Civilian Signalling on Social Media During Civil War

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

[1]Hertie School, Berlin

[2]University of California, Los Angeles

March 4, 2022

Abstract

Recent research has highlighted the mobilization potential of social media, which can offer citizens who were previously motivated to hide their true preferences an easier way to share their grievances and find common support. It is less clear how these changing dynamics of revealing preferences affect contentious processes beyond initial mobilisation. We argue that in conflict settings, previously shared social media posts indicating political loyalties can pose a severe risk for civilians. For example, anti-regime sentiments or display of digital support for opposition activities may prove to be life-threatening in government controlled areas. We expect civilians to strategically alter their social media usage, in particular when faced with profound changes to territorial control. We study dynamics of social media usage in Syria, focusing on the end of the siege of Aleppo in late 2016. Using geolocated tweets we compare Twitter users in Aleppo to users in other parts of Syria to understand how the siege impacted their online activity, sentiment, and emotion. The findings have important implications for our understanding of the risks - and the potential for civilian agency - of everyday digital communication in civil conflict.

*gohdes@hertie-school.org

†zst@luskin.ucla.edu

Introduction

For more than a decade, social media has been a central companion to contentious political processes. From isolated protests to country-wide uprisings, to organized armed conflict, non-state and government actors have learned that their actions are likely to be caught on cellphone camera, and that controlling the digital narrative in the chaos of conflict can offer decisive advantages. Growing research has helped advance our understanding of how conflict shapes social media, and conversely, how social media influences conflict dynamics. Studies have contended that social media reduces the cost of coordination (Little, 2015), increases the speed of its dissemination, and provides passive polling to conflict actors (Zeitzoff, 2017). These effects are generated through network dynamics (Enikolopov, Makarin and Petrova, 2016), communication, and media capabilities (Kwak et al., 2010, Shapiro and Siegel, 2015).

In addition to studying the effect of social media on conflict, data gleaned from social media platforms can be used to study network processes (Barberá et al., 2015, Romero, Meeder and Kleinberg, 2011) public opinion (Beauchamp, 2019), political representation (Barberá and Zeitzoff, 2018, Oklobdzija, 2018, Tucker, 2019), protests (Larson et al., 2019, Mooijman et al., 2018, Sobolev et al., 2020), and a wide range of other phenomena (Barbera and Steinert-Threlkeld, 2020).

Particular caution is required when interpreting social media in situations where everyday politics have turned violent. In the midst of conflict, information shared online can create 'a dangerous illusion of unmediated information flows.'(Lynch, Freelon and Aday, 2014, 3), that disregards important curation occurring by local and international stakeholders. While previous work has analyzed social media as a tool for mobilization (Steinert-Threlkeld, 2017), its use by rebel actors (see Jones and Mattiacci, 2019), and the ways in which it affects conflict dynamics (Zeitzoff, 2018), far less is known about how social media is used by individuals caught in the midst of conflict. In this paper we are interested in understanding everyday social media usage in the context of ongoing armed conflict.

Social media use during contentious politics

Social media is often an essential source of anti-regime information. This dynamic was especially true during the Arab Spring when governments were not yet employing coordinated harassment and derailing campaigns to affect non-state political coordination (Brym et al., 2014, Tufekci and Wilson, 2012). While early research and anecdotal evidences suggests that social media provides non-state actors with an advantage (Bennett and Segerberg, 2013, Gunning and Baron, 2013), those findings can likely be attributed to the fact that initial adopters were anti-status quo. Since then, pro-government actors have been actively catching up Sanovich, Stukal and Tucker (2018), Stukal et al. (2019).

More generally, social media may make protests harder to start since the state can monitor them for discontent and target key organizers preemptively(Gohdes, 2020, Xu, 2021). Where individuals are able to mobilize, however, the logistical benefits make protests larger than they otherwise would have been (Weidmann and Rod, 2018). Overall, existing evidence suggests that social media does not make protest more likely since it reveals pro-status quo signals and can be manipulated, but it facilitates ongoing protest via logistic coordination (Little, 2015, Steinert-Threlkeld, 2017).

There is growing awareness about how armed actors use social media during civil conflict. For rebel groups, social media provides a number of ways to improve their capabilities. Groups that may otherwise not have access to mainstream media can use social media do directly seek international support, publicize battlefield success, denounce those in power, and disseminate policy goals (Jones and Mattiacci, 2019). Social media allows groups to provide their own media, an especially important capability in countries with closed media systems (Sutton, Butcher and Svensson, 2014).

Beyond armed political conflict, social media has been shown to empower individuals in producing and consuming locally relevant information in response to violent events. In the context of Mexico's ongoing drug war, individuals have used Twitter to provide real-time alerts about new violence, with some channels becoming prominent enough to be functionally

similar to traditional media (Monroy-Hernández et al., 2013). Facebook, especially through its `Groups` feature, enabled the creation of local self-defense militias who united to defending themselves against the drug cartel *Knights Templar* (Savage and Monroy-Hernandez, 2015). Social media is also a place where emotional reactions to violence are publicly expressed. Researchers have documented the expression of fear, sadness, and anger in response to gun violence (Jones et al., 2016, Manuel and Valdes, 2015), in particular by those exposed to it (Saha and Choudhury, 2017). Similar behaviors are seen in response to terrorist attacks (Eriksson, 2018).

Case evidence suggests that social media make transitions from unrest to regular politics more fragile. Online debates can quickly turn into battles over the meaning of ongoing events, such as they occurred during Ukraine's Euromaidan protests (Driscoll and Steinert-Threlkeld, 2020, Metzger, Nagler and Tucker, 2015). Studies on the Arab Spring argue that social media undermined democratic transitions after the protests because it encouraged ideological enclosure, fueling paranoia and mistrust (Lynch, Freelon and Aday, 2016). Because social media can mobilize large numbers, it may discourage the development of organizational structures that facilitate the transition to democracy, making protest more effective in the short-term, but weakening movements in the long-term (Tufekci, 2017).

Analysis of strategic social media use has thus focused on onset of contentious politics; few studies examine civilians' use of social media beyond initial mobilization. Researchers have studied how social media is used to draw international attention to domestic issues (Najjar, 2010). Analyzing the strategy of doctor-activists, Alasaad (2013) shows how this specific group used Facebook and YouTube to spread international awareness about a leishmaniasis outbreak in Syria's Deir Ezzor province in 2013.

While we know that violent events are often reflected on social media, and conflict actors will strategically use platform, the role of social media used by civilians *during* conflict remains understudied. In this paper we approach civilian use of social media from signalling perspective. We assume that one strategic use of civilian social media use is related to

signally loyalty and support to armed group actors. In the following section we discuss the logic of signalling civilian loyalty in conflict, and develop theoretical expectations on how changes in conflict dynamics should influence civilian use of social media in conflict.

Social media as a tool for strategic singalling

Incentives to signal wartime support

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, and yet winning the 'hearts and minds' of civilians is commonly seen as a crucial step to winning wars (Beath, Christia and Enikolopov, 2012). 'Hearts and minds' as a strategy does not imply that individuals have to enthusiastically embrace an armed group. Instead it has been argued that 'calculated self-interest, not emotion, is what counts' (United States. Department of the Army. and United States. Marine Corps., 2007, 294). This understanding of support builds on the notion that civilians caught in conflict will primarily be driven by rational self-interest, which in turn should mean that, on average, non-combatants will be non-ideological about whom to support (Popkin, 1979).

A core way in which civilian support can manifest itself is through the willingness to collaborate with an armed actor. Kalyvas (2006) contends that incentives to collaborate are principally driven by the question of who controls a given territory, and that civilians will be more willing to collaborate with armed groups where they need not fear reprisals. In a nutshell, Kalyvas argues that

> Irrespective 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)

As a consequence, civilians should have an incentive to support armed groups when they anticipate this increasing their chances of security, and by consequence conflict dynamics are

expected to be a key explanation for variations in civilian support. Yet while extensive and important research has discussed the effects of violence on civilian preferences, few studies examine how civilians end up making their support for armed actors publicly known in the first place.[1] This question is important as displays of civilians support can take on many different forms. In situations of incomplete information - such as during a conflict - actors are likely to rely on visible signals to communicate their preferences (see more generally, 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).

Next to food, supplies, and shelter, non-material support in the form of public displays of loyalty, solidarity, and positive morale is a key method of showing support for armed actors. Such behavioral signs may include displaying pro-government slogans in windows, whether at home or one's business, is a common tactic individuals use to signal their preferences (Havel and Wilson, 1986).[2] Singing opposition songs or flying their flags is a more identifiable, and therefore riskier, signal (Pfaff, 1996). Behavioral signals, however, are only effective when the target can directly observe it.

An important recent study by Schubiger (N.d.) argues that civilians fearing collective targeting will display highly observable behavioral measures to signal their support of the government, in the hope of being spared excessive coercive force. Studying the civil war in Peru, she finds that communities victimized by the Peruvian state forces were more likely to mobilize against insurgents in an attempt to signal their non-affiliation with the insurgents. Importantly, she discusses how challenging it can be for civilians to signal their allegiance when caught in conflict, and that successful signals have to be highly visible in order to be

[1]The literature on civilian support is extensive and highly nuanced. For example, research by Lyall, Blair and Imai (2013) finds that group identity mediates civilian support for armed groups, where violence perpetrated by members of the civilians' 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.

