of loyalty for the state. Many authoritarian governments, including Syria's, routinely compel public displays of support, even if the extent to which these signals reflect true preferences tends to be uncertain (Kuran, 1991, Wedeen, 1999).

The significance of obtaining local support is also reflected in counterinsurgency strategy contending that improving civilians' attitudes towards the government are likely to reduce support for insurgencies (Berman, Shapiro and Felter, 2011). Winning the 'hearts and minds' of civilians does not imply, however, that individuals have to enthusiastically embrace an armed group. Instead 'calculated self-interest, not emotion, is what counts(U.S. Army / Marine Corps, 2007, 294)'. This understanding of support builds on the assumption that civilians caught in conflict will primarily be driven by rational self-interest, which implies that non-combatants will be non-ideological about whom to support (Popkin, 1979).

Civilians caught in the midst of conflict should therefore have an incentive to support armed groups when they anticipate that doing so will increase their chances of security. Prior research offers support for this assumption, and indicates that conflict dynamics are likely to impact civilians' incentives to show support for one of the conflict actors, or show none at all (Lyall, 2009). Civilians may engage in pro-government mobilization to credibly signal their allegiance with the government in an effort to avoid indiscriminate violence (Schubiger, 2021). For individuals with indeterminate beliefs showing public displays of support may be of particular importance; these 'gray people' (*al-ramadiyyin*) are vulnerable to both sides of a conflict if they do not make their loyalty explicit and unambiguous (Wedeen, 2019).

The argument about social media as a tool for strategically communicating public support is distinct from earlier claims about the coordinating effects of social media (Little, 2015). Initial enthusiasm about social media focused on its potential to create political coordination, allowing the disaffected who would otherwise feel alone to realize their preferences may be closer to their polity's median preferences than they thought (Kuran, 1991). However, social media likely no longer provides political coordination that favors one side to a conflict, as pro-status quo individuals regularly signal their support for current policies (Munger et al., 2019), and state-affiliated actors have been found to directly inject pro-status quo messages via automated accounts and paid content producers (Lukito, 2019). While the written language used in social media posts could reflect a strategic choice to signal allegiance to one side during a conflict, it appears instead that on social media, language serves as a focal point so that both sides understand each others' signal (Metzger, Nagler and Tucker, 2015).

Territorial changes, uncertainty, and social media signalling

In situations of incomplete information actors are likely to rely on visible, costly signals to communicate their support (Spence, 1973). More formally, signals 'are the stuff of purposive communication. Signals are any observable features of an agent which are intentionally displayed for the purpose of raising the probability the receiver assigns to a certain state of affairs (Gambetta, 2009, 170)'. These signals can be necessary to secure the provision of private goods (Kalyvas and Kocher, 2007, 182).

During armed conflict, the degree to which information about civilians' loyalties is both relevant and available fluctuates. A major factor affecting both the relevance and availability of information on civilian support are major changes in who controls a given locality. Territorial shifts from fragmented control, where at least two armed actors battle for control, to situations where one conflict party (re)gains complete control of an area such as a region, city, or neighbourhood, are likely to incentivize signalling. In the aftermath of such shifts, there will be significant uncertainty about local support for the winning side. At the same time, the victorious side will also be more interested in knowing who is loyal towards them and who should be punished for having collaborated with the enemy.

In the aftermath of such shifts, the value of showing support for the winning side thus likely increases. On social media, there should therefore be an increase in content supporting the victor. Individuals may signal via the topics of their posts. Changing the topic of what is publicly said is an overt signal of support for an armed actor. For example, the statement 'I support the military's success' is an overt display of support in response to a change in battlefield condition. Since individuals can post about any number of potential topics, expressing support or hostility after a change in battlefield conditions is a strategic decision.

Changes in sentiment also signal tacit support for changes in local control. Composing messages expressing either positive or negative sentiment is an intentional choice by users and reflects their willingness to publicly communicate preferences. Sentiment is a signal for three reasons. First, individuals who are loyal to the winning side will feel emboldened to share positive and supportive content. Positive and public support of the winning side by civilians who are traditionally supporters of the victor partly stems from the strategic objective of making the victor aware of the individual's ongoing loyalty. Second, previously neutral civilians are likely to express support for the victor. Third, even supporters of the losing party will falsify their preferences online, just like they have offline (Kuran, 1991). This logic leads to the following *content* hypotheses:

H1a. Changes in territorial control will lead to increased content supporting the winning side.

H1b. Changes in territorial control will lead to increase content containing positive sentiment.

Signalling can also occur through activity patterns regardless of the content of that activity. While individuals may switch from making neutral or negative statements to posting positive (i.e. pro-victor) content, such actions can feel too risky or emotionally difficult for many. Social media users may want to signal support for the regime by increasing their online activity. Where government forces are known to monitor social media, civilians are likely to be aware of the fact that the content they post online could lead to questioning or even arrest if it is deemed critical of regime activity (Tibken, 2016). For example, individuals in China who joined Twitter during COVID-19 lockdowns posted content whose topics were neutral or pro-regime, but they chose to follow accounts posting sensitive political information blocked within China (Chang et al., 2022). This logic leads to the following *activity* hypothesis:

H2. Changes in territorial control will lead to increased activity of users that support the winning side.

This assertion does not assume that civilian support will *always* and *only* be strategic; identity, ethnicity, and ideology can significantly affect who civilians choose to support and whether they will change their public behavior. For example, group identity mediates civilian support for armed groups, where violence perpetrated by members of the civilian's in-group is less likely to trigger support for out-group actors, but victimisation by out-group actors will trigger more support for in-group actors (Lyall, Blair and Imai, 2013). In addition, some individuals may react more emphatically to local changes in the security situation than others. Rather than assume that every civilian in a certain locality will alter their public support for an armed actor, we expect that, *on average*, changes in the security situation in the aftermath of shifts in territorial control will lead a higher proportion of locals to publicly support the winning party.

The costliness of social media support

Contrary to skeptics' beliefs, posting to social media is costly (Morozov, 2012). In fact, digital traces left by messages posted in public support for a conflict party provide a visible signal to both sides of a conflict; this visibility puts the poster in danger and therefore makes the signal credible.

Individuals will use social media as a means of signalling overt and tacit support in conflict situations where the following three criteria are met. First, public facing social media platforms are widely available and used throughout the conflict. Second, at least one of the conflict parties monitors social media for intelligence. The third criterion is that individuals are aware of the monitoring activities and have witnessed conflict actors gleaning information on defectors or dissidents from social media.

Where individuals are aware that conflict actors could weaponize their public social media content they are also aware that posting online is costly. Signalling preferences on social media during conflict can have direct implications for the user's security situation. Social media monitoring by state security forces is widespread, not only in conflict situations (Feldstein, 2021). Social media is monitored and analyzed in order to identify individuals who could pose a threat to the political status quo through communicating messages that are critical of the government, or by recruiting others to join them in calling for change. The networked character of social media allows authorities to also monitor the friends and followers of identified targets and to calibrate their repressive responses based on the information collected (Gohdes, 2020, Xu, 2021).

In the context of ongoing armed conflict, social media posts commenting on local violence and politics can pose a severe risk to civilians' security. Posting in support of a rebel group may lead to interrogations and other forms of coercion by government forces. Warring non-state parties may be just as interested in collecting information on civilian 'troublemakers' or 'regime collaborators' as state forces are. Following pro-government accounts or speaking out is support of the military online may therefore lead to increased risks of targeting through rebel groups. For example, when US forces withdrew from Afghanistan in 2021, reports of the Taliban combing through peoples' social media profiles for traces of interactions with the United States Army were so prevalent and fear-inducing that Facebook temporarily introduced new security features specifically for Afghan users (Glaser and Smith, 2021).

In addition, social media usage during ongoing conflict is likely to be more political than during peacetime. This is because users will be more likely to post political content related to the war when their daily lives are dominated by violent struggles and contentious episodes will increase the political valence of any form of online communication. In digitally mediated conflicts where social media is used by a large proportion of society, the *absence* of political positioning online as well as the absence of *any* messaging online sends a public signal that can potentially have adverse consequences.

The costliness of posting online will also depend on the ease with which conflict parties are able to access content and identify its location. In the context of armed conflict, conventional filter devices such as hashtags are inefficient because analysts, news outlets, diaspora, and foreign observers dilute the hashtag's usefulness for local monitoring. Some social media platforms allow users to add exact geo-location to their messages, thereby publicly revealing their location at the time of posting. Social media monitors can use simple search tools to filter content from specific localities, significantly facilitating their access to local information. Posting sensitive content while disclosing one's location is thus likely to make this content available to a broader public and therefore be more costly than posts without location information. Because users have the ability to selectively disclose their location, there should be variation in the type of content posted with location information versus content posted without it. This logic leads to the following *location* hypothesis:

H3. Content that includes individuals' location information is more likely to evidence strategic behavior.

The Syrian Conflict

We study the Syrian conflict, focusing specifically on social media content and activity during and after the regime's siege of Aleppo at the end of 2016. The Syrian conflict has been called the most socially mediated civil conflict in history (Lynch, Freelon and Aday, 2014, 5), with some commentators going to far as to claim that the Internet itself has become a weapon of war (Hashem, 2015). Despite experts' warning about social media and the seeming abundance of data giving observers the illusion of complete information about events in Syria (e.g. Lynch, Freelon and Aday, 2014), little research exists on how dynamics of the conflict itself, such as major changes in the territorial control, affect online signalling.

In March 2011, peaceful protesters took to the streets in the city of Daraa to protest the imprisonment and torture of young teenagers by the regime. The protests quickly spread to other cities and were met with extreme brutality by regime forces. The massacre of the first protesters in Daraa was shared widely on YouTube, Twitter, and Facebook. The continued violent crackdown on ever growing demonstrations across the country developed into an armed internal struggle, and by the beginning of 2012 the government commenced its use of heavy artillery fire, and aerial bombardments would form a core military strategy of the Assad regime. In September 2015, Russian forces commenced their military intervention in the conflict, supporting the Syrian regime's military offensive aimed at regaining territorial control.

Actors on all sides of the conflict have made use of social media to communicate their (change in) allegiances, spread propaganda, and communicate with both their domestic and foreign audience (Moss, 2018). As peaceful protest turned into armed resistance, local cells of the Free Syrian Army used Facebook groups to distribute news, while the Syrian Electronic Army used Facebook to identify activists and establish pro-regime signals (Shehabat, 2012). Civilians used social media platforms such as Facebook, Telegram, and Twitter to discuss the war's daily events, compete to establish dominant frames, and publicly signal support for various combatants (Walk et al., 2022). While major bans on social media platforms were lifted a few month prior to the outbreak of the conflict in 2011, Internet activity remained highly monitored and controlled by the Syrian regime (Freedom House, 2015). Internet access has been shut down countrywide numerous times, and the regime has strategically limited access in certain governorates as part of their broader repressive strategy (Gohdes, 2020).

The siege of Aleppo

The siege of Aleppo represented a culmination of extreme measures taken by the Syrian government and its allies to re-establish territorial control over Syria's most populous city. Syrian government forces and rebels battled for control of Aleppo starting in mid 2012 and following months of protests. By November 2012 the city was deadlocked into a regime and a rebel controlled area. Aleppo was at the center of some of the worst fighting, with the regime using barrel bombs to attack the rebel-held areas. With the Russian intervention in the conflict in September of 2015, regime forces were joined by Russian airpower in targeting rebel held areas in Aleppo.

By July 2016, the Syrian Army had cut the last supply line into the city, effectively placing the entire city under siege. Throughout the conflict, regime forces have repeatedly employed siege tactics as a mean of forcing rebel forces to surrender, such as in Deraa in April 2011, and in Rural Damascus in the Spring of 2013 (Todman, 2017). Sieges represent an extreme form of indiscriminate coercion (or: 'collective punishment') aimed at forcing the enemy to surrender a specific geographic location through endangering the lives of all inhabitants in this area.

