Data & Policy (2025), 7: e19
doi:10.1017/dap.2024.96

RESEARCH ARTICLE

State-linked manipulated media in the time of Covid-19: a look at Iran

Benjamin E. Bagozzi[1], Karthik Balasubramanian[2], Rajni Goel[2] and Chris Parker[3]

[1]Department of Political Science & International Relations, College of Arts & Sciences, University of Delaware, Newark, DE, USA.
[2]Information Systems & Supply Chain Management, School of Business, Howard University, Washington, DC, USA.
[3]Darden Graduate School of Business, University of Virginia, Charlottesville, VA, USA.
Corresponding author: Benjamin E. Bagozzi; Email: bagozzib@udel.edu

Received: 06 January 2024; **Revised:** 24 September 2024; **Accepted:** 20 November 2024

Keywords: Covid-19; manipulated media; text-as-data; topic modeling; Twitter

Abstract

What drives changes in the thematic focus of state-linked manipulated media? We study this question in relation to a long-running Iranian state-linked manipulated media campaign that was uncovered by Twitter in 2021. Using a variety of machine learning methods, we uncover and analyze how this manipulation campaign's topical themes changed in relation to rising Covid-19 cases in Iran. By using the topics of the tweets in a novel way, we find that increases in domestic Covid-19 cases engendered a shift in Iran's manipulated media focus away from Covid-19 themes and toward international finance- and investment-focused themes. These findings underscore (i) the potential for state-linked manipulated media campaigns to be used for diversionary purposes and (ii) the promise of machine learning methods for detecting such behaviors.

Policy Significance Statement

This paper develops and validates resources for the interpretable detection of manipulated social media. It also offers insights into the specific tactics that authoritarian governments favor when manipulating media to distract from domestic and international challenges. With this knowledge, government agencies confronted with manipulated media from abroad may be better equipped to identify the entities behind malicious accounts, their whereabouts, and their operations—which in turn could be used to disrupt and deter future manipulated media production. Furthermore, such agencies may be able to identify specific weaknesses that adversarial countries are concerned about based on the type of messaging the manipulated media contains. Finally, for social media companies—and those tasked with regulating or monitoring such companies—this paper offers resources and tactical insights that could be used for speedier and more comprehensive manipulated media scrutiny.

1. Introduction

Following Russia's widely publicized efforts to interfere in the 2016 US presidential elections (Eady et al., 2023; PBS, 2020), online state-linked manipulated media has proliferated at a global scale (Bradshaw and

This research article was awarded Open Data and Open Materials badges for transparent practices. See the Data Availability Statement for details.

Howard, 2019; Nemr and Gangware, 2019; Rulis, 2024). Beyond nation-states' learning from others' successes and failures, much of this proliferation can be attributable to the rise of social media. Twitter (although Twitter was re-branded to "X" in July 2023, we will refer to the company as Twitter because the data used in this paper was collected prior to the rebrand) has become an especially important social media platform for such state-linked manipulated media campaigns, with Twitter identifying over 37 distinct state-linked manipulated media campaigns spanning over 20 countries and 200-million tweets between 2018 and 2021 alone (Twitter, 2021).

State-linked manipulated media campaigns on Twitter continued to evolve during the Covid-19 pandemic, with numerous studies now showing significant state-linked efforts toward spreading misinformation about Covid-19 (Broniatowski et al., 2021; Moy and Gradon, 2020; Sciubba Caniglia, 2020). Yet we know comparatively little about how domestic Covid-19 experiences during the pandemic in-turn affect these same states' manipulated media campaigns. Given that the Covid-19 pandemic induced a variety of repressive offline behaviors among authoritarian-leaning states (Barceló et al., 2022; Grasse et al., 2021), the relative absence in understandings of how Covid-19 experiences may or may not affect nation-states' online manipulated media practices is surprising as the deliberate spread of this media by nation-states has broader implications for national security by it potentially influencing or diverting public opinions, and political processes.