[2]Whether or not these signals reflect true preferences is a different matter (Kuran, 1991).

received by the recipient. This explains the extreme measures Peruvian communities went to when organizing pro-government self-defence forces. Where armed forces rarely come into contact with civilians, such as when using indirect violence through airpower, civilians will have a much harder time having their signals received (Schubiger, N.d., 6).

We expect that through the rise of digital communication, in particular public communication on social media, public digital displays of allegiance, sentiment, or emotions will offer an opportunity to make such wartime signals more readily observable. The next section discusses these possibilities.

Signalling on social media

We argue that social media will we used as a device for strategic signalling in conflict situations where 1) at least one conflict side is monitoring social media for intelligence 2) where individuals are generally aware of the monitoring activities and have witnessed conflict actors gleaning information on defectors or dissidents from social media, and 3) where public facing social media platforms are widely available and used. In addition, we assume that individuals are more likely to use social media for political than apolitical purposes in the context of an ongoing conflict. Where individuals are readily aware of the fact that the audience for their public-facing social media content is likely to be consumed and weaponized by conflict actors signalling preferences online can be a strategic choice.

This focus on social media as a signalling tool is distinct from earlier claims about the coordinating effect 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, Spaiser et al., 2017), and the state directly injects pro-status quo messages via bots (automatic accounts) and trolls (paid content) (King, Pan and Roberts, 2017, Lukito, 2019). While the 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' signals (Driscoll and Steinert-Threlkeld, 2020, Metzger, Nagler and Tucker, 2015).

Changes in sentiment and emotion

Signalling on social media can broadly be understood by studying both the content and metadata of users. Content refers to actual statements and topics covered in social media posts. If users share content in direct support of an armed actor this would be equivalent to overt offline support activities, such as hanging pro-government posters or wearing a lapel pin. The metadata we can observe would pertain to users' social media activity, including posting frequency, number and types of accounts followed, as well as the start of termination of using a social media account. To compare this to offline forms of signalling, these might include the active participation in events organized by armed actors, reading their newspapers, listening to radio stations, and attending rallies. Individuals may engage in action that signals positive support of the regime, such as attending a pro-regime rally, that is negative, such as voicing dissent, or neutral, by abstaining from saying of behaving in ways that could be interpreted as taking a position towards the government. Positive or negative statements may signal support for the government. Such statements could be positive ("Long live our leader"), negative ("Death to the regime."), or neutral ("I like basketball."). The valence of such statements is determined in relation to the object of the statement or action, and is commonly referred to as sentiment.

Beyond sentiment, social media users may signal emotions through their use of social meida. Research on emotions generally distinguishes between reflex emotions and enduring emotions. Reflex emotions quickly arise in response to a direct stimulus, while enduring emotions, often called moods or affective orientations, form slowly and change less frequently (Jasper, 2006). Emotions such as fear, sadness, shame, anger, joy, and pride are generally associated with reflex, while archetypal enduring emotions include love, hate, and respect.

Because reflex emotions respond more directly to stimulus than enduring ones, they have received the most attention in studies of political mobilization. Anger is traditionally seen as the emotion most likely to translate a stimulus into mobilization (Valentino et al., 2011). For example, individuals may observe the arrest of a prominent opposition official or taxing the internet, and consequently take to the streets. Evidence from Zimbabwe suggests that fear causes individuals to focus on the costs of mobilization and therefore leads to demobilization (Young, 2019). Pearlman (2013) groups the reflex emotions into dispiriting ones that hinder mobilization (fear, sadness, and shame) and emboldening ones that facilitate it (anger, joy, and pride).

Joy and pride are the reflex emotions that should signal support for the situation in which an individual finds themselves. Joy is the short-term equivalent of the enduring emotion of happiness: it is the transitory positive emotion arising in response to a specific positive development. Pride arises when one connects an outcome to the actions of the self or a group with which the self identifies (Williams and DeSteno, 2008). For example, one may experience joy at the prospect of the end of indiscriminate violence or medical supply shortages. Pride may develop if one feels their actions contributed to the outcome towards which those actions were directed. Anger, fear, sadness, and shame should signal dissatisfaction with a conflict's result. None of these emotions should be interpreted as support for the victorious side in conflict situation. Instead,

Unlike previous research that focused on the consequences of emotions for contentious politics, we aim to study how changes in conflict dynamics affect the ways in which individuals publicly share their sentiment and emotion online. We expect to observe an increase in positive sentiment among social media users in the aftermath of major changes in conflict dynamics, such as changes in who controls a territory. We expect this for two reasons. First, individuals who are loyal to the 'winning' conflict side will likely share content with a positive sentiment and emotions, such as pride and joy. Second, we expect individuals who share no loyalty with the winning side to falsify their preferences online (Kuran, 1991) and

nonetheless express positive reflex emotions. This signalling via positive sentiment is likely to be a defense mechanism, and whether or not it reflects the signaller's true internal state is irrelevant so long as it convinces the receiving parties of its validity.

Changes in account characteristics

Signalling may also occur through activity patterns. The most visible change is the activity with which civilians use their social media account. While individuals may switch from making neutral or negative statements to posting positive (i.e. pro-government) content, such actions may feel too risky or too emotionally difficult for those who have suffered at the hands of a conflict actor. In addition, content that is critical of the government could be unearthed by regime forces when attempting to identify insurgent sympathizers. To avoid scrutiny of potentially incriminating content, users may instead radically reduce their social media activity, or stop using an account altogether.

Conversely, individuals may want to signal support for the regime by increasing their social media use. Where government forces are known to monitor social media (Tibken, 2016), civilians are likely to be aware of the fact that the content they post online may lead to questioning or even arrest if it is deemed critical of regime activity. When and where government forces capture or recapture a city or larger territory, the use of social media is likely to bring with it new risks that may motivate users to either delete previously posted content or delete their account completely. Conversely, individuals may be motivated to create new accounts to signal loyalty - or at least neutrality - towards the government. Where the government has recaptured territory, users may create new accounts to share positive content so as to signal loyalty (or at least cooperation) with the regime. While many people do use social media to lurk (passively observe), many others will use it to express pro-regime sentiment. Finally, individuals may change their account's screen name or biographical information to signal their support (or the absence thereof) in light of major conflict changes.

We expect that major changes in local conflict dynamics will have an impact on sentiment

of social media content shared from the specific local conflict location that experienced the change. Such changes may manifest themselves where a territorial control changes from being contest to being under full control of one conflict party. We study social media use during the Syrian Conflict, focusing specifically on changes in activity and content during and after the regime's siege of Aleppo in 2016.

Social Media and the Syrian Conflict

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). The massacre of the first protesters in Daraa and the torture of a 13 year old boy were shared widely on YouTube, Twitter, and Facebook. 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, which is pro-Assad, used Facebook to identify activists and establish pro-regime signals (Shehabat, 2012). Armed actors on all sides of the conflict made use of social media to communicate their (change in) allegiances, spread propaganda, and communicate with both their domestic and foreign audience (Moss, 2018). 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, Gohdes, 2020). Internet accessibility 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, 2015, 2020).

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, Price, Gohdes and Ball, 2015), little research exists on explicit ways in which dynamics of the conflict itself, such as changes in the composition of conflict parties and shifts in territorial control directly impact the nature of social media discourse.

The siege of Aleppo

Syrian government forces and rebels battled for control of Aleppo, Syria's most populous city, starting in mid 2012, following months of protests. By November 2012 the city was divided into a regime and a rebel controlled area, and caught in a deadlock. Since then, Aleppo has been at the center of some of the worst fighting, with the regime using barrel bombs to attack the rebel-held areas in an attempt to regain control. 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 Dera'a 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, in a coalition with Russian airforces, submitted the area to intense bombardment of for twelve days, targeting core civilian infrastructure, such as hospitals (Böttcher, 2017, 2-3). On December 13, a highly complex ceasefire was negotiated (ibid), which involved the hand-over of weapons and transfer of all remaining rebels to other territories, which resulted in an estimated relocation of one hundred thousand individuals (Atlantic Council, 2017, 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.

In the following analysis, we use December 15 as actual end of the siege, taking into account the additional days of population displacement, and the complexity of the ceasefire deal.[3]

[3]We rerun the analysis using both the official end on December 12 Bassam, McDowall and Nebehay (2016), and December 22 - when the Syrian Army declared having regained full control of the city as cut-off dates, and the results are very similar.

Data and Methods

Twitter

Twitter is the platform we use to study social media signalling in the Syrian conflict because it is the most likely to contain signalling. 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. We are also primarily interested in individual behavior, which would not be represented by aggregate Pages. Since about 90% of Twitter accounts are public, it is the most likely platform on which signalling will occur. Other research has shown that Syrians have used it throughout the country's protests and subsequent civil war. Accounts on it are also heterogeneous in their politics: there are, among others, secular revolutionaries, Islamists, suppporters of the Free Syrian Army, and pro-Assad accounts (Freelon, Lynch and Aday, 2015). In addition, the Twitter data used for this paper shows that new users from Syria have consistently joined and tweeted throughout the period under study; see Figures A1 and A2.

