

USING SOCIAL MEDIA DATA TO STUDY ARAB POLITICS

By Alexandra A. Siegel, Stanford University

From clerics with millions of online followers and government-sponsored bot armies, to activists organizing and individuals discussing politics, elites and everyday citizens across the Arab World are increasingly using social media tools to achieve their political goals. Taking advantage of the real-time digital footprints that these diverse actors leave online, social media data offers opportunities to measure political behavior in the Arab World and other comparative contexts. Additionally, social media use for political purposes has tangible offline consequences and is itself a political phenomenon of interest. Here I describe how social media data can be used to both study political behavior in the Arab World as well as explore the role that social media is playing in Arab politics. I then lay out a set of resources and tools for collecting and analyzing social media data, and discuss practical limitations and ethical and challenges in using this data for political science research.

Using social media data to study political behavior

In the Arab World, there is high social media penetration and online platforms are widely used by elites and everyday citizens to discuss politics and achieve political goals, making social media data a particularly valuable resource for political scientists.⁷³ In particular, there are several structural affordances of social media data that enable us to measure political behavior. First, because social media use is near constant—with hundreds of millions of users leaving digital traces on online platforms every day by posting, commenting, tweeting, liking, and sharing content—it provides researchers with real-time organic measures of behavior. Second, because of its networked structure, social media data offers measures of mass and elite behavior on the same platforms. Third, social media data enables us to access politically sensitive data from populations that are often difficult to reach in authoritarian

regimes or conflict settings using more traditional data sources.

A growing body of research from the Arab context highlights how these affordances facilitate new studies of mass and elite political behavior in the region. For example, recent work has used social media data to: explore the short-term dynamics of military conflict in Gaza,⁷⁴ develop real-time measure of transnational ideological diffusion across Islamist groups,⁷⁵ and investigate the dynamics of political polarization in post-coup Egypt.⁷⁶

Along these lines, my own work⁷⁷ takes advantage of the real-time and networked structure of social media data to assess when religious and political elites strategically incite sectarian tensions. Using millions of Arabic language tweets to construct measures of elite incitement, the analysis demonstrates that, while Saudi clerics and royal family members spread hostile sectarian rhetoric in the aftermath of foreign episodes of violence, they attempt to rein in this discourse following domestic episodes of violence.

In another recent project,⁷⁸ Jennifer Pan and I examine the political imprisonment of well-known Saudis to provide the first large-scale, systematic study of the effects of repression on online dissent. Analyzing more than 300 million tweets and Google search data from 2010 to 2017 using automated text analysis and crowd-sourced human evaluation of content, the paper tests whether repression has deterrent or backlash effects. We show that, although repression deterred imprisoned Saudis from continuing to dissent online following their releases, it did not suppress dissent overall. Observing repression increased dissent—including criticisms of the ruling family and calls for regime change—among the followers of those who were imprisoned, and drew public attention to arrested Saudis and their causes. By showing the varied effects of repression on online dissent, this work helps elucidate the relationship between repression and dissent in the digital age.

“*access to real-time, networked data facilitates analysis of the microdynamics of conflict and mass-elite interaction*”

As these examples illustrate, access to real-time, networked data facilitates analysis of the microdynamics of conflict and mass-elite interaction. Moreover, given how politically sensitive sectarianism and dissent are in the Gulf, it would have been extremely challenging to collect more traditional data that would provide insights into these political phenomena in Saudi Arabia.

Social media as a political tool

A second strand of literature focuses on the use of social media itself as a political tool in the Arab World. This includes work exploring the role of social media in organizing or sustaining protest, the use of social media by armed and extremist groups, and online disinformation and computational propaganda campaigns by governments and other powerful political actors.

