online dissent or criticism, collecting and analyzing social media data requires special attention to protect subjects, especially when making data available for replication following publication. For the analysis phase, data should be stored on encrypted and password protected computers, and the account names and account content produced by users should be stored in separate files.[107] Upon publication, researchers should make available the code used to query a given dataset either through an API or other method, analysis code for producing an aggregate dataset, and aggregate data for deriving any statistical results, rather than a full dataset of raw social media content. Researchers should also be careful when displaying example social media posts in their research —especially those that contain politically sensitive content—that they do not supply any identifying information.

Along these lines, when conducting surveys on Facebook, researchers can take care to ensure that their participants are not providing any trace data or identifiable information. For example, once a Facebook user clicks on an ad for a survey, they can be redirected to Qualtrics so that researchers cannot connect their responses to their Facebook accounts. Researchers should also disable Qualtrics tracking of respondents IP addresses to insure that information is not inadvertently collected about participants.

Finally, given how quickly the online sphere evolves, studying social media and politics requires regularly updated descriptive research to understand how conditions are shifting.[108] For example, platform use among a given population will likely change over relatively short time horizons. Findings about how diverse actors are using social media, or phenomena like the spread of disinformation or extremist content, for example, may therefore shift rapidly. Recognizing this, there is a great deal off value to designing projects that allow for scalable data collection so that researchers can continue to track particular online phenomena over time. Along these lines, it is also crucial that researchers using social media data regularly reconsider the ethics of their studies as contexts shift, working to ensure the safety and privacy of users whose data they analyze.

Conclusion

Not only can the real-time and networked structure of social media data provide insights about political behavior in the Arab World, but the use of these tools by diverse actors is also politically consequential in and of itself. Like any research approach, using social media data to study politics in the Arab World is not without challenges and limitations, but it nonetheless can be a valuable resource—particularly for scholars studying politically sensitive topics among hard to reach populations or well-known actors and groups. As computational social science approaches to collecting and analyzing data become increasingly accessible, they provide researchers with another set of tools that can be used on their own or integrated with traditional data sources—including survey data, event data, qualitative analysis, and ethnographic fieldwork—to improve our understanding of politics in the region.

QUANTITATIVE TEXT ANALYSIS OF ARABIC NEWS MEDIA

By Ala' Alrababa'h, Stanford University

Middle East scholars have long relied on Arabic news media to understand the priorities of Arab publics and regimes. In 1962, Palestinian American scholar Ibrahim Abu Lughod conducted one of the earliest quantitative analyses of newspapers from seven Arab countries, focusing on coverage of international events and major external powers, such as the United States and the Soviet Union. He claimed that this analysis of Arabic media shed light on "the degree to which the reading public is being oriented toward the outside" and that it could inform scholars "about the judgments and values that are being formed."[109] Abu Lughod relied on the tools available at the time, so he calculated the square centimeters of each newspaper dedicated to certain major powers and coded articles on the front pages by country, topic, and sentiment. Given the difficulty of

conducting a quantitative analysis with the available tools at the time, Abu Lughod limited his analysis to two weeks of newspaper coverage.

Recent developments in quantitative text analysis allow scholars to improve on such analyses of Arabic media. While these methods have been widely used in the analysis of social media in the Middle East[110] and authoritarian media outside of the Middle East,[111] they have not yet been widely applied to the analysis of Arabic news media. However, applying these new tools to Arabic news media holds much potential. Not only can Arabic news media act as an important source of data in a region where obtaining high quality data can be difficult, but the news media itself also has major effects on Arab politics.[112] In this article, I discuss ongoing research projects that utilize these tools in the Arabic news media to show their advantages and potential. I then go over some challenges associated with using quantitative text analysis with Arabic news media. Finally, I describe some of the relevant ethical considerations and technical details when using these tools.