Our primary reason to study Twitter data, however, relates to our ability to causally identify changes in user behavior. For this we require geographic information. This paper gathers geolocated tweets using a connection one of the authors maintains to Twitter's POST statuses/filter endpoint, the "streaming API" as it is often known. Only tweets with longitude and latitude coordinates are returned, and this process collects between one-third

and one-half of all tweets with coordinates Steinert-Threlkeld (2018).[4] 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 58,433 tweets from 3,269 accounts. In order to understand how Twitter users in Aleppo reacted to the end of the siege, we match accounts in Aleppo to accounts in other parts of Syria based on the content of tweets posted *prior* to the end of the siege in Aleppo (December 15th, 2016), using cosine similarity (Pan and Siegel, 2020). Note that we only match accounts that tweet in Arabic 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, we believe this approach provides us with a conservative approach to selecting users. This process identifies 335 Arabic-speaking Twitter accounts that were active and sent geolocated tweets from within Aleppo prior to the end of the siege, and matches them to the same number of accounts from other parts of Twitter that display the closest similarity to each account.[5]

To analyze the impact of the Aleppo siege's end on signalling we compare data from Aleppo with a control group of accounts that were active in Syria (outside of Aleppo) during the same time period. Even if one were to acquire data from the alternative platforms mentioned in the previous paragraph, the location of production of the content would need to be inferred. By contrast, Twitter assigns geographic information to tweets, often down to a specific latitude and longitude location. This geographic information facilitates the creation of a treatment and control group based on likely exposure to the siege.[6] The use of Twitter therefore provides the necessary amount of data to comparatively analyze signalling. The platform has also been used by other researchers to study periods of intense conflict

[4]Tweets with location coordinates represent 2-3% of all tweets, and Twitter matches the parameters of a request up to a 1% ceiling, so $\frac{1}{2} - \frac{1}{3}$ of geolocated tweets are returned.

[5]We use an alternative approach where we match on account characteristics, including user description, follower count, friend count, and favourite count. *Results to be added.*

[6]There is a large and impressive body of work examining different facets of the Syrian conflict (Pearlman, 2020, Wedeen, 2019), but those approaches do not fit our research question because studying signalling would have required researcher presence throughout Syria at dangerous moments.

that would otherwise have been inaccessible (Driscoll and Steinert-Threlkeld, 2020, Zeitzoff, 2011). Having identified users in Aleppo and matched them to users in Syria whose tweeting content is most similar before the siege's end, we use difference-in-differences design is then used. We define the treatment as the siege's end to study how this major change in the conflict's dynamic affected social media usage of users based in Aleppo, when compared to similar users in other parts of Syria. The primary models control for the date of the tweet (pre- or post-siege), and whether an account is based i Aleppo (or another part of Syria). For the analyses at the tweet level we control for the length of each tweet. Additional controls are added in later robustness checks. Standard errors are clustered by location, and the unit of analysis is the tweet.

To further examine signalling, we then download the entire tweet history of the treatment and control accounts. Since the streamed data are a sample, downloading all of an account's tweets provides two benefits. First, it reveals whether accounts signal differently based on whether or not a tweet is assigned a location. Second, this process will not return all the original accounts because some will have gone private or are no longer active on Twitter.[7] Comparing the streaming activity of the accounts no longer public with those still public will suggest whether or not user behavior changes as a result of the siege's ending.

Analysis of Signalling on Social Media

Three related dependent variables are created, two for sentiment and a third for emotion. The sentiment outcomes are the percentage of a tweet's words that positive and the ratio of positive to negative words in a tweet. The emotion outcome is the the ratio of positive to negative emotions expressed in a tweet.

A dictionary is used to estimate sentiment per tweet. We use EmoLex, also known as the National Research Council (NRC) Emotion lexicon; it is a dataset of 10,170 terms mapped to positive and negative sentiment as well as the emotions anger, anticipation, disgust, fear, joy,

[7]When querying an account that is private or deleted, Twitter will only reveal that it cannot provide data but not whether or not it is private or deleted.

sadness, surprise, and trust (Mohammad and Turney, 2013). We use the Arabic translation of the NRC Emotion Lexicon; these translations come from the original NRC team, and validation has shown the translated dictionary recovers true sentiment accurately (Salameh, Mohammad and Kiritchenko, 2015). The first dependent variable is the percent of words of positive sentiment, the count of the number of positive sentiment words divided by the number of words in the tweet. The second dependent variable is the ratio of positive to negative sentiment words per tweet.

Sentiment

Figure 1 shows weekly positive and negative sentiment estimates for the accounts in Aleppo and their matches elsewhere in Syria.The grey area denotes the time period leading up to the end of the siege, which is the period by which the accounts were matched using cosine similarity.

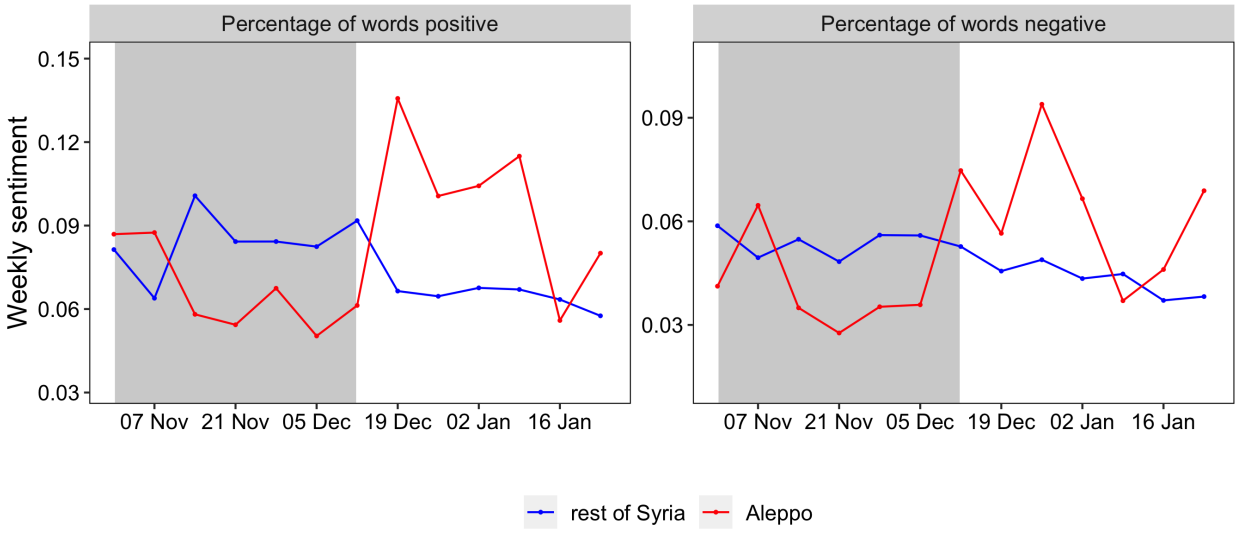


Figure 1: Sentiment of matched accounts that tweeted continuously in Aleppo versus the rest of Syria. The shaded area denotes the time period in which accounts were matched by cosine similarity.

To better understand sentiment, a third dependent variable is created, the ratio of positive to negative emotion words per tweet. The negative reflex emotions the NRC and its Arabic translation contains are anger, disgust, and fear, and it contains only one positive reflex

emotion, joy. We also include the positive emotion trust but exclude surprise because the siege's outcome was likely not surprising by December. The sum of joy and trust words is divided by the sum of anger, disgust, and fear words to create the ratio. A ratio is preferred in order to minimize the effect of measurement error in any of the five constituent emotions.

Table 1 presents the results of difference in difference regressions that attempt to showcase changes in the sentiment of tweets sent from Aleppo-based accounts in the aftermath of the siege.

Figure 2 shows the average positive sentiment for tweets from Aleppo and those that were matched from the rest of Syria. Figure 3 does the same but for the ratio of positive to negative words per tweet.

Table 1: Difference-in-differences analysis of sentiment and emotions. Post Siege: From 15 December 2016 onwards.