Some of the earliest research using social media data in the Arab World explored the effect of social media on protest dynamics during the Arab Spring period. A wide array of empirical research attempted to shed light on the question of whether social media instigated or facilitated protests, where it was most influential, and how its influence compared with other social, economic, political, or cultural factors. While many early articles argued that social media was the single most important driver of the Arab Spring protests,⁷⁹ later work tended to question the role of social media, highlighting that other factors, including offline networks and legacy media, were more important.⁸⁰ This echoes similar debates in the literature about the role of social media and protest in diverse contexts.⁸¹

After the initial optimism regarding the democratizing power of social media in the Arab World, more recent research has focused on the darker sides of social media use in the region. One strand of this research has focused on detecting

and understanding the role of disinformation in conflict settings, particularly the Syrian civil war, which has been dubbed the first “socially mediated civil conflict.”⁸² It has explored how diverse actors inside and outside of Syria have worked to spread disinformation online from government-funded anti-White Helmets narratives,⁸³ to disinformation campaigns developed by armed and extremist groups about the conflict.⁸⁴

Focusing on extremist groups’ use of social media, following the rise of ISIS, a number of studies used social media data to map the organization’s digital recruitment strategy and the remarkably successful broadcasting of its message on social media across social media platforms in the Arab World and globally.⁸⁵ Other work has used social media data to study the impact of events on radicalization,⁸⁶ as well as to predict the likelihood that individuals become radicalized over time.⁸⁷

Online information in conflict settings can be a matter of life and death, particularly for vulnerable populations. Under conditions of high anxiety and ambiguity, Syrian refugees have relied heavily on social media to access information during their journeys and upon arrival in host countries. In my own work analyzing data from public Facebook pages, I find that refugees rely heavily on unofficial sources of information, fueling the potential for rumors and disinformation. Other recent work in this space shows that frequent policy changes, information dissemination limits, and ad-hoc policy implementation often lead to rumors and disinformation among refugees.⁸⁸ Big data analysis of refugee communications aids our understanding of gaps in their information needs as well as where those populations are most vulnerable.⁸⁹ Using anonymized and aggregated digital trace data can also help researchers avoid some of the ethical challenges that may emerge when conducting research on refugees and other vulnerable populations.⁹⁰ Together, these studies provide valuable insight into how diverse actors—from governments and activists to armed groups and refugees—are using social media to pursue their

political goals.

Collecting and Analyzing Social Media Data

One of the primary advantages of social media data for political science research is that a great deal of data is free and publicly available and can be collected in a scalable manner. Twitter data is most widely used by social scientists due to its ease of collection and extensive metadata. While less popular in the Arab world than Facebook and WhatsApp, the platform is nonetheless widely used to discuss politics, and Gulf countries have some of the highest levels of Twitter penetration in the world.⁹¹

The most common way in which researchers access Twitter data is using application programming interfaces (APIs), which enable users to download data using individual access tokens. Twitter data can be queried through the Rest API,⁹² which allows researchers to search for specific information about users and tweets including user profile metadata, lists of followers and friends, and up to 3200 tweets generated by a given user. This can be done using publicly available statistical packages including `twitterR`, `rtweet`, and `netdemR`.

For collecting real-time Twitter data, researchers can use the Streaming API⁹³ to connect to a “stream” of tweets as they are being published, filtering by keywords, location, or sampling 1 percent of all tweets on Twitter. The R library `streamR` can be effectively used to access the streaming API. Researchers can also access historical Twitter data (non-real-time data) using Gnip’s Historical PowerTrack API,⁹⁴ which offers paid subscriptions to tweets and can be queried with keyword, location, and other metadata filters. Finally, as we enter what Deen Freelon has called a “post-API age,”⁹⁵ many researchers have developed tools to scrape Twitter directly,⁹⁶ avoiding rate limits and obtaining largely unlimited access to historical data.