In November 2016, regime forces circled the remaining densely populated rebel-held areas in the Eastern part of Aleppo, and, with support of Russian air forces, and submitted the area to intense bombardment for twelve days while targeting core civilian infrastructure such as hospitals (Böttcher, 2017, 2-3). On December 13, a complex ceasefire was negotiated that involved the handover of weapons and transfer of all remaining rebels to other territories, resulting in an estimated relocation of one hundred thousand individuals (Bassam, McDowall and Nebehay, 2016), which continued through to December 15. The siege left the city destitute with tens of thousands dead, and many more close to starvation. The siege's end marks the major shift in territorial control. We expect tht the siege's end is when strategic behavior should manifest itself.

Data and Methods

Twitter data is used to study social media usage. A sample of geotagged tweets and panel data based on that sample are created. Content is measured via a fine-tuned large language model (LLM) for topics and a separate LLM for sentiment, while activity is measured based on accounts' inferred affinity for the Assad regime.

Why Twitter?

Twitter is a predominately public site that Syrians have used to discuss publicly and contentiously the ongoing civil war (Walk et al., 2022). New users from Syria have consistently joined and tweeted throughout the period under study. With about 90% of Twitter accounts being public, it is the most likely platform on which users will strategically adapt their public messaging.[1] Furthermore, accounts on it are heterogeneous in their politics: Freelon, Lynch and Aday (2015) show that there are, among others, secular revolutionaries, Islamists, sup-

[1] YouTube is primarily a source of documentation of conflict but reveals little about *who* is participating, i.e. there is separation between the sharer and the content shared. Chat applications like Telegram and WhatsApp encourage communication within small groups; while important for some types of behavior related to conflict, signalling during major offline changes in control, the type of event studied in this paper, is less likely to be important in such a setting since any signalling will have a small audience. The same logic is true of Facebook: the vast majority of accounts are private, limiting the number of people who could receive a signal and therefore the utility of signalling. Facebook *Pages* tend to be public, but they are left-censored, mean that researchers only have access to Pages above a threshold amount of engagement. Individual behavior is also not represented by aggregate *Pages*.

porters of the Free Syrian Army, and pro-Assad accounts active during the early years of the Syrian conflict. Lastly, studying the change of user behavior online requires *geographic* information to analyze local-level changes in user behavior, and few other social media sites make this type of geographic self-reporting by users available for research.

Although Twitter is widely used in Syria and has been an essential communication tool during throughout the conflict it is important to note that individuals who use social media are not representative of the general population. According to the International Telecommunication Union (ITU), 32% of all Syrians had access to the Internet in 2016 (International Telecommunication Union (ITU), 2022). Social media users also tend to be younger, more educated, and tech savvy than those who do not post online. In Syria, the opening up of previously censored online spaces coincided with the initial uprising, and by consequence social media quickly became ingratiated as an important political outlet (Brownlee, 2018).

Data Collection

To test the theoretical expectations, geotagged content from Twitter is collected. Users in Aleppo are matched to comparable users in other parts of Syria based on their content. Each user's entire tweet history is then collected in order to reflect their entire public communication on the platform.[2]

Download geotagged tweets. First, geotagged tweets are downloaded from Twitter's POST statuses/filter endpoint, often known as the 'streaming API'. Geotagging refers to tweets that include longitude and latitude coordinates because the user tagged their location in the tweet. Accounts must first enable geotagging in their settings and then choose to assign a location to a tweet. Through the streaming API, between one-third and one-half of all tweets with coordinates are collected. The stored tweets are then queried for those sent from Syria, focusing on the three months before and after the siege's end (September 15, 2016 - March 15, 2017). This process results in more than 50,000 tweets from more than 3,000 accounts.

[2] See Figure A1 in the appendix for an overview of the data collection and labeling pipeline.

Match accounts. In order to understand how Twitter users in Aleppo reacted to the end of the siege, accounts in Aleppo are matched to accounts in other parts of Syria based on the content of tweets posted *prior* to the end of the siege in Aleppo.[3] Cosine similarity is used to measure distance between accounts (Pan and Siegel, 2020). Only accounts that tweet in Arabic are matched in order to avoid the inclusion of accounts by international aid workers or other non-local individuals tweeting from Aleppo. While this subset of users likely misses some local users who regularly tweet in other languages, this approach provides us with a conservative approach to selecting users: restricting by language increases the precision of the sample but reduces recall. This process identifies 335 Arabic-speaking Twitter accounts that were active and sent geotagged tweets from within Aleppo prior to the end of the siege and matches them to twice as many accounts from other parts of Twitter that display the closest similarity to each account. Matching ensures that the two groups are as similar as possible prior to the end of the siege.

Download panel tweets. The entire tweet history of each of the matched accounts is then downloaded, to obtain users' full history of public content, not just the geotagged posts. The *geotagged* sample includes 2,754 Arabic language tweets, while the *panel* data includes 33,455 Arabic language tweets.

This paper analyzes the *geotagged* and *panel* tweets separately in order to investigate the location hypothesis. Since users choose to geotag individual tweets, the decision to geotag is likely to be strategic: accounts may geotag tweets meant for the Assad regime (the victor after a major change in territorial control) while including no location information in tweets that contain neutral or anti-Assad information.

[3] Since individuals may have changed their behavior as the siege entered its final stage, only content that was posted before 12 December 2016, three days before the official end of the siege, is used.

Content Labeling

Topics and sentiment are assigned via fine-tuning of ARBERT, a large-language model trained on Arabic text. Section S3 provides more detail about the model and training data.

Table 1 shows summary statistics for the topics and sentiment for geotagged and panel content from the same set of accounts. The geotagged sample, at an aggregate level, shows the same average for pro-Assad content that the panel data has, but the panel data has more than double the average of anti-Assad content, when compared to the geotagged data. Evidently more anti-Assad content is posted in tweets that do not disclose location information, which is compatible with our expectation that overall, users will be more willing to criticize the government in less visible posts. The summary statistics also indicate that the majority of content posted does not take a stand for or against the Assad regime, which is in line with general findings on social media showing that most content is about entertainment and sports.

Table 1: Summary Statistics for Topics and Sentiment

	N	Mean	Std. Dev.	Min	Max
Geo-tagged Sample					
Anti-Assad	2754	0.011	0.106	0	1
Pro-Assad	2754	0.007	0.085	0	1
Sentiment	2754				
... ['negative']	957	34.7%			
... ['neutral']	691	25.1%			
... ['positive']	1106	40.2%			
Panel Sample					
Anti-Assad	33455	0.029	0.169	0	1
Pro-Assad	33455	0.007	0.082	0	1
Sentiment	33455				
... ['negative']	11788	35.2%			
... ['neutral']	7130	21.3%			
... ['positive']	14537	43.5%			

Table 2 shows four sample tweets and their value for topic, sentiment, and location.

Table 2: Example Tweets

Date	Original	Translated	Topic	Sentiment	Location
2017.12.10	كسر الزوبعهالسوداء بحلب وتحرير مواطنينا المحتجزين سلطان امزجه العصابات الارهابيه تداعت دول العدوان لباريس فعملائها ضعف اجرامهم قسرا	After breaking the black whirlwind in Aleppo and liberating our detained citizens under the authority of the mix of terrorist gangs, the countries of aggression collapsed to Paris, as their agents forcibly doubled their criminality.	Anti-Assad	Negative	Aleppo
2016.12.11	آفوضى ساعة يرتكب فيها من المظالم ما لا يرتكب في استبداد سنين.. ودعاة الفتن امنون في بيوتهم يحرضونْ الحسن البصري #درر	"Chaos is an hour in which grievances are committed that were not committed in the tyranny of years ... and the advocates of sedition are safe in their homes, instigating." Al-Hassan Al-Basri #Durar	Anti-Assad	Negative	Not Aleppo
2016.12.23	استشهد اليوم اخي اسامه معروف رحمه الله دفاعا مدينه دير الزور حاجز البانوراما وصد رفاقه بجيشنا العربي السوري هجوم للدواعش خزلهم الله	Today, my brother Osama Maarouf, may God have mercy on him, was martyred in defense of the city of Deir Ezzor, the Panorama checkpoint, and his comrades in our Syrian Arab army repelled an attack by ISIS, may God let them down.	Pro-Assad	Positive	Aleppo
2017.01.08	تحيهللبطلجبرانعواجي خاطر جبران عواجي بحياته لقتل الارهابي الثاني مسافه ٣ امتار يهاب الاحزمه الناسفه حفظك الل	Tribute to the hero, Gebran Awaji. Gebran Awaji risked his life to kill the second terrorist, 3 meters away. He fears explosive belts. May God protect you.	Pro-Assad	Positive	Not Aleppo

Note: Sample tweets selected by topic and location. English version translated via Google Translate.

Account labeling

To measure changes in activity, accounts are classified as one of 7 types based on reading 20 tweets from each account from before the end of the siege. The account labels include: 'pro-Assad', 'anti-Assad', 'Activist', 'Journalist', 'Media Organization', 'Bot', and 'Other'. The vast majority of accounts are classified as 'Other', indicating that they are neither publicly pro- or anti-Assad before the end of the siege. Other, anti-Assad, and pro-Assad are the most common, in that order, and the analysis presented below focuses on those three account types.[4]. We use the account labels in two steps of the analysis. In our investigation

[4] The sample includes one media organization, two activist accounts, and six journalists

of topics and sentiment, we replicate our main analyses but exclude all content from accounts that were labeled as pro-regime (i.e. pro-Assad) prior to the end of the siege. In a separate analysis of account type activity we model the probability of a tweet being from a pro-regime account, anti-regime account, or an Other account before and after the siege's end. Section S2.2 shows the codebook used.

Model Choice

We compare Twitter content and activity from Aleppo users with matched accounts that were active in Syria but outside of Aleppo during the same time period, which enables us to analyze the impact of the end of the siege in Aleppo on civilians' social media behavior. By comparing content and behavior before and after the end of the siege we study how major shifts in territorial control impact civilians' public messaging on social media in the wake of a changing security situation. We estimate a series of logistic regressions where the binary outcomes refer to the presence or absence of either a topic or a sentiment in a tweet. The models account for the timing of the tweet (pre- or post-siege), where an account is located (in Aleppo or another part of Syria), as well as an interaction term between timing and location. Standard errors are clustered by user. The main outcome of interest is the interaction term: whether post-siege users in Aleppo exhibited a significantly different social media behavior.

Results

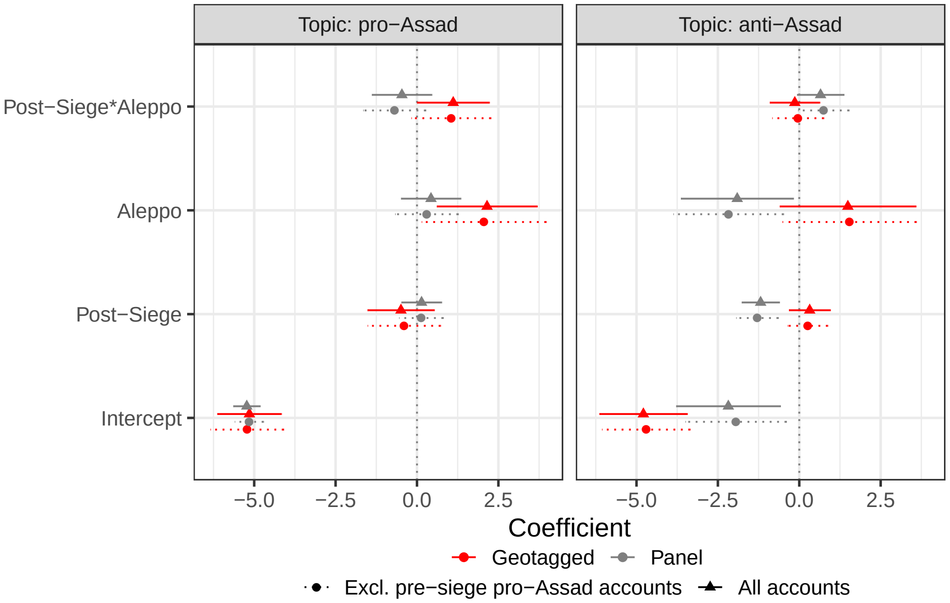
Evidence of content change is first presented and then followed by analysis of activity. Each set of results compares geotagged and panel data to test the location hypothesis.