	Positive Words (Percentage) (1)	Positive Words/ Negative Words (2)	Positive Emotions/ Negative Emotions (3)
Post-Siege	$-.0160^{***}$	$-.0101^{**}$	$-.0094^{*}$
	(.0041)	(.0040)	(.0054)
Aleppo	$-.0274^{**}$	$-.0173^{*}$	$-.0329^{**}$
	(.0107)	(.0104)	(.0140)
Words	$.0030^{***}$	$.0005^{*}$	$-.0006^{*}$
	(.0003)	(.0003)	(.0004)
Post-Siege* Aleppo	$.0633^{***}$	$.0436^{**}$	$.0487^{**}$
	(.0176)	(.0171)	(.0231)
Intercept	$.0441^{***}$	1.0261^{***}	1.0326^{***}
	(.0047)	(.0046)	(.0061)
N	2,247	2,247	2,247
R^2	.0565	.0061	.0047
Adjusted R^2	.0548	.0044	.0030

$^{*}p < .1$; $^{**}p < .05$; $^{***}p < .01$

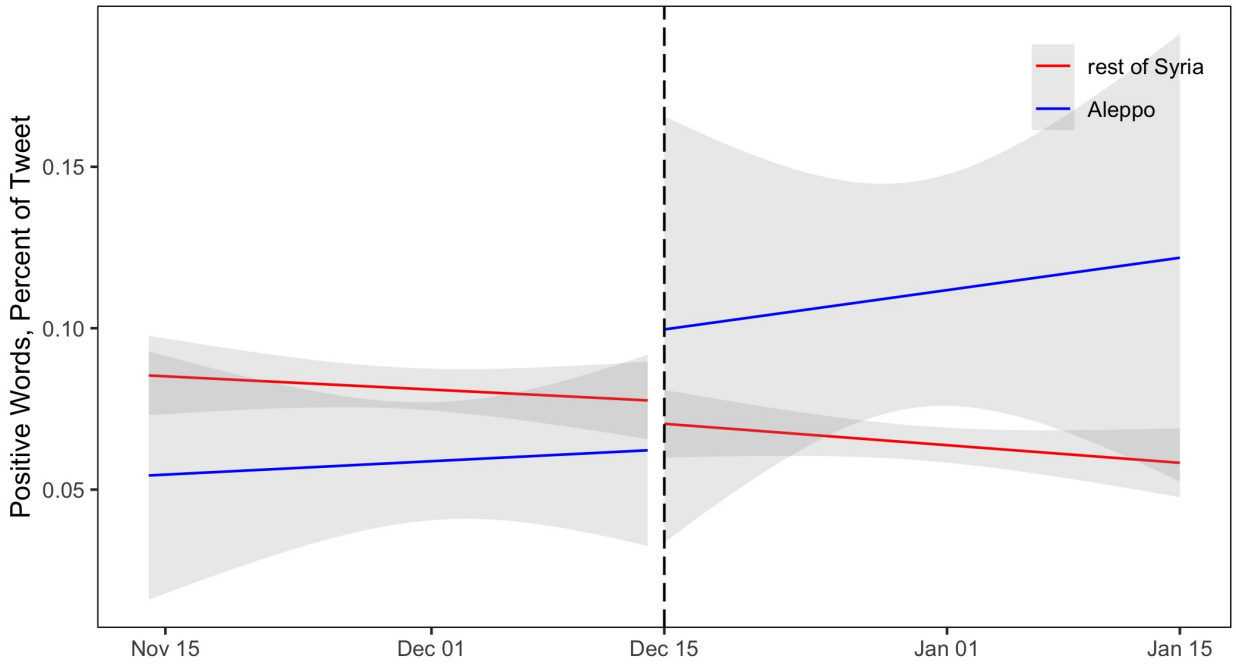


Figure 2: Difference in difference regression, comparing positive sentiment of tweets in Aleppo to paired accounts from the rest of Syria, before and after the end of the siege in Aleppo.

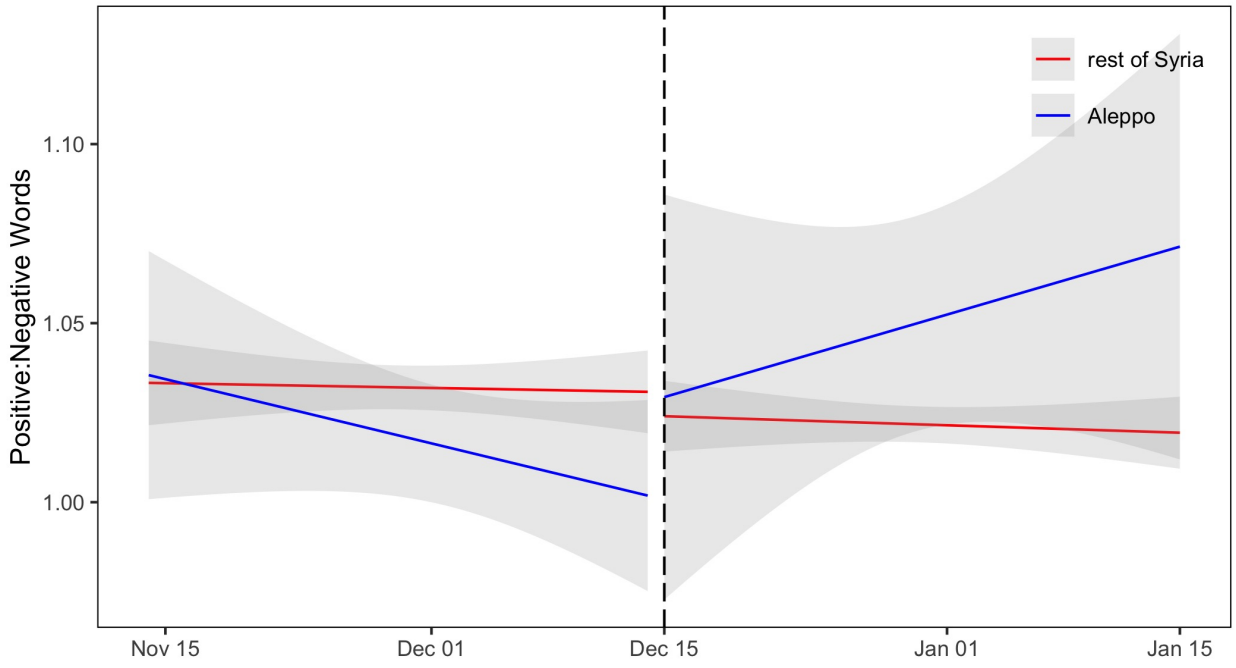


Figure 3: Difference in difference regression, comparing sentiment (positive to negative words ratio) of tweets in Aleppo to paired accounts from the rest of Syria, before and after the end of the siege in Aleppo.

Emotions

To better understand the mechanisms driving the increased expressed positiveness, we next analyze the effect of the siege's end on specific emotions: anger, fear, disgust, trust, and joy.Measuring emotion from text is challenging, and it is more difficult when the text is short and idiomatic like tweets. Emotions are also especially difficult to measure because they are interior states that are only sometimes expressed. Measuring emotions therefore almost always means measuring *expressions* of emotions, and the function connecting emotions to their expression is unknown. Since we expect emotions shared on social media to be strategic, however, the possible disconnect between internal and external states does not affect inference: the external state, the signal, is the piece of information the sender wants received and the only piece of information the receiver has about the sender's true beliefs. Measuring expressed emotion is therefore exactly what we want to measure.

A logistic model is created where the outcome is a 1 if the variable contains the emotion of interest and a 0 otherwise. Table 2 shows the percentage of tweet words that are positive, align with the positive emotions joy or trust, are negative, or align with the negative emotions anger, disgust, and fear. The results show a statistically significant increase in positive words and emotions but no statistically significant change for anger and disgust. The increase in negative words is half as large as the increase in positive ones, and the increase in fear is the smallest of all the statistically significant changes.

Topics

To better understand the pathways affecting changing sentiment and emotion, we build a supervised topic model. Specifically, we fine-tune a Bidirectional Encoder Representations from Transformers (BERT) model, 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

Table 2: Number of Words with the Following Emotions

	Dependent variable:						
	Pos. Words	Joy Words	Trust Words	Neg. Words	Anger Words	Disgust Words	Fear words
	(1)	(2)	(3)	(4)	(5)	(6)	(7)
Post-Siege	−.0696***	−.0539***	−.0641***	−.0455**	−.0262	−.0318*	−.0366**
	(.0195)	(.0184)	(.0191)	(.0188)	(.0164)	(.0170)	(.0178)
Aleppo	−.1120**	−.0905*	−.1136**	−.0458	.0338	−.0196	−.0728
	(.0508)	(.0481)	(.0499)	(.0492)	(.0427)	(.0445)	(.0464)
Words	.0318***	.0240***	.0306***	.0320***	.0232***	.0238***	.0269***
	(.0014)	(.0013)	(.0013)	(.0013)	(.0011)	(.0012)	(.0012)
Post-Siege*Aleppo	.2127**	.2158***	.2283***	.0963	−.0262	.0663	.1095
	(.0836)	(.0792)	(.0822)	(.0809)	(.0702)	(.0732)	(.0764)
Intercept	.1123***	.0226	.0474**	.0074	−.0668***	−.0455**	−.0360*
	(.0223)	(.0211)	(.0219)	(.0215)	(.0187)	(.0195)	(.0203)
Observations	2,247	2,247	2,247	2,247	2,247	2,247	2,247
R^2	.2024	.1405	.1959	.2121	.1589	.1543	.1755
Adjusted R^2	.2010	.1390	.1945	.2106	.1574	.1528	.1741
Note:							*p<0.1; **p<0.05; ***p<0.01

model that can then be fine-tuned (customized) to a specific task. Here, the customization consists of providing labeled training data, tweets, to generate a similar but new model.