Our paper addresses this lacuna by studying the effects of Iran's domestic Covid-19 experiences on a long running Iranian state-linked manipulated media campaign. Because this campaign had begun well before the start of the Covid-19 pandemic, we treat the pandemic as an exogenous shock and study whether Iran's domestic Covid-19 case rates influence the topical themes underlying this manipulated media campaign on Twitter. We apply a series of machine learning tools for variable selection and topic modeling, using a structural topic model (STM; Roberts et al., 2013) to evaluate the association between Iranian Covid-19 rates and a set of themes spread by Iranian state-linked manipulation agents on Twitter. We find that increases in domestic Covid-19 cases in Iran lead the Iranian state-linked manipulated media campaign under study to (i) decrease its focus on Covid-19-based content and (ii) increase its focus on economic content related to commodity prices and international financial markets.

As elaborated upon in our conclusion, the above findings have important implications for science and policy pertaining to state-linked manipulated media. Findings (i)–(ii) suggest that authoritarian governments such as Iran may at times use manipulated media campaigns as a diversionary tool—shifting attention away from pressing issues that threaten domestic stability. While authoritarian countries' use of diversionary tactics is widely recognized in the contexts of both international relations and domestic politics, our paper is one of the first to highlight the use of diversionary tactics within the contexts of Covid-19 and online manipulated media campaigns. Our findings likewise underscore the value of covariate-informed topic modeling for the detection of shifts in ongoing manipulated media campaigns. As a growing number of governments and social media companies endeavor to combat manipulated media campaigns, these results accordingly highlight a suite of tools that may allow future governmental or industry actors to better anticipate, detect, and/or counter such campaigns.

2. Theoretical motivation

Why would countries like Iran expend energy and resources on a manipulation campaign on Twitter? To answer this question, we first must define manipulated media. While manipulated media can be defined in a number of ways, we will use Twitter's definition for the purposes of this article. We use Twitter's definition because we are analyzing the data from state-linked accounts suspended by Twitter for violating their manipulated media policies. Twitter defines "platform manipulation as using Twitter to engage in bulk, aggressive, or deceptive activity that misleads others and/or disrupts their experience" (Twitter, 2023). Note that "misinformation" is a subset of manipulated media by Twitter's definition—misinformation on social media can be defined as "all false or inaccurate information that is spread in social media" (Wu et al., 2019).

Much research has been dedicated to developing strategies and tools to detect, verify, mitigate, and limit the spread of several varieties of misinformation, rumors and fake news (Dang et al., 2019; da Silva et al., 2019; Zhang and Hara, 2020; Bagozzi et al. 2024). In this respect, da Silva et al. (2019) present an extensive survey of machine learning applications, using different techniques and conceptual models, that have been developed to tackle the challenges of manipulated media detection. Our work utilizes previously identified Twitter manipulated media data to illustrate a potential strategic usage of targeted social media manipulation. Specifically, we consider the potential for manipulated media being used as a diversionary tool. Diversion is a social influence tactic that involves redirecting an audience's attention away from an issue or argument by introducing a new topic or by distracting that audience with a separate issue. We analyze topics and timing of manipulative tweets and illustrate a novel perspective of how new forms of 21st century diversion are being conducted.

Our expectations for diversion in this context draw from extant political science and international relations research. Diversionary war theory posits that governments and their leaders will often value foreign conflict as a means for diverting domestic attention away from domestic challenges, and as such, will at times strategically engage in conflictual international behaviors in order to divert public attention away from pressing domestic issues (Amarasinghe, 2022). To this end, leaders will commonly size upon or spark international conflicts as a means of nudging domestic actors away from exacting costs on their regime or leadership (Enterline, 2010). Much research has been conducted on the possible pathways by which leaders may undertake such diversionary behavior, from using military force abroad, to foreign policy adjustments, and other tactics—typically with the fundamental goal of distracting domestic audiences from domestic economic and/or political turmoil (Enterline, 2010; Kanat, 2011, 2014; McLaughlin Mitchell and Thyne, 2010; Smith, 1996; Oakes, 2012; Yeh and Wu, 2020). In this vein, diversionary tactics tend to be multifunctional. They can simultaneously serve to divert blame to other (international or domestic) actors, distract from domestic challenges, boost leaders' popularity, and/or demonstrate leaders' competence in the face of domestic problems (Oakes, 2012).