Applying quantitative text analysis to Arabic news media

While quantitative text analysis tools have only recently been used by political scientists to analyze Arabic media, several ongoing research projects reflect the potential of these tools. These projects demonstrate how quantitative text analysis can be used for exploratory research into trends in Arabic media, systematic testing of hypotheses, and identifying texts that would be useful for close qualitative reading.

> ***"the ability to process large amounts of text allows us to explore trends in media and understand how regimes and foreign powers use the media to influence regional publics.***

First, the ability to process large amounts of texts can allow scholars to explore general trends in media. In an analysis of regime-controlled media in Syria, Lisa Blaydes and I examine trends in the discussion of topics related to foreign threats and the Syrian regime.[113] While many scholars argue that regimes sometimes initiate conflicts to enhance their domestic standing, diversionary wars are rare because of their high cost. Instead, we show how the Syrian regime used diversionary rhetoric in state media. Using hand-coding of articles, unsupervised-learning methods—in which the computer discovers the topics in the texts with little intervention by the researchers[114]—and qualitative analysis of articles, we demonstrate how Syrian state media has long relied on foreign threats for domestic control and how the sources of foreign threats changed over time. The quantitative analysis of texts allows us to cheaply analyze trends over a decade, using tens of thousands of articles.

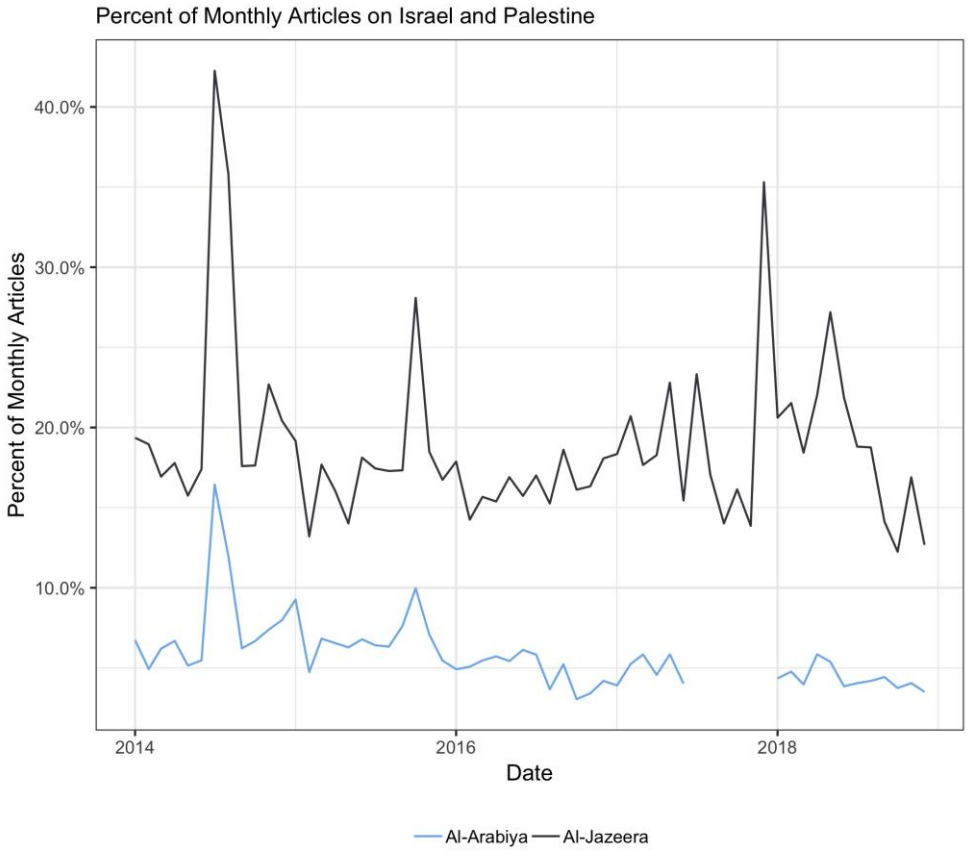
Relatedly, analyzing trends in Arabic media can show the ways in which foreign countries intervene in the region. Metzger and Siegel explore how the Russian-controlled *RT* has attempted to influence global discourse on the Syrian conflict both through their English- and Arabic-language content.[115] They analyze Twitter data to show that *RT* was the most shared Arabic-language news source about Syria, and they conduct an analysis of the content to illustrate how *RT's* coverage was favorable of Russia's intervention. This research demonstrates how foreign countries use Arabic language news to influence the narratives around their intervention in the region. This research also shows the interplay between traditional news media and social media.