We first translated 2,000 tweets in our sample into English and label them for being about one of ten categories: pro-Assad, anti-Assad, military, Russia, Aleppo, an extremist group, an opposition group, entertainment, religion, or anything else. The BERT is then fine-tuned using the original Arabic of the labeled tweets, and one tuned model is created for each topic. Once tuned, the resulting topic model is applied to the Arabic tweets in the full sample.

This work is currently ongoing. The BERT has been trained on the 10 labels and is being applied to the corpus. We expect to have these results by May 1, 2022.

Account Characteristics

Table 3 shows an initial analysis of changes in signalling behavior,the change in the number of tweets and users per day after the siege's end in Aleppo. There are both fewer users and

tweets, though the result is more precise for the decrease in the number of users. Combined with Table 1, these results suggest cross-cutting effects: many users provide a negative signal by using Twitter less frequently, but those that remain send positive signals (increased positive emotions).

Table 3: Change in Tweets and Users per Day

	Users	Tweets
	(1)	(2)
Post-Siege	.8508***	4.0141*
	(.2972)	(2.1523)
Aleppo	−4.5050***	−28.3176***
	(.3136)	(2.2712)
Post-Siege*Aleppo	−1.6970***	−5.4372*
	(.4419)	(3.2004)
Intercept	6.7742***	31.5484***
	(.2118)	(1.5339)
N	115	115
R^2	.2021	.2370
Adjusted R^2	.2006	.2356

*p < .1; **p < .05; ***p < .01

Next, we investigate whether user composition changes. To measure signalling from user behavior, we analyze account-level characteristics. Account popularity is approximated with their number of followers and friends (how many accounts the account follows). Activity on Twitter is measured using the average number of tweets per day, favourited tweets per day (a favourited tweet is one an account has marked so that Twitter saves it for easy retrieval), and retweets. Account age in days is also recorded in order to understand if new accounts join Twitter at different rates before or after the siege.

First, we identify users who tweet within 31 days before the siege's end but do not within 31 days after and a different subset of users who tweet after but not before. We then compare these two groups' production of positive emotions to users who tweet before and after the siege's end. The group which stops tweeting has a slightly higher ratio of positive to negative

words ($p = .06118$) than the persistent tweeters, but otherwise there is no difference across the three groups. There is also no statistically significant difference in emotion between the exclusive post-siege users and the exclusively pre-siege ones, though the pre-siege users had slightly more positive emotions ($p > .3174$).

Second, we analyze account characteristics before and after the siege. For this analysis, only one tweet per user per day is kept, and mean and median quantities of various account characteristics are tracked. Table 4 shows the results for the median quantities; results do not change for estimates using averages, but outcomes on social media are right-skewed, so the median provides a more accurate understanding of behavior. Models one and two examine account popularity; three and four look at new accounts; and five and six look at intensity of use. The only statistically significant treatment result is on median number of followers: accounts from Aleppo after the siege have more followers than before. The accounts do not follow more accounts, are not younger or older, and do not favorite or tweet at higher rates.

Table 4: Change in Account Characteristics

	Followers Median	Friends Median	Account Age Median	Users New	Favorite Rate Median	Tweets per User
	(1)	(2)	(3)	(4)	(5)	(6)
Post-Siege	79.9788	6.4219	206.8115**	.2359*	−.2189	.3963
	(81.3086)	(46.5812)	(102.3960)	(.1256)	(1.7418)	(2.9341)
Aleppo	−821.4355***	75.6774	176.0323*	−2.3871***	3.1262*	4.5833
	(81.9514)	(46.9494)	(103.2055)	(.1266)	(1.7556)	(2.9573)
Post-Siege*Aleppo	235.5292**	4.8382	−108.7823	−.2379	−1.5797	.8751
	(114.9878)	(65.8758)	(144.8098)	(.1776)	(2.4633)	(4.1494)
Intercept	1,511.6770***	225.5000***	2,627.5320***	2.4516***	2.6120**	12.3740***
	(57.9484)	(33.1983)	(72.9773)	(.0895)	(1.2414)	(2.0911)
N	126	126	126	126	126	126
R^2	.5750	.0447	.0599	.8679	.0367	.0475
Adjusted R^2	.5646	.0212	.0368	.8646	.0130	.0240

*p < .1; **p < .05; ***p < .01

Robustness

To verify the results we now turn to a series of robustness checks. Table A3 in Section S3 shows that neither user nor district fixed effects change results.

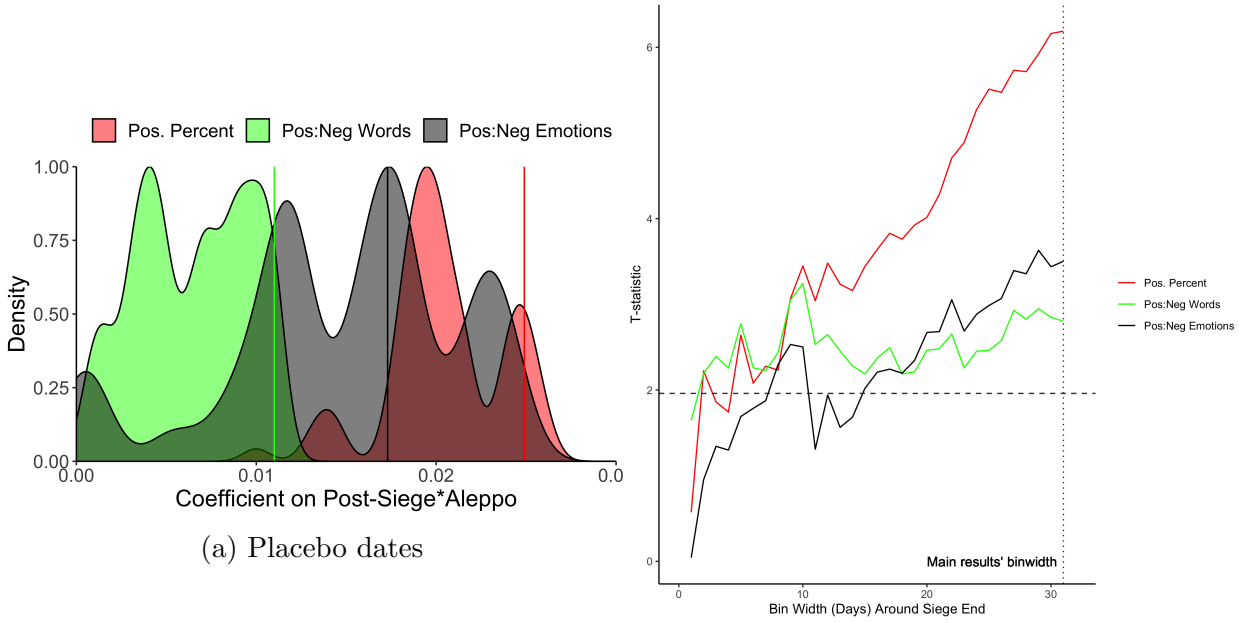
We then conduct a placebo test of 1,000 randomly selected start dates within 30 days of

December 15th and compare the treatment coefficients for each dependent variable in Table 1 to the distribution of 1,000 placebo coefficients. The coefficient for the percent of words that are positive is in the 89.9^{th}th percentile and that for the ratio of positive to negative words is in the 96.1^{st} percentile; the ratio of emotions is in the 62^{nd} percentile, which is not surprising given the difficulty of measuring emotions in text. Figure 4a shows this result; each vertical line corresponds to the treatment coefficient from Table 1.

In addition, we vary the bandwidth from the original model's 31 pre- and post-treatment days. Starting with a bandwidth of 1 day on each side of the treatment reveals if the siege end's effects are immediate. Figure 4b shows the result of this test for each dependent variable. The end of the siege immediately causes an increase in positive emotions, but not until eight days post-siege are the three dependent variables statistically significant. The percent of positive words per tweet and the ratio of positive and negative words attain statistical significance on the second day and are both consistently significant by the fourth day. The significance as the bandwidth widens out to 31 days. In results not shown, the percent of positive words and the ration of positive and negative emotions maintain their significance for at least six months, the widest bandwidth we analyze, while the ratio of positive to negative words oscillates around $t = 1.96$ starting on the 49^{th} post-siege day.

Since Twitter users decide to geotag per tweet, it could be the case that tweets in the original dataset are only geotagged when they contain a positive signal. To check for this possibility, we downloaded the tweet history of the accounts that still exist or public and repeat the regressions in Table 1, counting any tweet as being from Aleppo if its author was in Aleppo in the streamed data. The results are shown in Table 5. The end of the siege is now only statistically significant for the percent of words that are positive. To investigate these differing results, we compare the emotions of users in the streamed sample who are no longer available to the accounts still available. There is no difference in the use of positive words or the ratio of positive to negative words, though users who are no longer available do have a statistically higher amount of positive to negative emotions ($p = .0255$).