Twitter data is particularly well suited to time series analysis of changing rhetoric and engagement behavior over time. The text analysis

methods—both supervised and unsupervised—described in Rich Nielsen’s contribution “What Counting Words Teaches us About Middle East Politics” in this newsletter are well suited to categorizing tweets as belonging to particular topics or expressing sentiments. A well-developed set of free tools have also been built for cleaning and analyzing Twitter data.⁹⁷ That said, because tweets are very short, some automated text analysis approaches like topic modeling often do not work particularly well on tweets, and human validation is particularly crucial when evaluating model performance on such short texts. Moreover, textual analysis of social media data from the Arab World requires special care due to combination of Modern Standard Arabic text, text in multiple dialects, transliterated text (Arabizi), and text in English and French, not to mention Internet slang, hashtags, emojis, URLs, and other social media specific symbols. It is therefore particularly important that researchers pre-process their text carefully and transparently.⁹⁸

Additionally, because Twitter’s data structure enables us to measure both *connections* among users (friend-follower networks) and *interactions* among users (retweets, likes, and replies), this data is also particularly well suited to network analysis. In particular, we can identify influential nodes in networks of political discussion, track how information spreads through a network, and measure how closely particular users are tied together in a given network. Tools for network analysis visualization are freely available through R and Python, as well as using Gephi,⁹⁹ an open-source network analysis and visualization software package.

While Facebook is the most popular platform in the Arab World, collecting Facebook data has become increasingly difficult over time as there is currently no Terms of Service compliant way to access Facebook data—including data from public pages. Applications have been developed to scrape public data in violation of Facebook’s terms of service, but they are frequently shut down by the platform. Social Science One’s recent partnership¹⁰⁰ with Facebook has opened the door for academic researchers to obtain limited access

to Facebook data through the Crowdtangle platform,¹⁰¹ which research teams currently must apply to access. Opportunities for research on Facebook moving forward may depend both on the development of these partnerships and continuing debates in the social sciences over the ethics of scraping publicly available data in the post-API age.

Facebook's Ads feature also offers valuable opportunities to safely survey hard to reach populations on politically sensitive topics in the Arab World. For example, recent work has used this feature to conduct a Facebook survey experiment on Egyptian Facebook users evaluating the persuasiveness of competing information provided by a human rights organization and the Egyptian security forces at shaping attitudes toward state-sponsored violence.¹⁰² By enabling researchers to conduct surveys on large numbers of individuals without collecting or supplying identifying information, these online tools have potential for conducting low-cost surveys in the region. Recent work seeking to validate the use of these online surveys in developing countries with traditional survey data to help researchers address concerns about representativeness and reliability is promising,¹⁰³ though more research is needed to validate these tools in the MENA context.

YouTube, an underexplored platform for research on politics both in the Arab World and globally, has a very generous API.¹⁰⁴ Using YouTube's public API, researchers can access all data going back to 2006, including automatically generated transcripts of videos and user comments. Given that political and religious elites, as well as armed groups and extremist actors regularly produce content on YouTube, we can use text analysis tools to explore how content produced by diverse actors changes over time, as well as how everyday Youtubers engage with content produced by particular actors using similar techniques to those described in the discussion of analyzing Twitter data. Similarly, Instagram also offers a treasure trove for researchers, particularly as accessible tools have increasingly been developed for image analysis.¹⁰⁵ The platform is widely used in the Arab

World, by elites and everyday citizens alike, and although its API is increasingly restricted,¹⁰⁶ it is still possible to collect data from public accounts.

Limitations and Challenges

Despite the opportunities that social media data affords, it also brings a unique set of challenges for researchers. Most importantly, social media data is almost by definition not representative. While social media penetration is high in the Arab World, particularly in the Gulf, it is not used uniformly across the region and we, of course, cannot assume that behavior on any given social media platform, or surveys conducted on a particular platform, are representative. This is especially the case given the rise of bots and trolls, which can easily flood or mischaracterize mass behavior online. Moreover, as people in the Arab World are increasingly moving sensitive political conversations to closed groups and private or encrypted messages, publicly available social media data is not necessarily even representative of online discourse more broadly.