In addition to exploratory research, quantitative text analysis can be used for hypothesis testing. Koehler-Derrick, Nielsen, and Romney (2017) ask whether culture or political institutions explain the supply of conspiracy theories in Egypt.[116] They also examine whether conspiracy theories function as tools for messaging or distraction by the regime. Using a combination of supervised learning and case studies, they analyze state and independent media in Egypt. They find that political institutions explain the supply of conspiracy theories and that the promotion of conspiracy theories often acts as a messaging tool from the government to communicate fears and priorities.

Similarly, in a research project about the use of transnational media as public diplomacy tools, I analyze hundreds of thousands of articles from Saudi-owned *Al-Arabiya* and Qatari-owned *Al-Jazeera*, two of the most prominent news channels in the Middle East.[117] I provide an argument about how Qatar has attempted to mobilize Arab publics, especially in rival countries, while Saudi Arabia has attempted to pacify Arab publics to reduce the risk of revolutions. I test this argument by analyzing *Al-Arabiya* and *Al-Jazeera's* coverage of foreign countries, including the rhetoric and topics they employ in an attempt to pacify or mobilize Arab publics.

For instance, one of the particularly mobilizing topics for Arab publics is the Israeli-Palestinian conflict. Figure 1 shows the percent of articles in *Al-Arabiya*'s and *Al-Jazeera*'s Arabic websites that mention the words "Israel" or "Palestine." As can be seen in this figure, *Al-Jazeera* consistently discusses this conflict at a much higher rate than *Al-Arabiya*, suggesting the relative importance of this topic in *Al-Jazeera*.

Figure 1: Percent of monthly articles in Al-Arabiya and Al-Jazeera that mention "Israel" or "Palestine"



Finally, in my dissertation project, I examine how Arab regimes responded to the rise in information technology and independent news media. Before the rise in this technology, regimes state media was the main source of information. While people may not always believe the regime's propaganda, Lisa Wedeen argued that the "regime's power resides in its ability to sustain national fictions, to enforce obedience, to make people say and do what they otherwise would not."[118] But the rise of independent media and the ability to access it cheaply using the Internet posed a challenge to

regime media. I claim that regimes responded to this by using state media to psychologically manipulate their publics, especially by playing into their fears, with the goal of reducing dissent. I test this through an analysis of news media from several Arab countries to examine how coverage varies during peaceful times and during periods of dissent.

In summary, the ability to process large amounts of text can be helpful to political scientists, because it allows us to explore trends in media and understand how regimes and foreign powers use the media, particularly state-owned news outlets, to influence regional publics.

Challenges of quantitative text analysis with Arabic media

While quantitative text analysis has much potential when applied to Arabic media, there are important challenges to consider when obtaining and analyzing this media. For instance, especially when compared to U.S. media, media in the Arab world has been digitized only relatively recently, which temporally limits the scope of using these methods to study Arabic media. In addition, despite improvements over the years, Optical Character Recognition (OCR) tools remain underdeveloped in Arabic and it is often difficult to convert Arabic PDF images to text.

In addition to challenges with obtaining data, challenges also arise when analyzing the data. On the bright side, much of the traditional news media across Arab countries is written using Modern Standard Arabic (MSA), which allows for comparative study of these media. But scholars still need to be careful when analyzing these texts, because local colloquial words still often make it to these papers. For instance, Arabs in several countries use the English word "van" to refer to small transportation buses. This often makes it to Arabic media (فان) but this is also the same word for something mortal or perishable (Other forms of transportation like taxis, or Tok Tok, also often make it to Arabic media).