Figure 4: Results are Robust to Varying Treatment Date and Window



(a) Placebo dates

(b) Bandwidth

Note: Figure 4a shows the distribution of the treatment coefficient for the three dependent variables evaluated in Table 1. The treatment day is one of any date thirty-one days before or after December 15, 2016. Each density plot is of the distribution of coefficients from these new samples, and their vertical line corresponds to the coefficient from Table 1. Figure 4b varies the bandwidth around the treatment date from one to thirty-one days. The coefficients from Table 1 are at the 96.1, 89.9, and 62 percentiles for the placebo dates and are significant very soon after the siege's end.

Table 5: User Panel

	Positive Words (Percentage)	Positive Words/ Negative Words	Positive Emotions / Negative Emotions
Post-Siege	$-.0043^{***}$	$.0011$	$.0070^{***}$
	$(.0016)$	$(.0016)$	$(.0020)$
Aleppo	$.0323^{***}$	$.0241^{***}$	$.0300^{***}$
	$(.0015)$	$(.0015)$	$(.0018)$
Word	$-.0011^{***}$	$-.0013^{***}$	$-.0017^{***}$
	$(.0001)$	$(.0001)$	$(.0001)$
Post-Siege*Aleppo	$.0082^{***}$	$.0026$	$-.0035$
	$(.0021)$	$(.0020)$	$(.0025)$
Intercept	$.0775^{***}$	1.0373^{***}	1.0288^{***}
	$(.0016)$	$(.0016)$	$(.0020)$
N	42,928	42,928	42,928
Adjusted R^2	.0328	.0204	.0181

$^{*}p < .1$; $^{**}p < .05$; $^{***}p < .01$

Conclusion

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 conflicts. Much progress has been made in understanding the ways in which ICTs 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. In this paper we build on extensive work on the logic of violence in civil war that highlights the importance of civilian agency and strategy when confronted with prolonged violence. We argue that because civilians oftentimes purposefully and strategically signal political loyalties to ensure their survival, social media is likely to be used in similar ways. We assumed that civilians using social media in the context of ongoing civil conflict are likely to be well aware the risks of being monitored, which suggests that content, including sentiment and emotions, shared online are likely to reflect strategic decision-making.

We test our expectation by analyzing the ways in which civilians adapt their twitter activity in the aftermath of one of the most extreme forms of territorial contestation, a siege. We match Twitter accounts in Aleppo to similar users in other parts of Syria and examine changes in sentiment following the end of the siege using difference-in-difference regression. The results suggest that continuously active accounts in Aleppo posted significantly more positive content in the aftermath of siege. The results also suggest that the overall sentiment of Twitter users continuously based in Aleppo turned more positive in the aftermath of the siege, both when compared to themselves, and when compared to the general trend of similar accounts in the rest of Syria, but that this changes only last for a few weeks.

While these results are preliminary and require further investigation, they offer first support for our theoretical expectation that civilian behavior on social media is likely to be influenced by local level changes in conflict dynamics. Shifts in territorial control at the local level significantly change civilians' perceived and actual security situation, and oftentimes requires adaptation in how, where, and when political loyalties should be revealed. We hope

our project will help us better understand how such strategic decisions manifest on social media, and in what ways such changes on social media may bias or distort the ways in which local conflict dynamics are understood and portrayed more broadly.

References

Alasaad, Samer. 2013. "War diseases revealed by the social media: massive leishmaniasis outbreak in the Syrian Spring." *Parasites & vectors* 6(1):1–3.

Atlantic Council. 2017. "Breaking Aleppo." *The Atlantic Council of the United States* pp. 1–70.

Barberá, Pablo, Ning Wang, Richard Bonneau, John T. Jost, Jonathan Nagler, Joshua Tucker and Sandra González-Bailón. 2015. "The Critical Periphery in the Growth of Social Protests." *PloS ONE* 10(11):1–15.
URL: *https://doi.org/10.1371/journal.pone.0143611*

Barberá, Pablo and Thomas Zeitzoff. 2018. "The New Public Address System: Why Do World Leaders Adopt Social Media?" *International Studies Quarterly* 62(1):121–130.

Barbera, Pablo and Zachary C Steinert-Threlkeld. 2020. How to Use Social Media Data for Political Science Research. In *The SAGE Handbook of Research Methods in Political Science and International Relations*. London: SAGE Publications Ltd chapter 23, pp. 404–424.

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*

Beath, Andrew, Fotini Christia and Ruben Enikolopov. 2012. *Winning Hearts and Minds through Development? Evidence from a Field Experiment in Afghanistan*. Policy Research Working Papers The World Bank.
URL: *http://elibrary.worldbank.org/doi/book/10.1596/1813-9450-6129*

Beauchamp, Nicholas. 2019. "Predicting and Interpolating State-Level Polls Using Twitter Textual Data 00 State-level public." *American Journal of Political Science* 61(2):490–503.

Bennett, W Lance and Alexandra Segerberg. 2013. *The Logic of Connective Action*. Number June 2013 Cambridge: Cambridge University Press.

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

Brym, Robert, Melissa Godbout, Andreas Hoffbauer, Gabe Menard and Tony Huiquan Zhang. 2014. "Social media in the 2011 Egyptian uprising." *The British journal of Sociology* 65(2):266–292.

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.

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.
URL: *https://www.tandfonline.com/doi/full/10.1080/1060586X.2019.1701879 https://doi.org/10.1080/1060586X.2019.1701879*

Enikolopov, Ruben, Alexey Makarin and Maria Petrova. 2016. "Social media and protest participation: Evidence from russia.".

Eriksson, Moa. 2018. "Pizza, beer and kittens: Negotiating cultural trauma discourses on Twitter in the wake of the 2017 Stockholm attack." *new media & society* 20(11):3980–3996.

Freedom House. 2015. "Syria." *Freedom on the Net 2015* .
URL: *https://freedomhouse.org/sites/default/files/resources/FOTN 2015$_S$yria.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.".

Gohdes, Anita R. 2015. "Pulling the plug: Network disruptions and violence in civil conflict." *Journal of Peace Research* 52(3):352–367.

Gohdes, Anita R. 2020. "Repression Technology: Internet Accessibility and State Violence." *American Journal of Political Science* .
URL: *https://onlinelibrary.wiley.com/doi/abs/10.1111/ajps.12509*

Gunning, Jeroen and Ilan Zvi Baron. 2013. *Why Occupy a Square: People, Protests and Movements in the Egyptian Revolution*. London: Hurst Publishers.

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.

Jasper, James M. 2006. Motivation and Emotion. In *The Oxford Handbook of Contextual Political Analysis*, ed. Robert E Goodin and Charles Tilly. Number January 2022 pp. 1–16.

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.

Jones, Nickolas M, Sean P Wojcik, Josiah Sweeting and Roxane Cohen Silver. 2016. "Tweeting negative emotion: An investigation of Twitter data in the aftermath of violence on college campuses." *Psychological Methods* 21(4):526–541.

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.

King, Gary, Jennifer Pan and Margaret E Roberts. 2017. "How the Chinese Government Fabricates Social Media Posts for Strategic Distraction, Not Engaged Argument." *American Political Science Review* 111(3):484–501.

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

Kwak, Haewoon, Changhyun Lee, Hosung Park and Sue Moon. 2010. What is Twitter, a Social Network or a News Media ? In *International World Wide Conference*. pp. 591–600.

Larson, Jennifer M, Jonathan Nagler, Jonathan Ronen and Joshua A Tucker. 2019. "Social Networks and Protest Participation: Evidence from 130 Million Twitter Users." *American Journal of Political Science* 63(3):690–705.

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.
URL: *https://doi.org/10.1080/10584609.2019.1661889*

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.
URL: *http://journals.cambridge.org/article*$_S$0003055413000403

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.

Lynch, Marc, Deen Freelon and Sean Aday. 2016. How Social Media Undermines Transitions to Democracy. Technical report PeaceTech Lab.

Manuel, José and Delgado Valdes. 2015. Psychological Effects of Urban Crime Gleaned from Social Media. In *Proceedings of the Ninth International Conference on Web and Social Media*. pp. 598–601.

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

Mohammad, Saif M and Peter D Turney. 2013. "Crowdsourcing a word-emotion association lexicon." *Computational Intelligence* 29(3):436–465.

Monroy-Hernández, Andrés, Scott Counts, Danah Boyd, Emre Kiciman and Munmun De Choudhury. 2013. The New War Correspondents: The Rise of Civic Media Curation in Urban Warfare. In *Proceedings of the 2013 conference on Computer Supported Cooperative Work*. pp. 1443–1452.

Mooijman, Marlon, Joe Hoover, Ying Lin, Heng Ji and Morteza Dehghani. 2018. "Moralization in social networks and the emergence of violence during protests." *Nature Human Behaviour* 2(June):389–396.
URL: *http://dx.doi.org/10.1038/s41562-018-0353-0*

Moss, Dana M. 2018. "The ties that bind: Internet communication technologies, networked authoritarianism, and 'voice' in the Syrian diaspora." *Globalizations* 15(2):265–282.
URL: *https://doi.org/10.1080/14747731.2016.1263079*

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.

Najjar, Abeer. 2010. "Othering the Self: Palestinians Narrating the War on Gaza in the Social Media." *Journal of Middle East Media* 6(1):1–30.