Changes in Topic and Sentiment

Figure 1 plots the coefficients of logistic regressions that model the difference in content posted in Aleppo in the aftermath of the siege. The left panel shows the probability that a tweet will be pro-Assad, the right one anti-Assad. The red coefficients are based on the geotagged sample of tweets, while the grey shows the panel sample of tweets. The coefficients with confidence intervals that are plotted with solid lines include all accounts in our sample. In the aftermath of the siege, geotagged (in red) tweets in Aleppo are significantly more likely to be pro-Assad than geotagged tweets of similar accounts in other parts of Syria. Geotagged tweets show no change in likelihood of being anti-Assad. This topic change in geotagged tweets in Aleppo in the aftermath of the siege is compatible with our theoretical expectation that users in Aleppo will have an incentive to post more content in support of the 'winning side' following their recapturing of the territory, when compared to similar accounts in other parts of Syria. It is noteworthy that no such pattern is discernible for tweets that do not include location information; in fact the coefficient for the panel models are negative. When looking at the right panel, the data show that the panel data actually shows an *increase* in the probability of content being anti-Assad in the Aleppo in the aftermath of the siege. Recall that both models include the same users, but separately analyze their tweets with (geotagged) and without (panel) location information.

A relevant concern theses findings raise is whether the increased probability in pro-Assad content in Aleppo post-siege is the result of civilians who were pro-government all along suddenly being emboldened by the retaking of the city through the military. We account for this concern by excluding accounts (from Aleppo and other parts of Syria) from our analysis that we coded as being pro-regime prior to the end of the siege. By excluding these accounts we focus on civilians who were previously either anti-Assad or neutral. The coefficients with dotted confidence intervals replicate the geotagged and panel results just presented, for both pro- and anti-Assad topics, but without accounts that were previously publicly pro-regime. The coefficients are very similar to those including longtime supporters of Assad, indicating that the results are not driven by a few longtime supporters of the regime who started mass posting about their loyalty towards the regime once the siege ended.

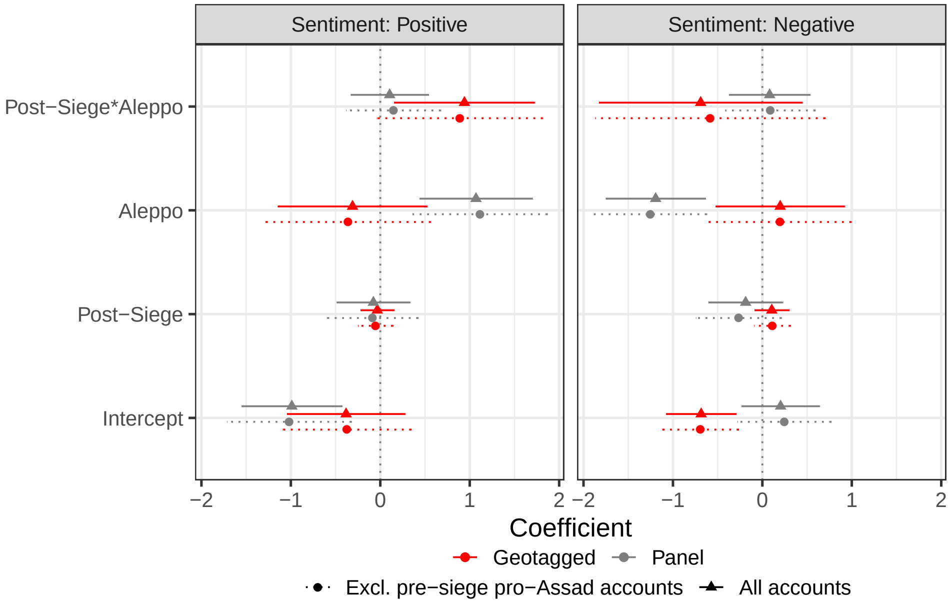
Figure 1: Changes in Topics in Aleppo



Note: Results for changes in topic for tweets from users in Aleppo compared to matched users outside of Aleppo. Post-Siege*Aleppo is of main interest; the red point estimate is for the geotagged data, the grey is for panel data. The estimates with dotted confidence intervals replicate the main results, but exclude pre-siege end pro-regime accounts. The difference in inference when comparing the geotagged and panel results is suggestive of signalling behavior. See Tables A6 and A15 for regression tables.

Figure 2 presents the results for changes in positive and negative sentiment. In the geotagged sample, content posted post-siege from Aleppo has a higher probability of being positive than content from comparable accounts in other parts of Syria. The sign for negative sentiment is negative, though it is not statistically significant. Similar to the pro-Assad topic results, this pattern does not exist for the panel tweets: in the panel data, there is no change in expressed sentiment. The sentiment treatment coefficients are especially informative when compared to the coefficients of Aleppo. In the geotagged sample, tweets from Aleppo are just as likely to be positive or negative as those from the matched users outside of Aleppo while the panel tweets are more likely to contain positive sentiment and less likely to contain negative sentiment than tweets from elsewhere.

Figure 2: Changes in Sentiment in Aleppo



Note: Results for changes in sentiment for tweets from users in Aleppo, compared to matched users outside of Aleppo. Post-Siege*Aleppo is of main interest; the red point estimate is for the geotagged data, the grey is for panel data. The estimates with dotted confidence intervals replicate the main results, but exclude pre-siege end pro-regime accounts. The difference in inference when comparing the geotagged and panel results is suggestive of signalling behavior. See Tables A7 and A16 for regression tables.

As with the topic analyses, we replicate our sentiment models with samples that exclude account that were pro-regime prior to the end of the siege. The results further suggest that the changes in behavior are not driven by a small selection of legacy supporters of the government; Aleppo-based accounts that were previously neutral or even anti-Assad are more likely to post positive content after the siege ends.

The results for topic and sentiment are highly suggestive of strategic behavior. The clearest evidence comes when comparing the geotagged topic results to the panel: tweets are tagged to Aleppo when they are pro-Assad, but when they are not geotagged they are less likely to be pro-Assad. In fact, the panel tweets (non-geotagged) reveal an increase in anti-Assad topics. The same is true for sentiment. In other words, a combatant would receive one signal about Aleppo residents if they searched for tweets from Aleppo but an opposite one if they read non-geotagged tweets from accounts from Aleppo.

Yet overall, panel tweets are more likely to be positive in Aleppo than in other parts of the country. Negative sentiment shows no significant change in probability in content from post-siege Aleppo.

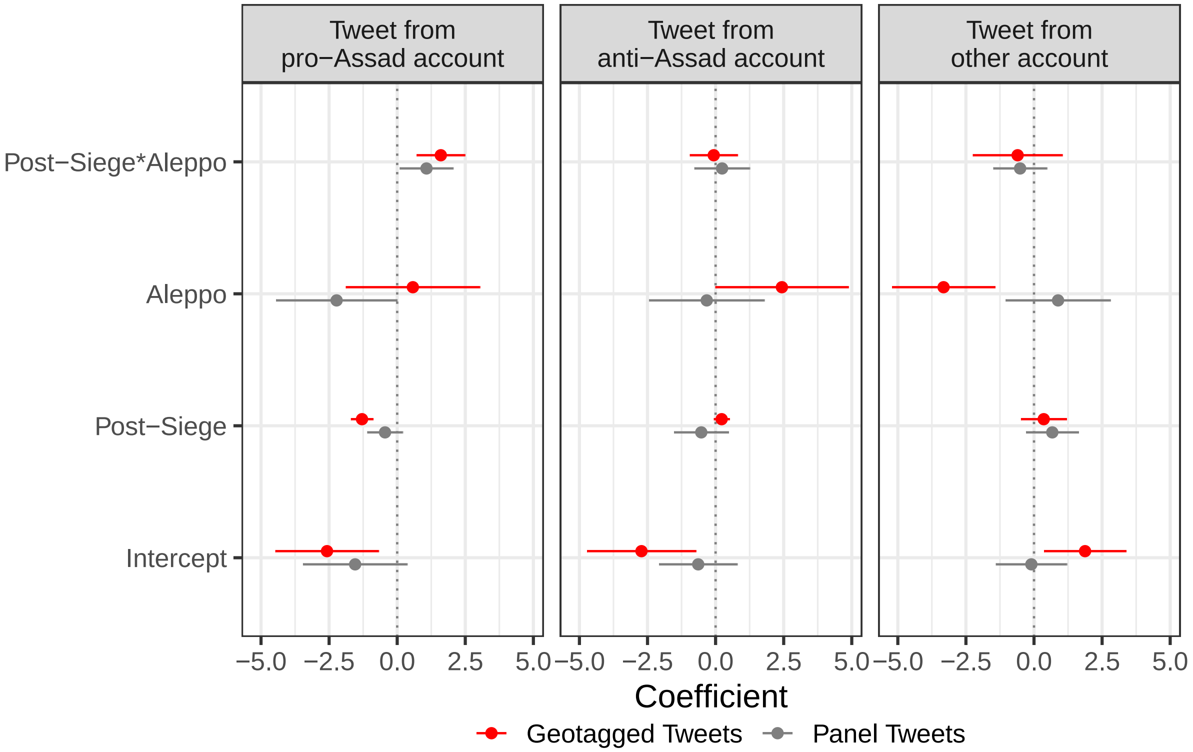
Changes in content from account types

The account labels allow us to investigate whether the account types who are active before and after the siege change in Aleppo and elsewhere in the country. Figure 3 shows the probability of a post being sent from either a pro-Assad (i.e. pro-regime) account, an anti-Assad account, or Other accounts. As before, we distinguish between content that includes location information and content that does not. In the aftermath of the siege, tweets in Aleppo are more likely to have been sent from an account that was labeled as being pro-regime prior to the end of the siege, when compared to similar accounts in other parts of Syria. Interestingly, the pattern holds for both geotagged and panel tweets, suggesting that longtime supporters of the regime were more likely to have authored a tweet in Aleppo following the retaking of the city by regime forces. This finding is compatible with our expectation that regime supporters in Aleppo likely felt emboldened by the renewed regime presence in the city following the ceasefire. It is also compatible with the expectation that regime supporters would deem it more necessary to signal their loyalty with the regime so as to ensure their security.

The middle and left panel of Figure 3 show that there is no change in the probability of a tweet being authored by anti-Assad or other accounts in Aleppo following the end of the siege. As with the findings on pro-Assad accounts, these patterns are true for both geotagged and panel data. This suggests that activity changes do not differ between content with and without location

Overall, these results show that tweets are more likely to have been authored by regime supporters in Aleppo once the regime takes over again, but they also show that anti-Assad and other types of accounts are not less likely to post. Taken together with the results on tweet topic and sentiment, in particular the analyses that exclude pro-Assad accounts, the

Figure 3: Change in probability of content being from different account types



Note: Results for changes in probability of a tweet being from either pro-Assad, anti-Assad, or other accounts, when comparing user posting from Aleppo to matched users outside of Aleppo. Post-Siege*Aleppo is of main interest; the red point estimate is for the geotagged data, the grey is for the panel data. See Table A8 for regression results.

findings are indicative of the strategic considerations we hypothesize about in this paper. They suggest that in Aleppo, following the siege, more pro-Assad and more positive content was posted in tweets that included location information, and that this pattern can also be found in accounts that were *not* overtly pro-regime prior to the end of the siege. Panel tweets, those that do not include location info, follow very different patterns. Here we see no increase in pro-Assad content, and instead find an increase in the probability of anti-Assad topics. When looking at the composition of tweets from different types of users before and after the siege, we find an increase in the likelihood of content stemming from longtime regime supporters, yet we do not find a decrease in other types of accounts. Taken together the findings indicate that pro-regime supporters likely felt emboldened and incentivized to post content in support of the regime following their recapturing of Aleppo, and that previously neutral or anti-Assad accounts were motivated to post more pro-regime and more positive content following the siege's end.