However, there are plenty of opportunities to use social media for research that do not require the data to be representative. For example, research using social media data to study the behavior of particular actors—for example known religious or political elites, activists, media outlets, extremist groups, or the engaged followers of any of these actors does not require representative data. Additionally, when researchers are interested in studying online phenomena specifically such as the spread of online hate speech, extremist content, or disinformation, then social media users of public platforms are the population of interest. Finally, when we are conducting research on hard to reach populations or politically sensitive topics, we may still gain valuable non-representative insights from social media users' behavior that nonetheless advance our knowledge of particular political phenomena in the region.

Second, there are important ethical challenges to working with social media data. As governments in the Arab World increasingly criminalize and punish

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 of 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

contain different parts of speech and have an entirely different meaning as a result of how words are connected in Arabic. For example, the sequence *وجيئته* can mean "and his/its side" or "I/She sent him/it [in the direction of]" and can be segmented into entirely different parts of speech.

⁶³ For an overview of text as data methods, see: Justin Grimmer and Brandon M. Stewart, "Text as Data: The Promise and Pitfalls of Automatic Content Analysis Methods for Political Texts," *Political Analysis* 21, no. 3 (2013): 267-97. Supervised methods include the use of human coders and thus rely on the researcher to know *ex-ante* what to code for in the data. Unsupervised methods generate the topics from the data but require researcher decisions about issues such as the number of topics to generate.

⁶⁴ Siegel and Pan's research is discussed in Alexandra Siegel's contribution to this newsletter.

⁶⁵ Lisa Blaydes, Justin Grimmer, and Alison McQueen. "Mirrors for Princes and Sultans: Advice on the Art of Governance in the Medieval Christian and Islamic Worlds," *Journal of Politics* 80, no. 4 (2018): 1150-1167; Jennifer A. London, "Re-imagining the Cambridge School in the Age of Digital Humanities," *Annual Review of Political Science* 19, no. 1 (2016): 351-373; Tamar Mitts, "Terrorism and the Rise of Right-Wing Content in Israeli Books," *International Organization* 73, no. 1 (2019): 203-24.

⁶⁶ Amy Catalinac, "From Pork to Policy: The Rise of Programmatic Campaigning in Japanese Elections," *The Journal of Politics* 78, no. 1 (2016): 1-18.

⁶⁷ Mathilde Emeriau, "Learning to be Unbiased: Evidence from the French Asylum Office," *Working Paper* (2019); Margaret E. Roberts, Brandon M. Stewart, Dustin Tingley, Christopher Lucas, Jetson Leder-Luis, Shana Kushner Gadarian, Bethany Albertson, David G. Rand, "Structural Topic Models for Open-Ended Survey Responses," *American Journal of Political Science* 58, no. 4 (2014): 1064-1082.

⁶⁸ Justin Grimmer, *Representational Style in Congress: What Legislators Say and Why It Matters* (Cambridge: Cambridge University Press, 2013).

⁶⁹ Azusa Katagiri and Eric Min, "The Credibility of Public and Private Signals: A Document-Based Approach," *American Political Science Review* 113, no. 1 (2019): 156-72; Benjamin Liebman, Margaret Roberts, Rachel Stern, and Alice Wang, "Mass Digitization of Chinese Court Decisions: How to Use Text as Data in the Field of Chinese Law," *21st Century China Center Research Paper* no. 2017-01 (2017).

⁷⁰ Another interesting question is how the choice of language in social media corresponds to social class and other individual characteristics? Fred Schaffer's ethnographic study of conceptions of democracy in Senegal suggests that these language choices are closely related to class and have important implications for how people understand and engage with important political concepts like democracy. See: Frederic C. Schaffer, *Democracy in Translation: Understanding Politics in an Unfamiliar Culture* (Ithaca: Cornell University Press, 1998).

⁷¹ Freya Pratty, "Arabic and AI: Why voice-activated tech struggles in the Middle East," *Middle East Eye*, September 10, 2019. Examining Danish, German, Spanish, French and Polish, de Vries, Schoonvelde, and Schumacher make the case for Google Translate. However, I have not found a comparable analysis of the performance of Google Translate for Arabic. See: Erik de Vries, Martijn Schoonvelde, and Gijs Schumacher, "No Longer Lost in Translation: Evidence That Google Translate Works for

Comparative Bag-of-Words Text Applications," *Political Analysis* 26, no. 4 (2018): 417-30.