Oklobdzija, Stan. 2018. Dark Parties: Citizens United, Independent-Expenditure Networks and the Evolution of Political Parties. In *Political Networks Workshops & Conference*.

Pan, Jennifer and Alexandra Siegel. 2020. "How Saudi Crackdowns Fail to Silence Online Dissent." *American Political Science Review* 114(1):109–125.
URL: *https://www.cambridge.org/core/product/identifier/S0003055419000650/type/journal*$_{a}$*rticle*

Pearlman, Wendy. 2013. "Emotions and the Microfoundations of the Arab Uprisings." *Perspectives on Politics* 11(2):387–409.

Pearlman, Wendy. 2020. "Mobilizing From Scratch: Large-Scale Collective Action Without Preexisting Organization in the Syrian Uprising." *Comparative Political Studies* p. 001041402091228.
URL: *http://journals.sagepub.com/doi/10.1177/0010414020912281*

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.

Price, Megan, Anita Gohdes and Patrick Ball. 2015. "Documents of war: Understanding the Syrian conflict." *Significance* 12(2):14–19.

Romero, Daniel M, Brendan Meeder and Jon Kleinberg. 2011. Differences in the Mechanics of Information Diffusion Across Topics: Idioms, Political Hashtags, and Complex Contagion on Twitter. In *Proceedings of the 20th International Conference on World Wide Web*. ACM pp. 695–704.

Saha, Koustuv and Munmun De Choudhury. 2017. Modeling Stress with Social Media Around Incidents of Gun Violence on College Campuses. In *Proceedings of the ACM on Human-Computer Interaction*. Vol. 1 pp. 1–27.

Salameh, Mohammad, Saif M Mohammad and Svetlana Kiritchenko. 2015. Sentiment after translation: A case-study on Arabic social media posts. In *Human Language Technologies: The 2015 Conference of the North American Chapter of the Association for Computational Linguistics*. pp. 767–777.

Sanovich, Sergey, Denis Stukal and Joshua A. Tucker. 2018. "Turning the virtual tables: Government strategies for addressing online opposition with an application to Russia.".

Savage, Saiph and Andres Monroy-Hernandez. 2015. Participatory Militias: An Analysis of an Armed Movement's Online Audience. In *Proceedings of the 18th ACM Conference on Computer Supported Cooperative Work & Social Computing*. pp. 724–733.

Schubiger, Livia Isabella. N.d. "State Violence and Wartime Civilian Agency: Evidence from Peru." *Journal of Politics*. Forthcoming.

Shapiro, Jacob N and David A Siegel. 2015. "Coordination and security: How mobile communications affect insurgency." *Journal of Peace Research* 52(3):1–11.

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

Sobolev, Anton, Jungseock Joo, Keith Chen and Zachary C Steinert-Threlkeld. 2020. "News and Geolocated Social Media Accurately Measure Protest Size.".

Spaiser, Viktoria, Thomas Chadefaux, Karsten Donnay, Fabian Russmann and Dirk Helbing. 2017. "Communication power struggles on social media: A case study of the 2011–12 Russian protests." *Journal of Information Technology & Politics* 14(2):132–153.
URL: *http://dx.doi.org/10.1080/19331681.2017.1308288*

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

Steinert-Threlkeld, Zachary C. 2017. "Spontaneous collective action: Peripheral mobilization during the arab spring." *American Political Science Review* 111(2):379–403.

Steinert-Threlkeld, Zachary C. 2018. *Twitter as data*. Cambridge University Press.

Stukal, Denis, Sergey Sanovich, Richard Bonneau and Joshua A Tucker. 2019. "Social Media Bots for Autocrats : How Pro-Government Bots Fight Opposition in Russia." *Working Paper* pp. 1–41.

Sutton, Jonathan, Charles R Butcher and Isak Svensson. 2014. "Explaining political jiu-jitsu: Institution-building and the outcomes of regime violence against unarmed protests." *Journal of Peace Research* 51(5):559–573.

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.

Tucker, Joshua A. 2019. "Who Leads? Who Follows? Measuring Issue Attention and Agenda Setting by Legislators and the Mass Public Using Social Media Data." *American Political Science Review* 113(4):883–901.

Tufekci, Zeynep. 2017. *Twitter and Tear Gas: The Power and Fragility of Networked Protest*. New Haven: Yale University Press.

Tufekci, Zeynep and Christopher Wilson. 2012. "Social Media and the Decision to Participate in Political Protest: Observations From Tahrir Square." *Journal of Communication* 62(2):363–379.

United States. Department of the Army. and United States. 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, Nicholas A, Ted Brader, Eric W Groenendyk, Krysha Gregorowicz and Vincent L Hutchings. 2011. "Election Night's Alright for Fighting: The Role of Emotions in Political Participation." *Journal of Politics* 73(1):156–170.

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

Weidmann, Nils B and Espen Geelmuyden Rod. 2018. *The Internet and Political Protest in Autocracies*. Oxford University Press.

Williams, Lisa A and David DeSteno. 2008. "Pride and Perseverance: The Motivational Role of Pride." *Journal of Personality and Social Psychology* 94(6):1007–1017.

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.

Young, Lauren E. 2019. "The Psychology of State Repression: Fear and Dissent Decisions in Zimbabwe." *American Political Science Review* 113(1):140–155.

Zeitzoff, Thomas. 2011. "Using Social Media to Measure Conflict Dynamics: An Application to the 2008-2009 Gaza Conflict." *Journal of Conflict Resolution* 55(6):938–969.
URL: *http://jcr.sagepub.com/content/55/6/938.abstract*

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

Zeitzoff, Thomas. 2018. "Does Social Media Influence Conflict? Evidence from the 2012 Gaza Conflict." *Journal of Conflict Resolution* 62(1):29–63.

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.

Appendix for

"Civilian Behavior on Social Media During Civil War"