Additional tests

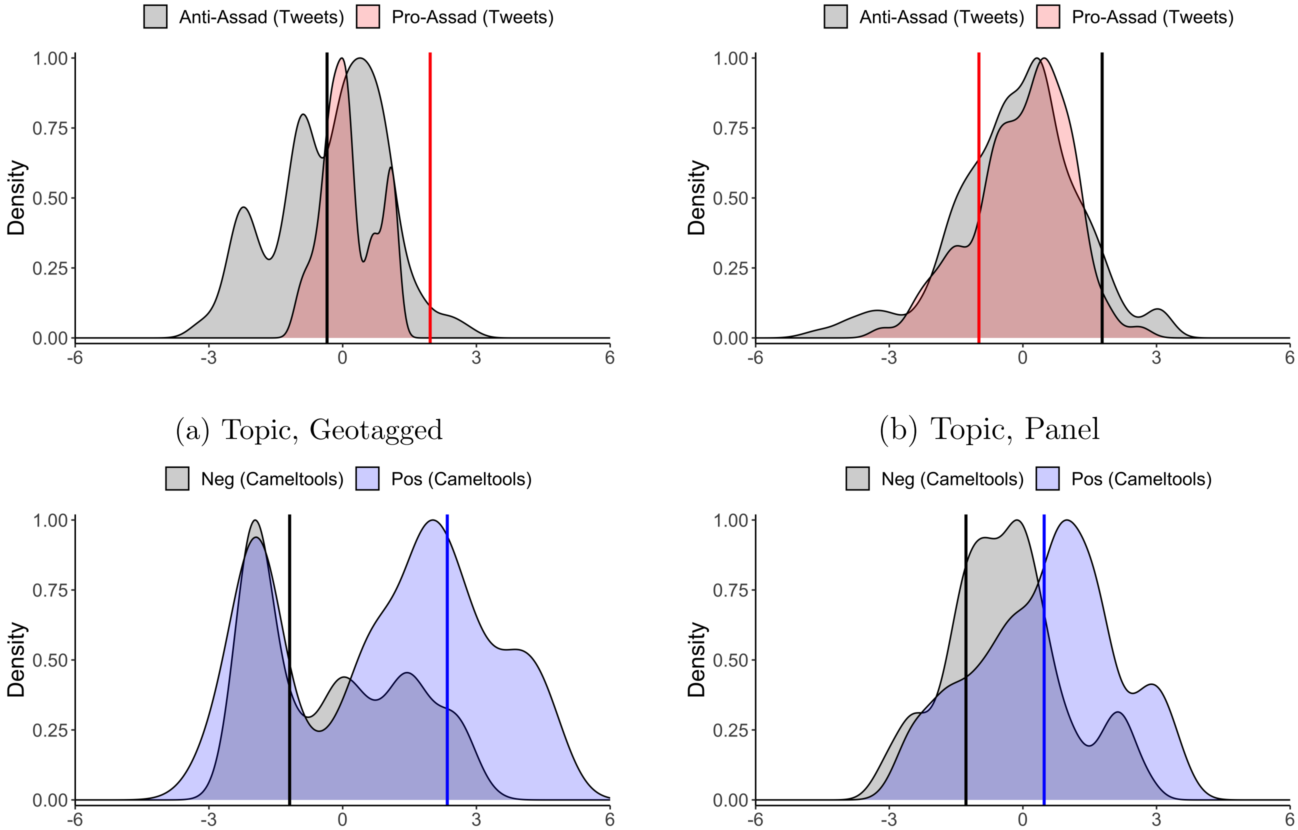
We run a series of additional tests to ensure that our findings are not dependent on a specific cut-off date for the siege's end, a specific comparison sample of non-Aleppo based accounts, or on changes that also occurred in other parts of Syria. In addition, we also investigate whether account in our sample are inauthentic, and find no evidence for this concern.

Though the siege's end was formally announced on December 15, 2016, there was growing awareness in the proceeding days that rebel forces were losing, so the official end of the siege was not an unforeseen event that would trigger an immediate change from one day to the next. In our main analysis we use December 15 as the best approximation of the end, but want to ensure that our results are not driven by choosing this day as the cut-off. To probe the sensitivity of our results with respect to the exact date of the end of the siege we replicate our results for the geotagged and panel data but remove all tweets posted in the two days leading up the end of the siege, the day of, and the two days following the end of the siege (13-17. December 2016). Figures A2 and A3 show that when we remove tweets that were published immediately before, during, and after the end of the siege our results stay largely the same, suggesting that they are not sensitive to a specific cut-off date.

To further analyze the robustness of our results we replicate our results with time-shifted placebo dates for the end of the siege, to see what the probability would be to find similar results to our findings if we were to randomly pick another date and compare content in Aleppo to the rest of Syria. We randomly choose any date between June 1 2015 and November 13 2016 (which 31 days before the siege's end), as the cut-off, rerun our analyses, and record the resulting test statistic for the `post-siege*Aleppo` parameter. This process is repeated 1,000 times per dependent variable (pro-Assad and anti-Assad topics, positive and negative sentiment), and that distribution is compared to our parameter for when the cut-off is December 15, 2016. Figure 4 show the density of the student's t-statistics for the placebo models, while the vertical lines correspond to the value for the actual end of the siege.

For the topic results (top row), the test statistic for $Post - Siege * Aleppo$ for Pro-Assad tweets is in the 89.6th percentile in the geotagged sample (left figure) and 31.6th in the panel (right figure); for Anti-Assad, 57.9th and 74.3rd respectively. That the treatment test statistic is not unlikely for the sentiment results makes sense given their small values in the original model.

Figure 4: Placebo Treatment Dates



(a) Topic, Geotagged

(b) Topic, Panel

(c) Sentiment, Geotagged

(d) Sentiment, Panel

Note: The x-axis is the student's t statistic. Each histogram represents 1,000 placebo dates, with the vertical lines corresponding to the actual student's t using December 15, 2016 as the treatment date.

We conduct a second type of placebo test by exchanging the accounts from Aleppo with comparable accounts from Hama during the same period under investigation, and compare those Hama-based users to accounts in the rest of Syria. The rationale behind conducting this test is to check whether the changes detected in Aleppo also occurred in anther part of Syria in the aftermath of the siege. Tables A11 and A12 show that no differences in topic or sentiment are detected in Hama in the aftermath of the siege in Aleppo.

The comparison group of users from the rest of Syria who are similar to Aleppo-based civilians may include users from quite disparate regions of Syria, with some being geographically closer to Aleppo than others, thereby introducing potential spillover effects. We construct a geographically concentrated comparison group comprising of users based only in Damascus. Tables A9 and A10 replicate the results of the main analysis, but compare the Aleppo accounts to Damascus-based accounts. The outcome of these analyses are comparable to the main results, suggesting that the changes detected in Aleppo are not dependent on the comparison group chosen in the main analysis.

Users could also act strategically by changing their screen name. When a user changes their screen name but keeps an account active, Twitter does not change the user identification number. This user ID is delivered with every tweet. Since this paper's geotagged data was collected in real time, we can therefore observe how many users change their screen name and when. Of the seven users who change their screen name within 31 days before or after the siege's end, three do so after and four before. Five of the seven changes occur between 12.06.2016 and 12.24.2016. These seven users who change their screen name tweet 398 times within the two-month window. Of the fifteen from Aleppo, 46.67% are from after the siege, while only 41.51% of tweets from outside Aleppo are from after the siege's end. Of the 166 from after the siege's end, 4.21% are from Aleppo versus 3.44% that are not. Neither result is statistically significant. A linear probability model suggests that these accounts are 50% more likely to tweet after the siege's end, though there is no difference for accounts based on their location. Overall, this evidence is weakly suggestive of strategic behavior. The timing of screen name changes is suggestive of strategic behavior, but a sample of seven is too small to conduct statistical tests.

A further concern is that accounts studied in this paper are inauthentic. While previous work has found that very few bot accounts geotag their tweets (Driscoll and Steinert-Threlkeld, 2020), inauthenticity also comes from state-linked information operations. This concern is especially pertinent in Syria given the Russian government's heavy offline involvement in the conflict and its widespread use of online information operations to advance foreign policy goals. In the context of the siege of Aleppo, such a campaign could manifest as creating the impression of widespread support of the Assad government and positive sentiment after the siege of Aleppo.

To understand if Russian state-backed influence operations affect this paper's findings, we turn to 13,221,152 tweets Twitter has identified as part of Russian state-linked information operation campaigns. These tweets are from 10 datasets representing 10 takedowns that Twitter released. As verified researchers, we were able to see the unhashed version, meaning we could search tweets for specific users Twitter has identified as accounts linked to Russian information operation campaigns. In the 13.2 million tweets, 73,591 from 2,103 accounts contain one of the words "Syria", "Damascaus", Aleppo", or "siege" in English, Arabic, or Russia. None of those tweets, and therefore none of the accounts, are found in this paper's data.

Conclusion

The findings presented in this paper provide evidence for significant changes in content, sentiment, and activity that are compatible with theoretical expectations of strategic signalling. In Aleppo, civilians on social media were more likely to share pro-Assad content in the aftermath of the siege in posts where they also included exact location information. In posts that did not include location information there is an increase in anti-Assad content. The changes in content and the diverging findings depending on in- or excluding location details provides evidence for the fact that the same group of users reacts strategically to the new security situation under regime control, while being aware of the costliness of sharing exact location information. The data also reveal an increase in the likelihood of users sharing positive content in Aleppo once the regime won control in mid-December. Importantly, the results do not change when excluding users who were publicly pro-regime prior to the end of the siege, and the results survive a series of placebo tests.

Changes in content on social media also have second order effects on those using online platforms to monitor and make impactful decisions from afar. When content and activity on social media is strategically modified in the context of shifts in conflict dynamics, then information collection based on open source information may misrepresent the opinion and needs of the local population. Misrepresentations of the context in turn may lead to suboptimal policy responses by humanitarian actors as well as misrepresentations in international news coverage. Where the security situation prohibits immediate access to conflict locations, journalists, aid workers from local and international organizations, and human rights documentation groups tend to complement their situation reports with social media content (Dubberley, Koenig and Murray, 2020, Singer and Brooking, 2018). International observers who followed social media that was tagged as originating from Aleppo in the aftermath of the siege may have concluded that the local population was in fact jubilant of the retaking of the city by pro-regime forces.

Where social media is used to cut through the fog of war, understanding the ways in which the conflict itself may form when and why civilians post online is more important than ever. This paper's analysis shows that the increase in positive and pro-Assad content was only present in intentionally geotagged posts, suggesting that civilians modified their content to improve their personal security. In content without location information there is no increase in pro-Assad content; instead, anti-Assad posts increase.

More than a decade after the first protests in the Middle East and North Africa were recorded and broadcast on social media, digital communication has become an everyday reality of modern day contentious politics. Much progress has been made in understanding the ways in which these no longer new tools are used for protest mobilization and coordination, yet beyond initial conflict onset, the everyday use of social media by civilians caught in the midst of war remains understudied. This paper points to a future direction of research:

studies should theorize about and test civilian behavior on social media during conflict.

References

Abdul-Mageed, Muhammad, Abdelrahim A Elmadany and El Moatez Bill. 2021. ARBERT & MARBERT: Deep Bidirectional Transformers for Arabic. In *59th Annual Meeting of the Association for Computational Linguistics*. pp. 7088–7105.

Baly, Ramy, Alaa Khaddaj, Hazem Hajj, Wassim El-Hajj and Khaled Bashir Shaban. 2019. "ArSentD-LEV: A Multi-Topic Corpus for Target-based Sentiment Analysis in Arabic Levantine Tweets.".