Relatedly, there are often spelling differences even when using Modern Standard Arabic—for instance, Egyptian media often omits the dots below the letter *ya* (ي) in Arabic when it is at the end of the word. This could make it potentially challenging to conduct a comparative analysis between, say, Egyptian media and Saudi media, because the same words may be spelled differently.

Scholars who conduct text analysis in Arabic should also watch for names, which often have meanings, especially when taken to their roots. For instance, the name Salman (for instance, the Saudi king, King Salman), when stemmed, becomes selm (سلم), which is the same three-letter root for the word "peace." This does not mean that these methods are not useful in Arabic, but it does suggest the importance of thinking carefully about the technical aspects of, for example, stemming the words—possibly including a list of words that the program should *not* stem—validating the results, and conducting close qualitative analysis in addition to any quantitative one.

Another challenge with obtaining and analyzing Arabic media relates to the political environment. Scholars should be aware of censorship, including self-censorship, practiced in many Arab countries. Media, including independent outlets, are often restrained by a series of formal laws that punish incitement and spreading falsehoods (as defined by the governments).[119] This makes media even less representative of the publics' opinions and priorities. Of course this can still provide an opportunity to analyze media to examine the priorities of regimes and what they allow to be published.

This discussion suggests the importance of validating the results of text analysis in Arabic and combining the quantitative analysis with a qualitative reading of the texts. It also suggests the importance of being clear and careful about the goals of Arabic media analysis, keeping in mind the important role of censorship.

Technical tools and ethical considerations

When scraping news websites, it is important to

take into consideration ethical and legal issues. First, intellectual property and copyright laws apply to these websites, so authors should not republish the texts of news articles (at least without explicit consent from the news source). Note that this is tricky, because it could make replication harder. Some researchers suggest that publishing the stemmed document-term matrix may be ethical, but there are no clear standards in the field yet.

A related concern is reading and understanding the terms and conditions of a website before scraping, as some explicitly ban it. In theory, scholars should use the Application Programming Interface (API) when scraping; however, in practice, I am not aware of any news websites in Arabic with APIs. These may be developed by some of these websites in the future.

Finally, it is important to make requests at a reasonable pace. Many Arabic news websites, especially small independent ones, cannot handle a lot of traffic. So if research make too many repeated scraping requests, researchers could unintentionally slow down or even shut down the websites, and this could be interpreted as a Denial of Service attack by the scholar.

There are many tools that make scraping relatively easy in R and Python. In R, *rvest* is a powerful library that makes it easy to scrape many news websites. After downloading the data, scholars often need to manipulate it. Sometimes websites include JavaScript, making it difficult to scrape using *rvest*. The open-sourced tools *Selenium* in Python or *Rselenium* in R can be particularly helpful to deal with these websites, as they allow the researcher to write code that browses the Internet like a user. There are several packages in R that are helpful to process and clean text data, including *tidytext*, *broom*, and *stringr*/*stringi*. As change in trends over time is often important with news media, *lubridate* is a particularly powerful package to manipulate dates. With Arabic texts, Nielsen's *arabicStemR* is particularly helpful for removing stop words and stemming. *Farasa* is another tool that allows parsing of Arabic texts, which also includes identifying parts of speech in Arabic sentences.

Conclusion

Conducting political research on the Middle East can often be challenging. Many regimes do not maintain easily accessible archives, restrict survey work, and sometimes even put the safety of researchers at risk. Arabic media thus offers a valuable source of data to learn about many questions related to regime behavior, intervention by foreign powers, public diplomacy, and political opposition in Arab countries. Yet the role of Arabic media goes beyond providing a source to study these topics. Research has shown that Arabic news media itself plays an important role in influencing regional politics. Quantitative text analysis provides researchers with the tools to use vast amounts of texts in order to study some of these topics and, in particular, the role of news media in politics. While there are important challenges and limitations, ongoing research projects that apply quantitative text analysis to Arabic media demonstrate the potential of these tools.