⁷² It would be interesting to see applications and evaluations of recent advances in language models, such as word embeddings, on Arabic texts. For an application in political science, see: Yaoyao Dai, "Measuring Populism in Context: A Supervised Approach with Word Embedding Models," *Working Paper* (2019).

Siegel notes:

⁷³ 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).

⁷⁴ Zeitzoff, Thomas. "Using social media to measure conflict dynamics: An application to the 2008–2009 Gaza conflict." *Journal of Conflict Resolution* 55, no. 6 (2011): 938-969.

⁷⁵ Kubinec, Robert, and John Owen. "When Groups Fall Apart: Measuring Transnational Polarization with Twitter from the Arab Uprisings." Unpublished Manuscript (2018).

⁷⁶ Weber, Ingmar, Venkata R. Kiran Garimella, and Alaa Batayneh. "Secular vs. islamist polarization in egypt on twitter." In Proceedings of the 2013 IEEE/ACM International Conference on Advances in Social Networks Analysis and Mining, pp. 290-297. ACM, 2013; Lynch, Marc, Deen Freelon, and Sean Aday. "Online clustering, fear and uncertainty in Egypt's transition." *Democratization* 24, no. 6 (2017): 1159-1177; Siegel, Alexandra, Joshua Tucker, Jonathan Nagler, and Richard Bonneau. "Tweeting beyond Tahrir: Ideological diversity and political tolerance in Egyptian twitter networks." Unpublished Manuscript (2019).

⁷⁷ Siegel, Alexandra. *Sectarian Twitter Wars: Sunni-Shia Conflict and Cooperation in the Digital Age*. Vol. 20. Carnegie Endowment for International Peace, 2015.; Siegel, Alexandra, Joshua Tucker, Jonathan Nagler, and Richard Bonneau. "Socially Mediated Sectarianism." *Unpublished Manuscript*. (2018).

⁷⁸ Pan, Jennifer and Siegel, Alexandra A. "How Saudi Crackdowns Fail to Silence Online Dissent." Unpublished Manuscript. 2019.

⁷⁹ For example: Howard, Philip N., and Muzammil M. Hussain. "The upheavals in Egypt and Tunisia: The role of digital media." *Journal of democracy* 22, no. 3 (2011): 35-48; Howard, Philip N., Aiden Duffy, Deen Freelon, Muzammil M. Hussain, Will Mari, and Marwa Maziad. "Opening closed regimes: what was the role of social media during the Arab Spring?." *Available at SSRN 2595096* (2011).

⁸⁰ See Smidi, Adam, and Saif Shahin. "Social Media and Social Mobilisation in the Middle East: A Survey of Research on the Arab Spring." *India Quarterly* 73, no. 2 (2017): 196-209. for an overview and Aday, Sean, Henry Farrell, Marc Lynch, John Sides, and Deen Freelon. "New media and conflict after the Arab Spring." *United States Institute of Peace* 80 (2012): 1-24. for an example of empirical evidence using Twitter data to make this argument.

⁸¹ For an overview of this debate and the empirical evidence on both sides, see: Tucker, Joshua A., Jonathan Nagler, Megan MacDuffee, Pablo Barbera Metzger, Duncan Penfold-Brown, and Richard Bonneau. "Big data, social media, and protest." *Computational Social Science* 199 (2016).

⁸² Lynch, Marc ; Freelon, Deen and Aday, Sean. 2014. *Syria's Socially Mediated Civil War*. United States Institute of Peace.

⁸³ Starbird, Kate, Ahmer Arif, Tom Wilson, Katherine Van Koevering, Katya Yefimova, and Daniel Scarnecchia. "Ecosystem or echo-system? Exploring content sharing across alternative media

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): 5319

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