Multiple studies have hypothesized and explored the unique methods being utilized by governments as diversionary tactics. Increasingly, such studies have come to identify tactics that fall outside the scope of war or militarized conflict. Carter (2020), for example, develops a theory of "diversionary cheaptalk" whereby leaders criticize foreign nations during domestic economic challenges as opposed to pursuing militarized conflict. Amarasinghe (2022) likewise claims that governments, during times of domestic turmoil, engage in diversionary tactics that prioritize verbal aggression in foreign interactions rather than violent interstate conflict. The authors consider international football losses as "exogenous" sentiment shocks in relation to domestic instability and governments' associated international interactions. They find that "shocks" associated with such football losses increase domestic turmoil and in turn causally increase governments' international interactions (Amarasinghe, 2022). This exemplifies a short-term diversionary tactic that can be characterized as both relatively low-cost and low-risk—similar to the social media manipulation diversion strategy we explore in this work.

Research furthermore suggests that authoritarian governments will be especially likely to leverage misinformation and related media-based content for diversionary purposes (Alrababa'h and Blaydes, 2021; Gray, 2010; Koehler-Derrick et al., 2022). Emphasizing the costliness to using actual armed conflict for diversionary purposes, Alrababa'h and Blaydes (2021) posit—and find—that Syria's government instead favored tactics of media manipulation concerning external threats and conspiracy theories in diversionary manners. Drawing upon this logic, Koehler-Derrick et al. (2022) go on to argue that authoritarian governments will be especially likely to spread diversionary conspiracy theories in state-controlled media when they are under threat, owing to the fact that in these contexts the benefits to this form of media manipulation offset any costs of tactical backfire. Rozenas and Stukal (2019) likewise find evidence to suggest that authoritarian governments increasingly favor tactics of media manipulation rather than censorship, in particular, using media manipulation to attribute bad domestic news to external factors.

A separate stream of research has examined social media messaging to explore the presence of discursive deflection. One study examines the tweets of President Trump (Ross and Rivers, 2018), suggesting that two of Trump's tweet strategies included diversion and deflection. A second recent study

https://doi.org/10.1017/dap.2024.96 Published online by Cambridge University Press

presents evidence suggesting that leaders use Twitter to divert media from topics that are potentially harmful or threatening to them (Lewandowsky et al., 2012). Another contemporary social media-focused study of diversionary tactics investigates the use of bots in efforts to suppress or prevent online and offline opposition activities, essentially leveraging the power of social media against (potential) mass political protests and related political instability in authoritarian regime contexts (Stukal et al., 2022). These authors observe evidence of strategies such as bot volumes, retweet diversity, cheer leading, and negative campaigning prior to, during, and following opposition activities. Moreover, it is unlikely that such efforts will be received merely as cheap talk, given broader extant findings concerning the influence of tweets on international politics, international conflict, and international investment (e.g., Duncombe, 2019; Tan and Tas, 2021; Valle-Cruz et al., 2022; Zeitzoff, 2018).

Social media manipulation research also gives consideration to the specific topics most prevalent in the media being spread. For example, studies have investigated the magnitude, propagation, and/or content of manipulated media regarding Covid-19 on Twitter (Himelein-Wachowiak et al., 2021; Yini Zhang and Lukito, 2023). Lee et al. (2023) trace specific Covid-related conspiracy content through social media to understand differences in manipulated media tactics used therein. In line with several studies reviewed in our introduction, Verrall (2022) similarly discusses the use of Covid-19 disinformation, misinformation, and malinformation (DMM) during the Covid-19 pandemic and DMM's evolving capabilities. Yet research into manipulated media during Covid-19 primarily focuses on how manipulated media spreads Covid-19 misinformation or disinformation, rather than considering how manipulated media may be used by governments to divert attention from Covid-19-related challenges. We hence extend this research by investigating the potential for diversionary tactics of manipulated media to increase in response to domestic Covid-19 challenges. Based upon the literature reviewed further above, we focus on an authoritarian country's state-directed manipulated media tactics (i.e., Iran). Given the recent findings concerning increases in commodity futures-based fake news during the Covid-19 pandemic and its associated period of heightened global market uncertainty (Banerjee et al., 2024), we anticipate international finance/investment-focused media manipulation to be especially likely in this context.