S1 Account Activity

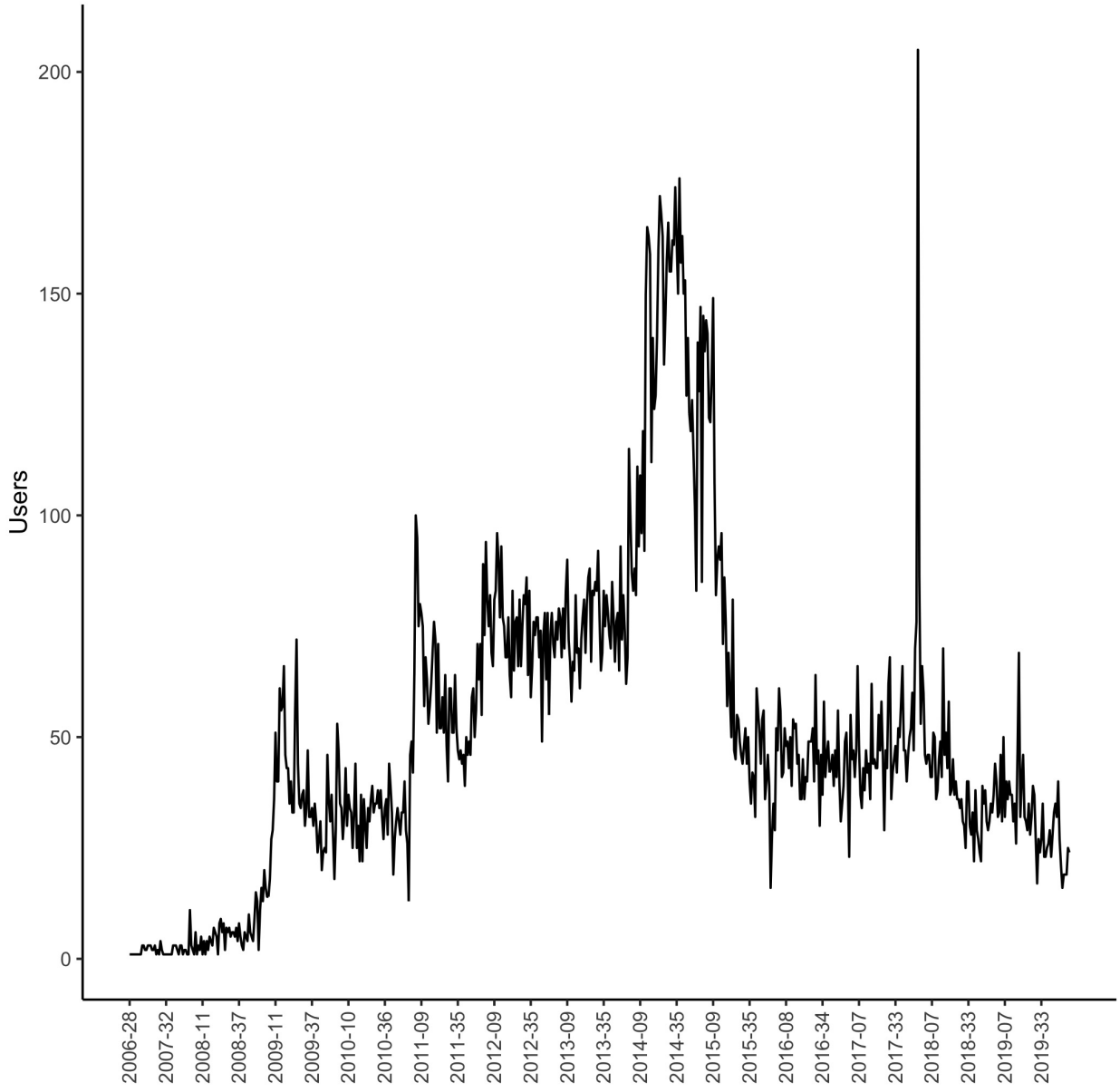


Figure A1: New Users Consistently Join Twitter in Syria

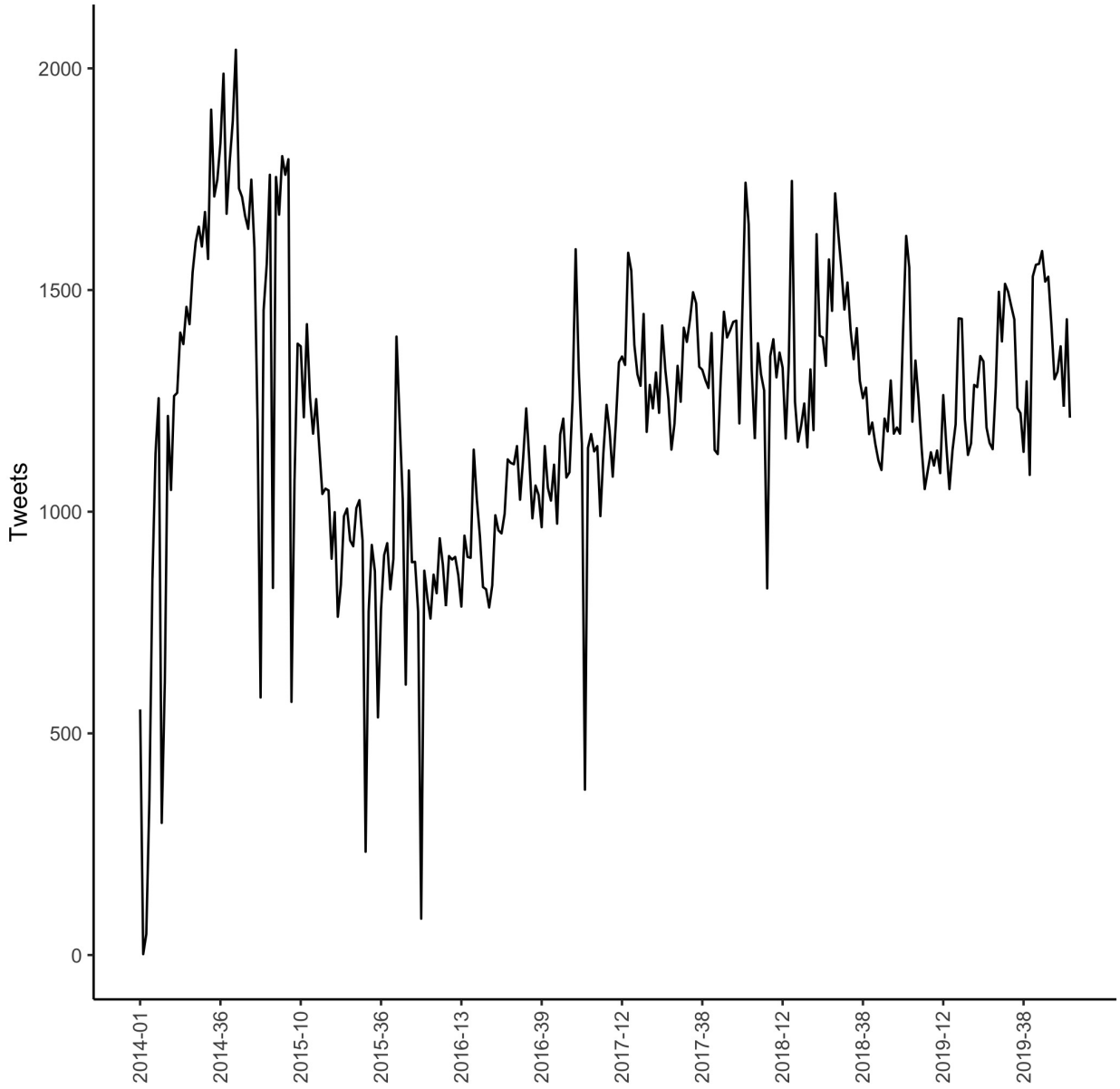


Figure A2: Users Consistently Tweet in Syria

S2 Additional Regressions

Table A1: Change in Emotion Probability, OLS

	Dependent variable: Tweet Contains the Emotion						
	Pos.	Joy	Trust	Negative	Anger	Disgust	Fear
	(1)	(2)	(3)	(4)	(5)	(6)	(7)
Post-Siege	−.0724***	−.0500***	−.0603***	−.0575***	−.0067	−.0324**	−.0299*
	(.0177)	(.0159)	(.0170)	(.0167)	(.0136)	(.0146)	(.0154)
Aleppo	−.1105***	−.0437**	−.0712***	−.0717***	−.0178	−.0688***	−.0385**
	(.0220)	(.0197)	(.0211)	(.0208)	(.0169)	(.0182)	(.0191)
Words	.0289***	.0216***	.0282***	.0300***	.0210***	.0205***	.0231***
	(.0011)	(.0010)	(.0011)	(.0011)	(.0009)	(.0009)	(.0010)
Post-Siege*Aleppo	.1521***	.0782***	.1153***	.0728**	.0076	.0392	.0439
	(.0309)	(.0277)	(.0297)	(.0292)	(.0237)	(.0256)	(.0269)
Constant	.1007***	.0090	.0258	.0053	−.0805***	−.0196	−.0289*
	(.0189)	(.0170)	(.0181)	(.0179)	(.0145)	(.0156)	(.0165)
Observations	3,890	3,890	3,890	3,890	3,890	3,890	3,890
Adjusted R^2	.1546	.1111	.1575	.1751	.1333	.1142	.1278

Note: *p<0.1; **p<0.05; ***p<0.01

Table A2: Change in Emotion Probability, Logistic Model

	DV: Tweet Contains the Emotion						
	Pos.	Joy	Trust	Neg.	Anger	Disgust	Fear
	(1)	(2)	(3)	(4)	(5)	(6)	(7)
Post-Siege	−.3490***	−.3021***	−.3164***	−.3070***	−.0392	−.2105**	−.1847*
	(.0870)	(.0980)	(.0912)	(.0922)	(.1149)	(.1040)	(.1006)
Aleppo	−.5340***	−.2447**	−.3629***	−.3713***	−.1149	−.4798***	−.2260*
	(.1098)	(.1212)	(.1141)	(.1155)	(.1436)	(.1361)	(.1258)
Words	.1312***	.1175***	.1352***	.1463***	.1507***	.1287***	.1312***
	(.0059)	(.0061)	(.0060)	(.0062)	(.0073)	(.0066)	(.0064)
Post-Siege*Aleppo	.7407***	.4617***	.6022***	.3896**	.0362	.2778	.2690
	(.1524)	(.1687)	(.1583)	(.1614)	(.1991)	(.1885)	(.1749)
Intercept	−1.8009***	−2.5372***	−2.2583***	−2.3985***	−3.6912***	−2.9615***	−2.8530***
	(.0956)	(.1110)	(.1025)	(.1047)	(.1430)	(.1232)	(.1175)
Observations	3,890	3,890	3,890	3,890	3,890	3,890	3,890
Log Likelihood	−2,302.2430	−1,937.7260	−2,152.5500	−2,103.1250	−1,490.3210	−1,695.0420	−1,843.9520
Akaike Inf. Crit.	4,614.4860	3,885.4510	4,315.0990	4,216.2500	2,990.6420	3,400.0840	3,697.9040

Note: *p<0.1; **p<0.05; ***p<0.01

S3 Additional Robustness Checks

Table A3: Main Results Are Robust to User and District Fixed Effects

	perc.pos	pos.neg.ratio	pos.emotion.ratio
	(1)	(2)	(3)
Post-Siege	$-.0119^{***}$	$-.0084^{**}$	$-.0080$
	(.0043)	(.0042)	(.0056)
Aleppo	$-.0720$	$-.0359$	.0391
	(.0849)	(.0836)	(.1120)
Words	$.0021^{***}$	.0005	$-.0004$
	(.0003)	(.0003)	(.0004)
Post-Siege*Aleppo	$.0804^{***}$	$.0585^{***}$	$.0495^{*}$
	(.0203)	(.0200)	(.0268)
Intercept	.0600	1.0041^{***}	1.0201^{***}
	(.1040)	(.1024)	(.1371)
N	2,247	2,247	2,247
User FE	Y	Y	Y
District FE	Y	Y	Y
R^2	.1075	.0315	.0445
Adjusted R^2	.0855	.0076	.0210

$^{*}p < .1$; $^{**}p < .05$; $^{***}p < .01$