URL: *http://arxiv.org/abs/1906.01830*

Bassam, Laila, Angus McDowall and Stephanie Nebehay. 2016. "Battle of Aleppo ends after years of bloodshed with rebel withdrawal.".

URL: *https://www.reuters.com/article/us-mideast-crisis-syria/battle-of-aleppo-ends-after-years-of-bloodshed-with-rebel-withdrawal-idUSKBN1420H5*

Berman, Eli, Jacob N. Shapiro and Joseph H Felter. 2011. "Can Hearts and Minds Be Bought? The Economics of Counterinsurgency in Iraq." *Journal of Political Economy* 119(4):766–819.

Bisbee, James and Jennifer M. Larson. 2017. "Testing social science network theories with online network data: An evaluation of external validity." *American Political Science Review* 111(3):502–521.

Böttcher, Annabelle. 2017. "News Analysis Humanitarian Aid and the Battle of Aleppo." 105(1):1–6.

Brownlee, Billie Jeanne. 2018. Mediating the Syrian Revolt: How new media technologies change the development of social movements and conflicts. In *The Syrian Uprising: Domestic Factors and Early Trajectory*, ed. Raymond Hinnebusch and Omar Imady. Routledge pp. 188–206.

Chang, Keng Chi, William R. Hobbs, Margaret E. Roberts and Zachary C. Steinert-Threlkeld. 2022. "COVID-19 increased censorship circumvention and access to sensitive topics in China." *Proceedings of the National Academy of Sciences of the United States of America* 119(4):e2102818119.

Cullinane, Susannah. 2018. "Egyptian activist detained after social media video post criticizes government.".

URL: *https://edition.cnn.com/2018/05/11/middleeast/egypt-amal-fathy-activist-detained/index.html*

Driscoll, Jesse and Zachary C. Steinert-Threlkeld. 2020. "Social media and Russian territorial irredentism: some facts and a conjecture conjecture." *Post-Soviet Affairs* 36(2):101–121.

Dubberley, Sam, Alexa Koenig and Daragh Murray, eds. 2020. *Digital witness: using open source information for human rights investigation, documentation, and accountability.* Oxford University Press.

Feldstein, Steven. 2021. *The Rise of Digital Repression: How Technology is Reshaping Power, Politics, and Resistance.* Oxford University Press.

Freedom House. 2015. "Syria." *Freedom on the Net 2015* .

URL: *https://freedomhouse.org/sites/default/files/resources/FOTN 2015_Syria.pdf*

Freelon, D, M Lynch and S Aday. 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.

Gambetta, Diego. 2009. Signaling. In *The Oxford Handbook of Analytical Sociology.* p. chapter 8.

Glaser, April and Saphora Smith. 2021. "Taliban violence drives Afghans to wipe social media profiles.".

URL: *https://www.nbcnews.com/tech/tech-news/taliban-violence-drives-afghans-wipe-social-media-profiles-rcna1731*

Gohdes, Anita R. 2020. "Repression Technology: Internet Accessibility and State Violence." *American Journal of Political Science* 64(3):488–503.

Hashem, Mohamed. 2015. "Q&A: In Syria the 'internet has become a weapon' of war." *Al-Jazeera* .

URL: *http://www.aljazeera.com/indepth/features/2015/06/qa-syria-internet-weapon-war-150619215453906.html*

Havel, Václav and Paul Wilson. 1986. "The power of the powerless." *International Journal of Politics* 15(4):23–96.

IFEX. 2018. "Bahrain vows greater crackdown on online criticism amid new arrests - IFEX.".

URL: *https://ifex.org/bahrain-vows-greater-crackdown-on-online-criticism-amid-new-arrests/*

International Telecommunication Union (ITU). 2022. "Individuals using the Internet (% of population).".

URL: *https://data.worldbank.org/indicator/IT.NET.USER.ZS*

Jones, Benjamin T and Eleonora Mattiacci. 2019. "A Manifesto, in 140 Characters or Fewer: Social Media as a Tool of Rebel Diplomacy." *British Journal of Political Science* 49(2):739–761.

Kalyvas, Stathis. 2006. *The Logic of Violence in Civil War*. New York: Cambridge Univeristy Press.

Kalyvas, Stathis N. 2008. Promises and pitfalls of an emerging research program: the microdynamics of civil war. In *Order, Conflict, and Violence*, ed. Stathis N. Kalyvas, Ian Shapiro and Tarek Masoud. Cambridge University Press pp. 1–436.

Kalyvas, Stathis N. and Matthew Adam Kocher. 2007. "How Free is Free Riding in Civil Wars?: Violence, Insurgency, and the Collective Action Problem." *World Politics* 59(2):177–216.

Kuran, Timur. 1991. "Now Out of Never: The Element of Surprise in the East European Revolution of 1989." *World Politics* 44(1):7–48.

Little, Andrew T. 2015. "Communication Technology and Protest." *Journal of Politics* 78(1):152–166.

Lukito, Josephine. 2019. "Coordinating a Multi-Platform Disinformation Campaign: Internet Research Agency Activity on Three U.S Social Media Platforms, 2015 to 2017." *Political Communication* pp. 1–18.

Lyall, Jason. 2009. "Does Indiscriminate Violence Incite Insurgent Attacks?: Evidence from Chechnya." *Journal of Conflict Resolution* 53(3):331–362.

Lyall, Jason, Graeme Blair and Kosuke Imai. 2013. "Explaining Support for Combatants during Wartime: A Survey Experiment in Afghanistan." *American Political Science Review* 107(04):679–705.

Lynch, Marc, Deen Freelon and Sean Aday. 2014. "Blogs and Bullets III: {S}yria's Social Mediated War." *United States Institute of Peace, Peaceworks* 91.

Metzger, Megan, Jonathan Nagler and Joshua a. Tucker. 2015. "Tweeting Identity? Ukrainian, Russian, and #Euromaidan." *Journal of Comparative Economics* 44(1):16–40.

Morozov, Evgeny. 2012. *The Net Delusion: The Dark Side of Internet Freedom.* New York City: PublicAffairs.

Moss, Dana M. 2018. “The ties that bind: Internet communication technologies, networked authoritarianism, and voice’ in the Syrian diaspora.” *Globalizations* 15(2):265–282.

Munger, Kevin, Richard Bonneau, Jonathan Nagler and Joshua A Tucker. 2019. “Elites Tweet to Get Feet Off the Streets: Measuring Regime Social Media Strategies During Protest.” *Political Science Research and Methods* 7(4):815–834.

Pan, Jennifer and Alexandra Siegel. 2020. “How Saudi Crackdowns Fail to Silence Online Dissent.” *American Political Science Review* 114(1):109–125.

Pfaff, Steven. 1996. “Collective Identity and Informal Groups in Revolutionary Mobilization: East Germany in 1989.” *Social Forces* 75(1):91–117.

Popkin, Samuel L. 1979. *The Rational Peasant: The Political Economy of Rural Society in Vietnam*. Berkeley: University of California Press.

Qin, Bei, David Strömberg and Yanhui Wu. 2017. “Why Does China Allow Freer Social Media? Protests versus Surveillance and Propaganda.” *Journal of Economic Perspectives* 31(1):117–140.

Schon, Justin. 2021. “How narratives and evidence influence rumor belief in conflict zones: Evidence from Syria.” *Perspectives on Politics* 19(2):539–552.

Schubiger, Livia Isabella. 2021. “State Violence and Wartime Civilian Agency: Evidence from Peru.” *Journal of Politics* 83(4):1383—-1398.

Shehabat, Ahmad. 2012. “The social media cyber-war: the unfolding events in the Syrian revolution 2011.” *Global Media Journal: Australian Edition* 6(2).

Singer, Peter Warren and Emerson T. Brooking. 2018. *LikeWar: The weaponization of social media.* Eamon Dolan Books.

Spence, Michael. 1973. “Job Market Signaling.” *The Quarterly Journal of Economics* 87(3):355–374.

Stukal, Denis, Sergey Sanovich, Richard Bonneau and Joshua A. Tucker. 2022. "Why Botter: How Pro-Government Bots Fight Opposition in Russia." *American Political Science Review* pp. 1–15.

Tibken, Shara. 2016. "How a Facebook page sent one Syrian dissenter to prison." *CNET* .
URL: *https://www.cnet.com/news/how-a-facebook-page-sent-one-syrian-dissenter-to-prison/*

Todman, Will. 2017. "Isolating dissent, punishing the masses: siege warfare as counter-insurgency." pp. 1–32.

U.S. Army / Marine Corps. 2007. *The U.S. Army/Marine Corps counterinsurgency field manual : U.S. Army field manual no. 3-24 : Marine Corps warfighting publication no. 3-33.5.* University of Chicago Press.

Valentino, Benjamin A, Paul Huth and Dylan Balch-Lindsay. 2004. "Draining the Sea: Mass Killing and Guerrilla Warfare." *International Organization* 58(02):375–407.

Walk, Erin, Elizabeth Parker-Magyar, Kiran Garimella, Ahmet Akbiyik and Fotini Christia. 2022. "Social Media Narratives across Platforms in Conflict: Evidence from Syria.".

Wedeen, Lisa. 1999. *Ambiguities of domination : politics, rhetoric, and symbols in contemporary Syria.*

Wedeen, Lisa. 2019. *Authoritarian Apprehensions: Ideology, Judgment, and Mourning in Syria.* University of Chicago Press.

Wong, Chun Han. 2021. "China Is Now Sending Twitter Users to Prison for Posts Most Chinese Can't See - WSJ.".
URL: *https://www.wsj.com/articles/china-is-now-sending-twitter-users-to-prison-for-posts-most-chinese-cant-see-11611932917*

Xu, Xu. 2021. "To Repress or to Co-opt? Authoritarian Control in the Age of Digital Surveillance." *American Journal of Political Science* 65(2):309–325.

Zeitzoff, Thomas. 2017. "How Social Media Is Changing Conflict." *Journal of Conflict Resolution* 61(9):1970–1991.

Appendix for

“Civilian Behavior on Social Media During Civil War”

Contents

S1 Data Collection and Labeling Pipeline

Figure A1 provides an overview of the data collection and and labeling pipeline. The four stacked boxes on the left summarize data collection process. The three middle columns describe the labeling of content and accounts. The final (right side) of the pipeline shows the analysis of the content and accounts that leads us to the results.