In these contexts, several characteristics of the Iranian state-linked misinformation campaign that we consider—and of Twitter access and English language usage in Iran more generally—require further discussion in order to establish the scope conditions for our theoretical expectations and analysis. First, we note that despite Twitter's being banned in Iran since 2009, past research suggests that it remains widely used and accessed in Iran (Hashemi et al., 2022; Jafari, 2020) and has played a pivotal role in Iran's contemporary political sphere (Faris and Rahimi, 2016; Khazraee, 2019). Perhaps correspondingly, prominent Iranian political elites now commonly hold official Twitter accounts, whereas many everyday citizens frequently bypass Iran's official Twitter ban via virtual private networks (Ziabari, 2023). Research has likewise found evidence of several key clusters of Persian Twitter users, including the Iranian diaspora, reformists, and an increasing number of conservative crowdfunded elites (Kermani and Adham, 2021). The latter group of elites has in turn been characterized as encompassing a wide range of pseudo-intellectuals, clerics, young devotees, and financed troll armies that together operate under a mandate of winning "a war of narratives against [Iran's] skeptical people and anyone else in the world who doesn't sympathize with the Islamic Republic" (Ziabari, 2023). Research that has compared English and Persian tweets in this vein has furthermore found subtle differences in types of influential Twitter accounts operating in each language, with Persian language tweets being more dominated by Iranian micro-celebrities and English-language tweets seeing the most retweets for institutional elites intersecting with journalists and news/media outlets (Khazraee, 2019).

Note also that the specific Iranian state-linked misinformation tweet campaign considered below is multilingual in its tweet content. While Twitter is unable to definitively classify the language of every tweet therein, the most commonly tagged languages are Spanish (35%), English (21%), Indonesian (10%), and Farsi (8%). Thus, the campaign under analysis can be assumed to have a substantial international focus, albeit with a secondary—likely elite oriented—domestic component. Given, moreover, that our eventual empirical focus is specifically on the English-language tweet subset of this misinformation campaign for purposes of comparability and exposition, our own analysis and

https://doi.org/10.1017/dap.2024.96 Published online by Cambridge University Press

corresponding scope conditions assume the intended audience of our tweets to be primarily international—including Iranian diaspora—and secondarily to include domestic Iranian elite. In the latter case, we can further note that English competency within Iran's broader population—though difficult to estimate precisely—is likely to be fairly limited during our time period of analysis, with the biggest exceptions being Iran's younger generations in (central and northern) Tehran, Isfahan, and Shiraz (Goodrich, 2020). This reinforces our characterization of this campaign's most likely target audience as being one of international actors and domestic Iranian elites.

Such international audiences and domestic elite actors are not always explicitly or exclusively considered within studies of diversionary war. Nevertheless, we contend that each target group aligns well with the diversionary strategies outlined further above. Indeed, the targeting of these specific actors with manipulated media can critically serve to divert the attention of potentially adversarial foreign governments, international media, and Iranian diaspora away from Iran's domestic Covid-19 challenges. Importantly, each of these international and elite actors can in turn influence Iran's broader domestic population through various channels—thus linking their sentiments and actions to overall domestic Iranian support for Iran's government. In this manner, Iran's broader domestic audiences can be seen as a second, and more indirect, target of our anticipated manipulated media tactics to the extent that international influences—and Iran's economic, business, and political elites—influence this broader population.

3. Data

In order to conduct this research, we used (dis)aggregated data across a wide variety of sources. Our dependent variables are all derived from the content of tweets from accounts that have been identified by Twitter as state-linked disseminators of manipulated media. Our primary independent variables include Covid-19 rates, commodity prices, measures of weather, and sociopolitical events. Each of those datasets is described below.