IDEOLOGICAL SCALING IN A POST-ISLAMIST AGE

By Nate Grubman, Yale University

In recent years, a growing body of scholarship has focused on the interaction of cultural identity and class as potential bases for partisanship in the Arab world. A number of puzzles have emerged: Following the 2010 to 2011 uprisings driven in large part by economic grievances, why did party systems in Egypt and Tunisia revolve more tightly around competing notions of religious and national identity than competing economic orientations?[120] Why did many of the poor turn to Islamist parties rather than Marxist-Leninist or Arab nationalist ones?[121] How can the perceived dominance of a secularist-Islamist cleavage and the popularity of Islamist parties be reconciled with the observation that citizens of Arab countries are concerned with the mundane economic issues that preoccupy other people of the world?[122]

domains." In *Twelfth International AAAI Conference on Web and Social Media*. 2018.
[84] Fisher, Ali. "How jihadist networks maintain a persistent online presence." *Perspectives on terrorism* 9, no. 3 (2015).
[85] See, for example: Berger, J. M. "Tailored online interventions: The Islamic state's recruitment strategy." *CTC Sentinel* 8, no. 10 (2015): 19-23; Siegel, Alexandra A., and Joshua A. Tucker. "The Islamic State's information warfare." *Journal of Language and Politics*17, no. 2 (2018): 258-280.; Ceron, A., Curini, L. and Iacus, S.M., 2019. ISIS at its apogee: the Arabic discourse on Twitter and what we can learn from that about ISIS support and Foreign Fighters. *Sage open*, *9*(1), p.2158244018789229.
[86] Mitts, Tamar. "From isolation to radicalization: anti-Muslim hostility and support for ISIS in the West." *American Political Science Review* 113, no. 1 (2019): 173-194.
[87] Magdy, W., Darwish, K. and Weber, I., 2015. # FailedRevolutions: Using Twitter to study the antecedents of ISIS support. *arXiv preprint arXiv:1503.02401*.
[88] Carlson, Melissa, Laura Jakli, and Katerina Linos. "Rumors and refugees: how government-created information vacuums undermine effective crisis management." *International Studies Quarterly* 62, no. 3 (2018): 671-685.
[89] https://digitalrefuge.berkeley.edu/
[90] Masterson, Daniel, and Mourad, Lama. "The Ethical Challenges of Field Research in the Syrian Refugee Crisis." 2019. APSA MENA Newsletter.
[91] Noman, H., Faris, R. and Kelly, J., 2015. Openness and Restraint: Structure, Discourse, and Contention in Saudi Twitter. *Berkman Center Research Publication*, (2015-16); Salem, Fadi. "The Arab social media report 2017: Social media and the internet of things: Towards data-driven policymaking in the Arab World (Vol. 7)." *Dubai: MBR School of Government* (2017).
[92] https://developer.twitter.com/en/docs.html
[93] https://developer.twitter.com/en/docs/tutorials/consuming-streaming-data.html
[94] https://developer.twitter.com/en/docs/tweets/batch-historical/api-reference/historical-powertrack.html
[95] Freelon, Deen. "Computational research in the post-API age." *Political Communication* 35, no. 4 (2018): 665-668.
[96] https://github.com/twintproject/twint
[97] https://smappnyu.org/research/data-collection-and-analysis-tools/
[98] For an overview of the need for careful decision-making and validation when pre-processing of text, see: Denny, M.J. and Spirling, A., 2018. Text preprocessing for unsupervised learning: Why it matters, when it misleads, and what to do about it. *Political Analysis*, *26*(2), pp.168-189.
[99] https://gephi.org/
[100] https://socialscience.one/our-facebook-partnership
[101] https://www.crowdtangle.com/
[102] Malik, Mashail and Williamson, Scott. "Contesting Narratives of Repression: Experimental Evidence from Sisi's Egypt" Unpublished working paper. 2019.
[103] Pham, Katherine Hoffmann, Rampazzo, Francisco, and Rosenzweig, Leah. 2019. "Social Media Markets for Survey Research in Comparative Contexts: Facebook Users in Kenya." Unpublished Working Paper.
[104] https://developers.google.com/youtube/v3/quickstart/python
[105] Torres, Michelle. "Give me the full picture: Using computer vision to understand visual frames and political communication." URL: http://qssi. psu. edu/new-faces-papers-2018/torres-computer-vision-and-politicalcommunication (2018).
[106] https://developers.facebook.com/docs/instagram-api