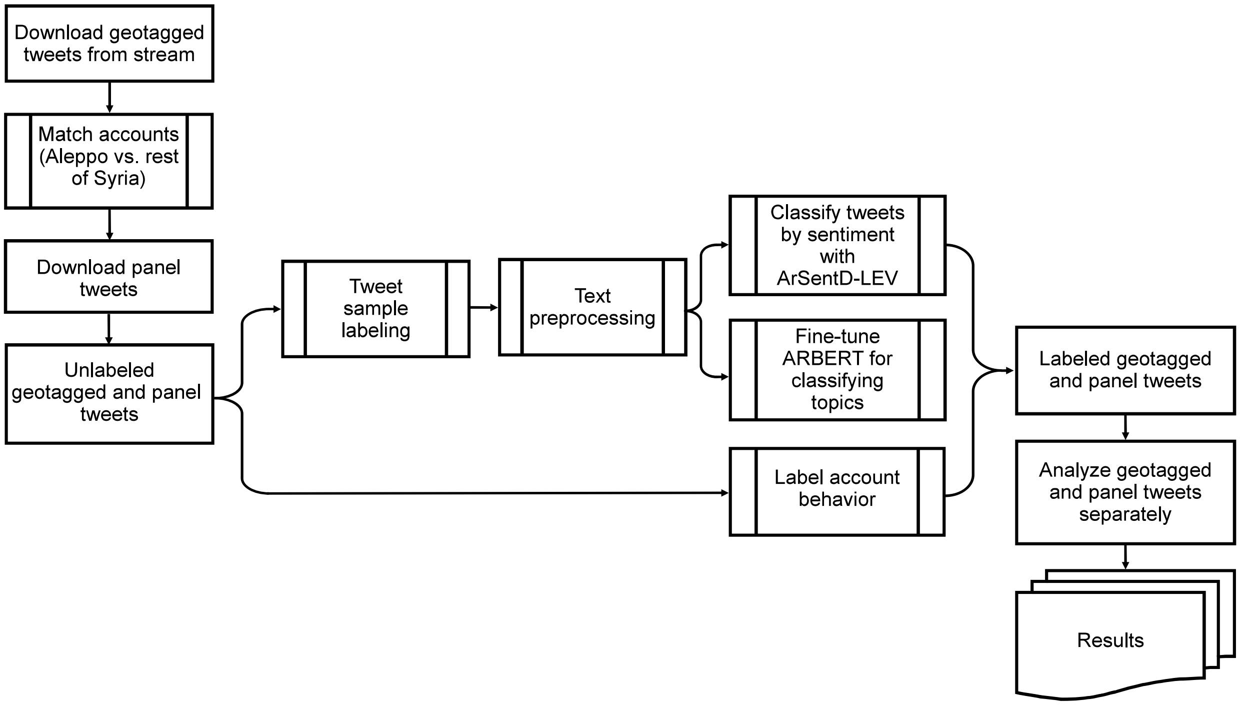

Download geotagged tweets from stream
Match accounts (Aleppo vs. rest of Syria)
Download panel tweets
Unlabeled geotagged and panel tweets
Tweet sample labeling
Text preprocessing
Classify tweets by sentiment with ArSentD-LEV
Fine-tune ARBERT for classifying topics
Label account behavior
Labeled geotagged and panel tweets
Analyze geotagged and panel tweets separately
Results

Figure A1: Data Collection and Labeling Pipeline

S2 Codebooks

S2.1 Codebook for Tweet Content

The following codebook was presented to the team of research assistants who labeled training data.

This document details the coding process for the Syrian Twitter project.

Fill in the cells of a spreadsheet. Each row is a tweet, and each column starting with F represents an emotion or topic. (Notice the first row of the Sheet now indicates which columns are emotions and which are tweets.) If a tweet contains an emotion or topic, enter a 1 in the corresponding cell; leave the cell blank otherwise. Emotions and topics are not mutually exclusive: a tweet can express fear and disgust or pride and surprise, for example. If you are unsure of an emotion or topic, give it a 1 but also assign a 1 to the other column.

The categories are:

- Topic
  - aleppo_notWar: 1 if the tweet is about Aleppo but not the civil war, blank otherwise.
  - aleppo_war: 1 if the tweet is about Aleppo and the civil war, blank otherwise.
  - Assad_anti: 1 if the tweet is against the Assad regime, blank otherwise. A tweet that supports the Free Syrian Army, ISIS, or any other group fighting the Assad regime counts as anti-Assad.
  - Assad_pro: 1 if the tweet is supportive of the Assad regime, blank otherwise.
  - humanitarian_group: 1 if the tweet is about a humanitarian group such as Doctors Without Borders, blank otherwise. For example, if the tweet mentions Doctors Without Borders, give it a 1.
  - war_foreignNotRussia: 1 if the tweet is about a countrys involvement in the civil war that is not Russia, blank otherwise. For example, a tweet mentioning Irans involvement in the civil war is a 1.
  - war_foreignRussia: 1 if the tweet is about Russias involvement in the civil war, blank otherwise. If the tweet is about Russia and Aleppo, give it a 1 for aleppo_war as well.
  - war_syriaNotAleppo: 1 if the tweet is about the civil war but not about Aleppo, blank otherwise.
  - entertainment: 1 if the tweet is about music, pop culture, sports, books, movies, etc.; blank otherwise.
  - foreign_news: 1 if the tweet is about a foreign country, blank otherwise. For example, a tweet mentioning the Houthi rebels in Yemen is a 1; so is a tweet not mentioning a conflict.
  - joke: 1 if the tweet contains a joke or is humorous, blank otherwise.

  - poetry: 1 if the tweet is about or contains poetry, blank otherwise. Poetry here is broadly construed. For example, an inspirational statement written in formal Arabic should receive a 1 for poetry.
  - religion: 1 if the tweet is about religion, blank otherwise.
  - other: 1 if the tweet is not about any of these topics or contains an emotion. If you have assigned a 1 to an emotion or tweet but are unsure of that choice, assign a 1 to other as well.

- Emotion
  - anger: 1 if the tweet expresses anger, blank otherwise.
  - annoyed: 1 if the tweet expresses annoyance, blank otherwise.
  - disgust: 1 if the tweet expresses disgust, blank otherwise.
  - fear: 1 if the tweet expresses fear, blank otherwise.
  - happy_or_joy: 1 if the tweet expresses happiness or joy, blank otherwise.
  - hate: 1 if the tweet expresses happiness or joy, blank otherwise.
  - hope: 1 if the tweet expresses hope, blank otherwise. Hope content must contain an orientation to the future.
  - love: 1 i the tweet expresses love, blank otherwise.
  - pride: 1 if the tweet expresses pride, blank otherwise.
  - respect: 1 if the tweet expresses respect, blank otherwise.
  - sadness: 1 if the tweet expresses sadness, blank otherwise.
  - shame: 1 if the tweet expresses shame, blank otherwise.
  - surprise: 1 if the tweet expresses surprise, blank otherwise.

- Notes
  - notes: enter any notes here you have about the tweet. A note could be about an ambiguity in the tweet, a question to raise at a meeting, or a topic to which you would like to call attention.

If a tweet contains no emotion, then do not enter a 1 in any cell. If a tweet does not contain one of the topics, enter a 1 for other.

As you code, let me know if you see emotions or topics in tweets that are not in the Sheet.

S2.2 Codebook for Account Type

The following codebook was used to label account types.

The Sheet accounts_to_code contains 20 tweets from accounts in Syria. Using those tweets (text column), and the user_name, user_username, and user_description columns, code each account into one of the following accountType categories:

- pro-Assad
- anti-Assad
- Activist
- Journalist
- Media organization
- Bot
- Other

First try to determine the account type based on the description. If the description is not clear, then read the tweets. (For the approximately 50 accounts that had fewer than 20 tweets in the original data, there will be repeated tweets; ignore them.) If the account type is not clear from those tweets, go to www.twitter.com, search for the user, read their description, and attempt to code the account from that description. If the account type is still unclear, make it the Other type. If you visit Twitter, make the checkedTwitter column Yes.

As soon as you are comfortable assigning an account type, do not read more tweets from that account. If you are comfortable with the account type just from the user_description column, choose an accountType for the first tweet only and leave the remaining tweets blank. Each account should have at least one tweet with an accountType and checkedTwitter value; to accelerate the coding process, not every tweet per account needs those values.

S3 BERT and ARBERT

BERT is a leading natural language processing neural network model developed at Google and released to the public at the end of 2019 (Devlin et al., 2019). Trained on the 800 million words of the BookCorpus (Zhu et al., 2015) and 2.5 billion in English Wikipedia, BERT's primary advance is to provide a general purpose language model that can then be fine-tuned (or customized) to a specific task. Here, the customization consists of providing labeled training data, tweets, to generate a similar but new model.

ARBERT customizes the BERT architecture using a 61 gigabyte (6.5 billion tokens) collection of Modern Standard Arabic text: 1,800 Arabic books from the public Arabic bookstore Hindawi; 5 million news articles data set from 10 major news sources covering eight Arab countries from (Abu El-khair, 2016); the Arabic Gigaword 5th Edition from the Linguistic Data Consortium (LDC), a comprehensive archive of news wire text from multiple Arabic news sources; OSCAR, the Modern Standard Arabic and Egyptian Arabic portion of the Open Super-large Crawled Almanach coRpus (Ortiz Suárez, Sagot and Romary, 2019); the Open Source International Arabic News Corpus (OSIAN) (Zeroual et al., 2019); and a December 2019 download of Arabic Wikipedia. The final model contains 163 million parameters.

S3.1 Data Cleaning and Processing

The tweets are cleaned before being analyzed by the fine-tuned ARBERT. Tweets can contain special characters such as emoticons and Unicode characters that occurs as a result of a miss-translation of the original characters in Python. They may also contain non-standard spellings and meta data such as URLs, mentions, and hashtags, all of which can add noise to the text and hinder the process of capturing meaningful information from the words present in it. The following steps were therefore applied to the entire corpus of tweets.

The tweets are set to lowercase, Unicode characters, trailing white spaces and consecutive characters occurring in a word are removed. URLS and mentions are removed, as is the hashtag symbol. The 365 most frequent emojis were translated into Arabic using translations provided by one of the coders; they were kept in the training tweets given to the human coders. Any punctuation between characters in a word is also removed.

In addition to these relatively simple steps are considerations unique to processing Arabic. Orthographic ambiguity means the form of characters and spelling of words can vary depending on their context. Morphological richness means the same words can have thousands of different forms. Dialectic variation and orthographic inconsistency mean the same word can be written different ways. This variation contributes to data sparsity, making it difficult for a model to captured semantic relationship between the words. To reduce the noise and sparsity of our data, we make use of the Camel Tools package to perform Unicode normalization, dediacritization, removing characters with no phonetic value, and orthographic normalization (Obeid et al., 2020).

S3.2 Content Labeling

Each tweet is assigned topic and sentiment labels. Each tweet can contain multiple topics, which come from a fine-tuned deep learning text model created for this project. Only one sentiment label is assigned for each tweet; these come from a fine-tuned deep learning model for Levantine Arabic (Baly et al., 2019, Antoun, Baly and Hajj, 2020).

To label the topics, the ARBERT language model (Abdul-Mageed, Elmadany and Bill, 2021), a neural network natural language processing model that is itself built upon a Bidirectional Encoder Representations from Transformers (BERT) model, the base neural network language model upon which most new language models build, is fine-tuned. ARBERT customizes the BERT architecture using a 61 gigabyte (6.5 billion tokens) collection of Modern Standard Arabic text.[1] We fine-tune ARBERT using a sample of 6,000 tweets from the collection used in this paper. Tweets from Syria from September 1, 2016 through March 31, 2017 were randomly chosen because they are more likely to contain relevant content than the broader set of text ARBERT uses. Customizing an already trained model is common with large language models (LLMs) and achieves high levels of accuracy with much smaller training datasets than the ones used to train BERT and its variants (Eljundi et al., 2019).

Three coders, native Arabic or second-generation Arabic speakers, were instructed to label tweets as containing one of fourteen topics and one of 13 emotions; see Section S2.1 for the codebook. The topics are the war in Aleppo, Aleppo but not the war, pro-Assad, anti-Assad, a humanitarian group, foreign involvement in the civil war that is not Russia, foreign involvement in the war that is Russia, the civil war outside of Aleppo, entertainment, international news, a joke, poetry, religion, and any other topic.[2] This paper only uses the Anti-Assad and Pro-Assad topics as they are the topics that most directly related to our theory, and emotions are not analyzed because classification accuracy was poor.

Once cleaned,[3] the labeled tweets are split into training and test samples. A model is trained separately for each topic since topics are not mutually exclusive. To confirm that a BERT model is the best approach, naive Bayes, random forest, support vector machine, and logistic classifiers were also developed.[4] They have much lower F1, precision, and recall than the fine-tuned ARBERT model. Table A1 shows the evaluation statistics for each topic. The evaluation statistics show that of the topics related to the war, the F1 scores are highest for those about the siege of Aleppo and Russian involvement.

Sentiment - whether a tweet expresses negative, neutral, or positive tone - is measured using an AraBERT fine-tuned for sentiment on ArSentD-LEV, a dataset of tweets in Levantine Arabic (Baly et al., 2019, Antoun, Baly and Hajj, 2020); 1,000 of that model's 4,000 training tweets are from Syria. In addition to fine-tuning on the same language as the tweets in this paper, ArSentD-LEV is preferred because it outperforms other Arabic sentiment classifiers (Obeid et al., 2020).