Tweets: Our dependent variables are derived from Twitter's curated repository of state-linked manipulated media. We will henceforth refer to these tweets of interest as "manipulative tweets." This particular Iranian manipulative tweet sample was identified and released by Twitter in 2021. It covers 560,571 total tweets made by 209 distinct accounts during the period January 1, 2011, to December 27, 2020. After implementing the processing steps described below, these tweets were used as dependent variables in both a preliminary Lasso analysis and our paper's eventual STM models, albeit at different aggregations.

For the Iranian manipulative tweet sample mentioned above, Twitter provides each identified tweet's original text alongside associated meta-data pertaining to tweet date, tweet language, tweet ID, user profile information, and replies/retweets—among other fields. For our state-linked Iranian manipulative tweets, we subset the full set of manipulative tweets to retain only those tweets that Twitter identified as being made in English. This ensures a consistent sample of tweets for our anticipated STM analyses, retaining 115,723 tweets from Twitter's full sample of 560,571 Iranian state-linked manipulative tweets.

Covid-19: Covid-19's effects permeated almost every aspect of society at the beginning of 2020. It is therefore unsurprising that there emerged an urgent need for a database of daily data on Covid-19 cases and deaths to be collected and maintained by an internationally recognized and trusted entity. The World Health Organization (WHO), an agency of the United Nations filled that role. While collecting health data during the global pandemic was almost surely imperfect—especially in countries without transparent governments, WHO data were the most comprehensive, global source for Covid-19 cases in the time period of interest. Therefore, we retrieved WHO's daily Covid-19 data for both Iran and globally starting from January 3, 2020 (when WHO began publishing global counts) to June 23, 2021 (after Iran's specific manipulated media operation described in the previous subsection had ceased). Briefly, we can note that Iran was among the first countries in the world to experience the Covid-19 pandemic and experienced several surges in daily cases thereafter, including in November 2020, April 2021, August 2021, and February 2022 (Moghanibashi-Mansourieh et al., 2023). Iran's response to the Covid-19 pandemic was similarly uneven. For example, the country's vaccination efforts throughout this period faced a number of challenges, especially early-on (Moghanibashi-Mansourieh et al., 2023, 541–542). While Iran

https://doi.org/10.1017/dap.2024.96 Published online by Cambridge University Press

implemented a number of rolling restrictions on public gatherings and public spaces during the Covid-19 pandemic, full-scale lockdowns were imposed more sparingly, often lasting for periods of 1–2 weeks at a time and designated to specific subregions of Iran (A3M, 2024).

To complement our daily Covid-19 data for Iran and globally, we also retrieved Covid-19 data for a set of countries that were representatives of Iran's geopolitical allies, its geopolitical adversaries, and its neighbors. The representative adversarial countries were the United States, Israel, the United Kingdom, and Saudi Arabia. Its representative "allied" countries were Russia, Syria, Lebanon (where Iran backs the Lebanese militant group, Hezbollah), Yemen (where Iran backs the Yemeni Houthi rebel group), and the Palestinian Territories. These data were again drawn from the WHO source mentioned above.

Commodity prices: In investigating the effects of Covid-19 on Iran's state-linked manipulated media campaign, this research also accounts for how commodity prices might be associated with Iran's manipulated media operations. One commodity dominates Iran's exports: oil. Therefore, we retrieve daily price, for open market days, data for the oil variant that has the most liquid market: Brent crude oil (in USD per barrel). Similarly, we collect daily price data from Iran's major food imports: rice (USD per 100 pounds), soybeans (USD per bushel), corn (USD per bushel), sugar (USD per pound), barley (USD per bushel), and cotton (USD per pound). All commodity price data, as well as the asset price data below, come from Bloomberg's direct data feed via the Chicago Board of Trade. These daily data span the entirety of the period for which the identified Iranian manipulated media operation was active.