[107] For example, there should be one file with the account name and a unique id, and another file with the id and content of the account. After the data analysis is complete, the file with the account names should be permanently deleted
[108] See Munger (2019) for an overview of this debate.

Alrababa'h notes:

[109] Ibrahim Abu Lughod, "International News in the Arabic Press: A Comparative Content Analysis," *Public Opinion Quarterly* 26 no. 4 (1962): 600
[110] See Alexandra Siegel's contribution in this newsletter.
[111] Gary King, Jennifer Pan, and Margaret E. Roberts, "How Censorship in China Allows Government Criticism but Silences Collective Expression," *American Political Science Review* (2013); Anjalie Field et al., "Framing and Agenda-setting in Russian News: a Computational Analysis of Intricate Political Strategies," EMNLP (2018). Jaros, Kyle and Jennifer Pan, "China's Newsmakers: How Media Power is Shifting in the Xi Jinping Era." *The China Quarterly* 233 (2018): 111-136; Arturas Rozenas and Denis Stukal. "How Autocrats Manipulate Economic News: Evidence from Russia's State-Controlled Television." *The Journal of Politics* 81.3 (2019).
[112] For instance, see Marc Lynch, *Voices of the New Arab Public: Iraq, Al-Jazeera, and Middle East Politics Today*. Columbia University Press, 2006 and Marc Lynch, "How the Media Trashed the Transitions," *Journal of Democracy* 26 no. 4 (2015): 90-99.
[113] Ala' Alrababa'h and Lisa Blaydes, "Authoritarian Media and Diversionary Threats: Lessons from Thirty Years of Syrian State Discourse," Working Paper https://blaydes.people.stanford.edu/sites/g/files/sbiybj1961/f/syria.pdf
[114] See Rich Nielsen's contribution in this newsletter for more discussion of unsupervised learning.
[115] Megan Metzger and Alexandra Siegel, "When State-Sponsored Media Goes Viral: Russia's Use of RT to Shape Global Discourse on Syria," Working Paper, https://alexandra-siegel.com/wp-content/uploads/2019/08/syria_RT.pdf
[116] Gabriel Koehler-Derrick, Richard A. Nielsen, David Romney, "Conspiracy Theories in the Egyptian State-Controlled Press," Working Paper, http://aalims.org/uploads/conspiracy_10april2017_AALIMS.pdf
[117] Ala' Alrababa'h, "Using Transnational Media as a Public Diplomacy Tool: Evidence from Al-Arabiya and Al-Jazeera," Working paper.
[118] Lisa Wedeen, "Acting "As If": Symbolic Politics and Social Control in Syria." Comparative Studies in Society and History 40.3 (1998): 5ß19
[119] For example, see Wafa Ben Hassine, "The Crime of Speech: How Arab Governments Use the Law to Silence Expression Online," *Electronic Frontier Foundation*, https://www.eff.org/files/2016/04/28/crime-of-speech.pdf and Matt J. Duffy, "Arab Media Regulations: Identifying Restraints on Freedom of the Press in the Laws of Six Arabian Peninsula Countries," *Berkeley Journal of Middle Eastern & Islamic Law*, 6 (2014): 1:31