[1] See Appendix S3 for more details on BERT and ARBERT.

[2] Light's Kappa values for inter-coder reliability all had sufficiently small p-values except for 'foreign news' and 'jokes' topics.

[3] See Appendix S3.1.

[4] See Appendix S3.4.

S3.3 Evaluation Statistics for Topics - ARBERT

Table A1: Evaluation Statistics for Topics - ARBERT

Topic	F1	Precision	Recall	Accuracy
Aleppo, not war	0.59	0.77	0.48	0.99
Assad, anti	0.54	0.52	0.56	0.95
Assad, pro	0.4	0.51	0.33	0.97
Humanitarian group	0.27	0.71	0.17	0.99
War, Aleppo	0.88	0.88	0.88	0.98
War, Russia	0.83	0.83	0.83	0.99
War, foreign not Russia	0.62	0.6	0.65	0.98
War, Syria not Aleppo	0.66	0.61	0.73	0.92
Entertainment	0.61	0.74	0.52	0.96
Foreign news	0.6	0.59	0.62	0.96
Joke	0.37	0.47	0.3	0.98
Poetry	0.48	0.72	0.36	0.93
Religion	0.89	0.94	0.85	0.93
Other	0.77	0.79	0.75	0.85

S3.4 Topic classification results on baseline Classifiers

The tables below show the classification report for the topic classification and emotion analysis using various baseline classifiers on the 6000 tweets.

Table A2: Evaluation Statistics for Topics - Support Vector Machine

Topic	F1	Precision	Recall	Accuracy
aleppo_notWar	0	0	0	1
aleppo_war	0	0	0	0.93
Assad_anti	0	0	0	0.99
Assad_pro	0	0	0	1
humanitarian_group	0	0	0	1
war_foreignNotRussia	0	0	0	1
war_foreignRussia	0	0	0	1
war_syriaNotAleppo	0	0	0	0.96
entertainment	0	0	0	0.98
foreign_news	0	0	0	0.99
joke	0	0	0	1
poetry	0	0	0	0.99
religion	0.18	0.13	0.3	0.67
other	0.02	0.01	0.42	0.51

Table A3: Evaluation Statistics for Topics - Random Forest

Topic	F1	Precision	Recall	Accuracy
aleppo_notWar	0	0	0	0.99
aleppo_war	0	0	0	0.98
Assad_anti	0	0	0	1
Assad_pro	0	0	0	1
humanitarian_group	0	0	0	1
war_foreignNotRussia	0	0	0	1
war_foreignRussia	0	0	0	1
war_syriaNotAleppo	0	0	0	0.98
entertainment	0	0	0	0.99
foreign_news	0	0	0	1
joke	0	0	0	1
poetry	0	0	0	1
religion	0.18	0.13	0.31	0.66
other	0.01	0.01	0.33	0.55

Table A4: Evaluation Statistics for Topics - Naive Bayes

Topic	F1	Precision	Recall	Accuracy
aleppo_notWar	0	0	0	1
aleppo_war	0	0	0	1
Assad_anti	0	0	0	1
Assad_pro	0	0	0	1
humanitarian_group	0	0	0	1
war_foreignNotRussia	0	0	0	1
war_foreignRussia	0	0	0	1
war_syriaNotAleppo	0	0	0	1
entertainment	0	0	0	1
foreign_news	0	0	0	1
joke	0	0	0	1
poetry	0	0	0	1
religion	0.09	0.09	0.09	0.78
other	0	0	0	0.98

Table A5: Evaluation Statistics for Topics - Logistic Regression

Topic	F1	Precision	Recall	Accuracy
aleppo_notWar	0	0	0	1
aleppo_war	0	0	0	0.93
Assad_anti	0	0	0	0.99
Assad_pro	0	0	0	1
humanitarian_group	0	0	0	1
war_foreignNotRussia	0	0	0	1
war_foreignRussia	0	0	0	1
war_syriaNotAleppo	0	0	0	0.97
entertainment	0	0	0	0.99
foreign_news	0	0	0	1
joke	0	0	0	1
poetry	0	0	0	1
religion	0.17	0.12	0.28	0.68
other	0.02	0.01	0.33	0.56

S4 Regression Tables for Main Analyses

S4.1 Changes in content

Table A6: Change in Topics related to the war, geotagged and panel data

	Anti-Assad	Anti-Assad	Pro-Assad	Pro-Assad
	(geo)	(panel)	(geo)	(panel)
Intercept	−4.788***	−2.179***	−5.147***	−5.228***
	(.688)	(.816)	(.502)	(.210)
Post-Siege	.320	−1.187***	−.492	.141
	(.324)	(.295)	(.523)	(.314)
Aleppo	1.492	−1.906**	2.152***	.430
	(1.067)	(.882)	(.786)	(.469)
Post-Siege*Aleppo	−.137	.653*	1.113**	−.463
	(.391)	(.367)	(.567)	(.469)
N	2,754	31,658	2,754	31,658
Log Likelihood	−166.730	−3,777.257	−106.022	−1,269.640
AIC	341.460	7,562.514	220.044	2,547.279

*p < .1; **p < .05; ***p < .01. Logistic regression. Standard-errors clustered by user.

Table A7: Change in Sentiment, geotagged and panel data.

	Positive	Positive	Neutral	Neutral	Negative	Negative
	(geo)	(panel)	(geo)	(panel)	(geo)	(panel)
Intercept	−.383	−.989***	−1.050***	−1.530***	−.684***	−1.530***
	(.337)	(.287)	(.294)	(.102)	(.199)	(.223)
Post-Siege	−.034	−.077	−.083	.375	.105	.375*
	(.095)	(.209)	(.092)	(.237)	(.098)	(.212)
Aleppo	−.310	1.070***	.133	.196*	.199	.196
	(.426)	(.322)	(.464)	(.112)	(.368)	(.284)
Post-Siege*Aleppo	.941**	.105	−.441	−.294	−.690	−.294
	(.401)	(.222)	(.367)	(.243)	(.580)	(.230)
N	2,754	31,658	2,754	31,658	2,754	31,658
Log Likelihood	−1,852.127	−20,682.490	−1,550.255	−16,452.530	−1,776.811	−16,452.530
AIC	3,712.255	41,372.980	3,108.510	32,913.070	3,561.623	32,913.070

*p < .1; **p < .05; ***p < .01. Logistic regression. Standard-errors clustered by user.

S4.2 Changes in account behavior

Table A8: Change in Tweets by Account Type, geotagged and panel Data

	pro-Assad account	pro-Assad account	anti-Assad account	anti-Assad account	Other account	Other account
	(geo)	(panel)	(geo)	(panel)	(geo)	(panel)
Intercept	−2.577***	−1.545	−2.722***	−.639	1.874**	−.097
	(.968)	(.977)	(1.020)	(.733)	(.769)	(.665)
Post-Siege	−1.291***	−.447	.222	−.526	.357	.671
	(.208)	(.332)	(.146)	(.510)	(.426)	(.491)
Aleppo	.576	−2.225**	2.434*	−.325	−3.321***	.883
	(1.255)	(1.133)	(1.247)	(1.080)	(.964)	(.981)
Post-Siege*Aleppo	1.601***	1.076**	−.068	.238	−.601	−.510
	(.452)	(.501)	(.447)	(.519)	(.838)	(.502)
N	2,989	33,484	2,989	33,484	2,989	33,484
Log Likelihood	−544.696	−7,755.533	−807.039	−19,089.850	−1,070.412	−21,063.490
AIC	1,097.393	15,519.070	1,622.078	38,187.690	2,148.824	42,134.990

*p < .1; **p < .05; ***p < .01. Logistic regression. Standard-errors clustered by user.

S5 Additional Analyses

S5.1 Comparing Aleppo with Damascus

Table A9: Change in Topics related to the war, Aleppo vs. Damascus accounts.

	Anti-Assad	Pro-Assad
	(geo)	(geo)
Intercept	-5.839^{***}	-4.919^{***}
	(.557)	(.559)
Post-Siege	.069	$-.851^{***}$
	(1.519)	(.105)
Aleppo	2.543^{**}	1.923^{**}
	(.990)	(.827)
Post-Siege*Aleppo	.113	1.472^{***}
	(1.535)	(.242)
N	1,463	1,463
Log Likelihood	−48.437	−72.915
AIC	104.875	153.830

*p < .1; **p < .05; ***p < .01
Logistic regression. Standard-errors clustered by user

Table A10: Change in sentiment, Aleppo vs. Damascus accounts.

	Positive	Neutral	Negative
	(geo)	(geo)	(geo)
Intercept	−.678	$-.717^{**}$	$-.684^{***}$
	(.438)	(.295)	(.224)
Post-Siege	$-.206^{*}$	.125	.072
	(.118)	(.139)	(.127)
Aleppo	−.015	−.199	.199
	(.510)	(.466)	(.383)
Post-Siege*Aleppo	1.113^{***}	$-.649^{*}$	−.657
	(.408)	(.383)	(.588)
N	1,463	1,463	1,463
Log Likelihood	−914.481	−927.881	−938.891
AIC	1,836.962	1,863.762	1,885.782

*p < .1; **p < .05; ***p < .01
Logistic regression. Standard-errors clustered by user

S5.2 Placebo accounts: Hama vs. matched accounts

Table A11: Change in Topics related to the war, Hama vs. matched accounts.

	Anti-Assad	Pro-Assad
	(geo)	(geo)
Intercept	-4.788^{***}	-5.147^{***}
	(.692)	(.504)
Post-Siege	.320	−.492
	(.326)	(.526)
Hama	-11.778^{***}	-12.419^{***}
	(1.062)	(.950)
Post-Siege*Hama	−.320	.492
	(1.183)	(1.253)
N	2,636	2,636
Log Likelihood	−145.517	−76.261
AIC	299.033	160.521

*p < .1; **p < .05; ***p < .01.
Logistic regression. Standard-errors clustered by user.

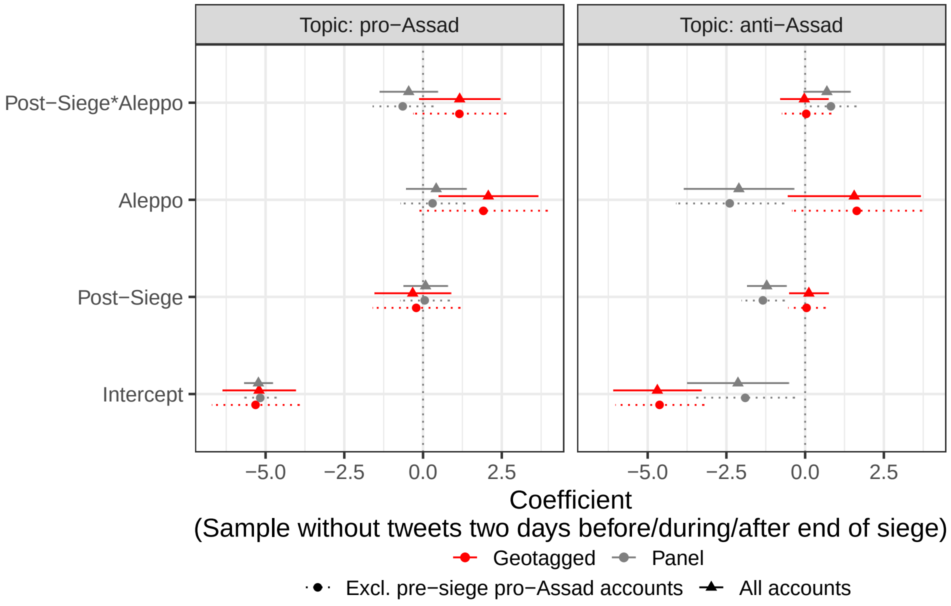
Table A12: Change in sentiment, Hama vs. matched accounts.

	Positive	Neutral	Negative
	(geo)	(geo)	(geo)
Intercept	−.383	-1.050^{***}	$-.684^{***}$
	(.339)	(.296)	(.201)
Post-Siege	−.034	−.083	.105
	(.096)	(.092)	(.099)
Hama	−.310	$.827^{*}$	−.568
	(.535)	(.442)	(.408)
Post-Siege*Hama	−.372	.306	.049
	(1.497)	(.795)	(.604)
N	2,636	2,636	2,636
Log Likelihood	−1,774.327	−1,486.002	−1,701.307
AIC	3,556.653	2,980.004	3,410.614

*p < .1; **p < .05; ***p < .01.
Logistic regression. Standard-errors clustered by user.

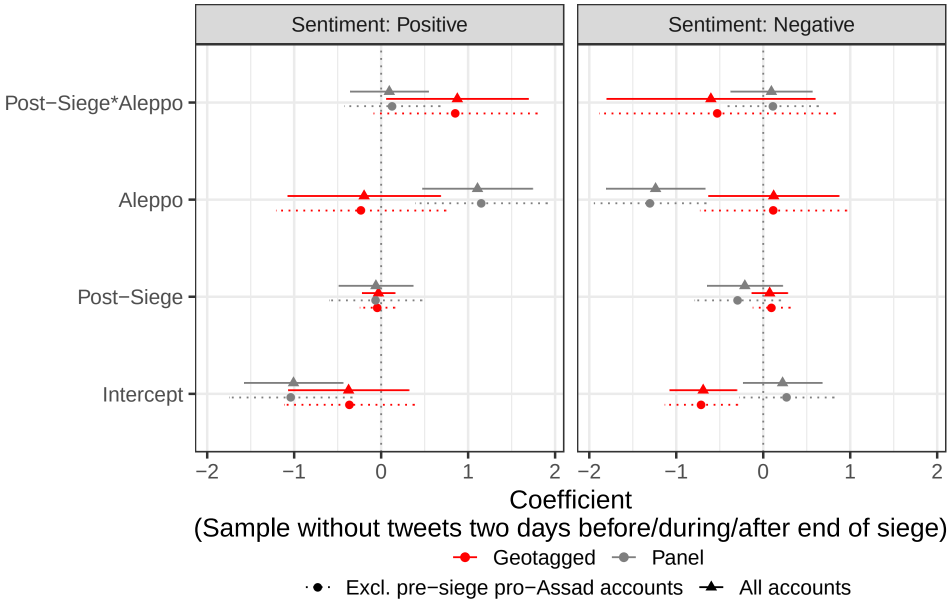
S5.3 Dropping days around the end of the siege

Figure A2: Changes in Topics in Aleppo



Note: Results for changes in topic for tweets from users in Aleppo, compared to matched users outside of Aleppo, **dropping tweet two days before, during, and two days after the end of the siege (12-17. December 2016)**. Post-Siege*Aleppo is of main interest; the red point estimate is for the geotagged data, the grey is for panel data. The estimates with dotted confidence intervals replicate the main results, but exclude pre-siege end pro-regime accounts. The difference in inference when comparing the geotagged and panel results is suggestive of signalling behavior. For full regression table, see Table A13.

Figure A3: Changes in Sentiment in Aleppo



Note: Results for changes in sentiment for tweets from users in Aleppo, compared to matched users outside of Aleppo, **dropping tweet two days before, during, and two days after the end of the siege (12-17. December 2016)**. Post-Siege*Aleppo is of main interest; the red point estimate is for the geotagged data, the grey is for panel data. The estimates with dotted confidence intervals replicate the main results, but exclude pre-siege end pro-regime accounts. The difference in inference when comparing the geotagged and panel results is suggestive of signalling behavior. For full regression table, see Table A14.

S5.3.1 Tables for Figures in Section S5.3

Table A13: Change in topics related to the war, geotagged and panel data, dropping days around the end of the siege.

	Anti-Assad	Anti-Assad	Pro-Assad	Pro-Assad
	(geo)	(panel)	(geo)	(panel)
Intercept	−4.693***	−2.133***	−5.208***	−5.228***
	(.713)	(.822)	(.592)	(.230)
Post-Siege	.118	−1.220***	−.330	.080
	(.317)	(.318)	(.619)	(.358)
Aleppo	1.558	−2.103**	2.072**	.417
	(1.075)	(.892)	(.805)	(.489)
Post-Siege*Aleppo	−.027	.691*	1.163*	−.456
	(.389)	(.380)	(.656)	(.470)
N	2,493	29,146	2,493	29,146
Log Likelihood	−150.132	−3,360.882	−95.834	−1,137.435
AIC	308.265	6,729.764	199.667	2,282.870

*p < .1; **p < .05; ***p < .01
Logistic regression. Standard-errors clustered by user.

Table A14: Change in sentiment, dropping days around the end of the siege.

	Positive	Positive	Neutral	Neutral	Negative	Negative
	(geo)	(panel)	(geo)	(panel)	(geo)	(panel)
Intercept	−.375	−1.008***	−1.053***	−1.535***	−.690***	−1.535***
	(.354)	(.290)	(.310)	(.103)	(.197)	(.232)
Post-Siege	−.031	−.060	−.049	.391	.074	.391*
	(.097)	(.218)	(.095)	(.249)	(.106)	(.222)
Aleppo	−.196	1.108***	.098	.205*	.120	.205
	(.449)	(.324)	(.469)	(.115)	(.383)	(.290)
Post-Siege*Aleppo	.876**	.094	−.500	−.308	−.602	−.308
	(.417)	(.230)	(.341)	(.254)	(.612)	(.240)
N	2,493	29,146	2,493	29,146	2,493	29,146
Log Likelihood	−1,680.127	−19,015.030	−1,409.468	−15,172.480	−1,599.692	−15,172.480
AIC	3,368.253	38,038.060	2,826.937	30,352.960	3,207.385	30,352.960

*p < .1; **p < .05; ***p < .01. Logistic regression. Standard-errors clustered by user.

S5.4 Changes in content, excluding pre-siege end pro-Assad accounts

Table A15: Change in Topics related to the war, geotagged and panel data, excluding pre-siege pro-regime (pro-Assad) accounts

	Anti-Assad	Anti-Assad	Pro-Assad	Pro-Assad
	(geo)	(panel)	(geo)	(panel)
Intercept	−4.706***	−1.947**	−5.220***	−5.156***
	(.690)	(.791)	(.574)	(.225)
Post-Siege	.260	−1.290***	−.397	.128
	(.313)	(.330)	(.571)	(.346)
Aleppo	1.542	−2.175**	2.056**	.298
	(1.048)	(.863)	(.975)	(.494)
Post-Siege*Aleppo	−.040	.747*	1.049*	−.691
	(.403)	(.400)	(.623)	(.490)
N	2,611	29,398	2,611	29,398
Log Likelihood	−164.834	−3,544.993	−93.645	−1,102.082
AIC	337.668	7,097.987	195.290	2,212.165

*p < .1; **p < .05; ***p < .01
Logistic regression. Standard-errors clustered by user.

Table A16: Change in sentiment, excluding pre-siege pro-regime (pro-Assad) accounts.

	Positive	Positive	Neutral	Neutral	Negative	Negative
	(geo)	(panel)	(geo)	(panel)	(geo)	(panel)
Intercept	−.373	−1.019***	−1.051***	−1.559***	−.693***	−1.559***
	(.366)	(.355)	(.320)	(.123)	(.217)	(.268)
Post-Siege	−.054	−.088	−.064	.495**	.111	.495**
	(.100)	(.260)	(.093)	(.221)	(.104)	(.244)
Aleppo	−.361	1.114***	.191	.227*	.197	.227
	(.471)	(.385)	(.502)	(.133)	(.406)	(.322)
Post-Siege*Aleppo	.888*	.147	−.462	−.387*	−.584	−.387
	(.473)	(.270)	(.406)	(.227)	(.655)	(.257)
N	2,611	29,398	2,611	29,398	2,611	29,398
Log Likelihood	−1,755.078	−19,223.650	−1,475.245	−15,436.130	−1,684.079	−15,436.130
AIC	3,518.157	38,455.300	2,958.490	30,880.250	3,376.157	30,880.250

*p < .1; **p < .05; ***p < .01. Logistic regression. Standard-errors clustered by user.

Supplementary Material References

Abdul-Mageed, Muhammad, Abdelrahim A Elmadany and El Moatez Bill. 2021. ARBERT & MARBERT: Deep Bidirectional Transformers for Arabic. In *59th Annual Meeting of the Association for Computational Linguistics.* pp. 7088–7105.

Abu El-khair, Ibrahim. 2016. "1.5 billion words Arabic Corpus." *arXiv e-prints* p. arXiv:1611.04033.

Antoun, Wissam, Fady Baly and Hazem Hajj. 2020. AraBERT: Transformer-based Model for Arabic Language Understanding. In *Proceedings of the 4th Workshop on Open-Source Arabic Corpora and Processing Tools, with a Shared Task on Offensive Language Detection.* pp. 9–15.

Baly, Ramy, Alaa Khaddaj, Hazem Hajj, Wassim El-Hajj and Khaled Bashir Shaban. 2019. "ArSentD-LEV: A Multi-Topic Corpus for Target-based Sentiment Analysis in Arabic Levantine Tweets.".
URL: *http://arxiv.org/abs/1906.01830*

Devlin, Jacob, Ming Wei Chang, Kenton Lee and Kristina Toutanova. 2019. BERT: Pre-training of deep bidirectional transformers for language understanding. In *NAACL HLT 2019 - 2019 Conference of the North American Chapter of the Association for Computational Linguistics: Human Language Technologies - Proceedings of the Conference.* Vol. 1 Association for Computational Linguistics (ACL) pp. 4171–4186.

Eljundi, Obeida, Wissam Antoun, Nour El Droubi, Hazem Hajj, Wassim El-Hajj and Khaled Shaban. 2019. hULMonA: The Universal Language Model in Arabic. In *Proceedings of the Fourth Arabic Natural Language Processing Workshop.* pp. 68–77.

Obeid, Ossama, Nasser Zalmout, Salam Khalifa, Dima Taji, Mai Oudah, Bashar Alhafni, Go Inoue, Fadhl Eryani, Alexander Erdmann and Nizar Habash. 2020. {CAM}e{L} Tools: An Open Source Python Toolkit for {A}rabic Natural Language Processing. In *Proceedings of the 12th Language Resources and Evaluation Conference.* Marseille, France: European Language Resources Association pp. 7022–7032.
URL: *https://aclanthology.org/2020.lrec-1.868*

Ortiz Suárez, Pedro Javier, Benoît Sagot and Laurent Romary. 2019. Asynchronous Pipeline for Processing Huge Corpora on Medium to Low Resource Infrastructures. In *7th Workshop on the Challenges in the Management of Large Corpora (CMLC-7)*, ed. Piotr Bański, Adrien Barbaresi, Hanno Biber, Evelyn Breiteneder, Simon Clematide, Marc Kupietz, Harald Lüngen and Caroline Iliadi. Cardiff, United Kingdom: Leibniz-Institut für Deutsche Sprache.
URL: *https://hal.inria.fr/hal-02148693*

Zeroual, Imad, Dirk Goldhahn, Thomas Eckart and Abdelhak Lakhouaja. 2019. {OSIAN}: Open Source International {A}rabic News Corpus - Preparation and Integration into the {CLARIN}-infrastructure. In *Proceedings of the Fourth Arabic Natural Language Processing Workshop.* Florence, Italy: Association for Computational Linguistics pp. 175–182.
URL: *https://aclanthology.org/W19-4619*

Zhu, Yukun, Ryan Kiros, Rich Zemel, Ruslan Salakhutdinov, Raquel Urtasun, Antonio Torralba and Sanja Fidler. 2015. Aligning books and movies: Towards story-like visual explanations by watching movies and reading books. In *Proceedings of the IEEE international conference on computer vision*. pp. 19–27.