Asset prices: Beyond commodity prices, we are also interested in how the prices of assets, such as precious metals, cryptocurrencies, and stocks, might be associated with our dependent variables. Therefore, we collect daily data on the price of gold, silver, and platinum. Bloomberg sources these prices from the Chicago Board of Trade. Given the online nature of manipulated media, we also considered the daily prices in USD (at midnight UTC-4) of the two largest cryptocurrencies by market capitalization as of January 2021: Bitcoin and Ethereum. Given that cryptocurrency valuations are quite volatile, we also examine the daily change in price of each of these cryptocurrencies. Finally, we include two measures of US stock market performance. The first measure is the daily closing price of the Vanguard Total Stock Market Index Fund ETF (VTI), a passively managed fund that includes the weighted share prices of large, medium, and small-capitalization publicly traded companies. We use VTI as it is a single-number proxy for the state of the United States economy. Additionally, we use the daily closing price of VOO, the exchange traded fund that tracks the weighted performance of the Standard and Poor's (S&P) 500—a very commonly used basket of 500 of the largest United States publicly-traded companies by market capitalization.

Events: To account for political events in Iran and globally, we leverage event data from the Integrated Crisis Early Warning System (ICEWS; Boschee et al., 2015). ICEWS machine codes events from international news(wire) sources according to specific source and target actors, and a wide range of event types encompassing (verbal and material) conflictive and cooperative actions. We use the raw ICEWS data to create several daily event count aggregations. We first calculate the total global daily events in ICEWS for our daily sample period as a measure of overall political interaction at a global scale. We then separately extract a similar daily aggregation for events of all types that involve an actor associated with the country of Iran as a source (initiator) or target of a particular event. Next, we separately subset and aggregate only those events associated with (i) non-state-initiated domestic protests within Iran and (ii) repressive events in Iran initiated by Iranian state-based actors targeting non-state-based Iranian actors. Past research has established the appropriateness of using ICEWS for studying repressive events (Bagozzi et al., 2021) and protests (Steinert-Threlkeld, 2017)—including Covid-19-era protests (Mitternich, 2020).

Additional variable details: Alongside the variables summarized above, we consider control variable measures of Iran's daily temperatures and precipitation levels, obtained from the National Oceanic and Atmospheric Administration's online repository of global historical weather data. We likewise include a month (time) counter variable that spans the full duration of our state-linked manipulative tweet corpus. This control—alongside our aforementioned temperature measure—accordingly allow us to

https://doi.org/10.1017/dap.2024.96 Published online by Cambridge University Press

account for both longer-term and cyclical (i.e., seasonal) temporal dependence in our data. To further control for political dynamics, we also consider dichotomous election indicators for domestic elections in Iran, the United States, and Israel, set equal to one beginning in the first month of a given country's election year and remaining one until the exact day of that country's election. For several of the variables mentioned further above, we interpolate missing values, primarily for our ensuring complete coverage for the daily aggregations used in our preliminary Lasso analysis, rather than for our STM analyses. Remaining missing values lead us to listwise delete observations lacking full coverage on our variables within our final STM analyses. This primarily arises toward the end of our time series. Finally, many of the variables outlined above are noticeably skewed. We accordingly log a majority of the covariates considered here. See Table 1 for more details on the variables that were logged in our analyses.

4. Analysis and results

4.1. Covariate selection via Lasso

We use STMs (Roberts et al., 2013) to evaluate the association between (i) Iranian Covid-19 rates and (ii) several specific themes of manipulated media spread by Iranian state-based agents on Twitter. To do so, we first must identify a reasonable set of control variables for inclusion within our eventual STMs. The present subsection describes this process.

We employ an auxiliary Lasso model to evaluate all variables described further above in relation to an aggregated version of our Iranian manipulative tweets. Based upon our Lasso model, we then retain the subset of the variables that reliably predict our aggregated version of Iranian manipulative tweets as our STMs' control variables. Importantly, and because the aggregated version of Iranian manipulative tweets considered here is not the primary dependent variable of interest in our ultimate STM analysis, this Lasso step allows us to identify a subset of variables for inclusion as controls within our STMs in a manner that is at least partially removed from our final data and analysis framework.

While our primary STM analysis considers the individual texts of each retained tweet, our current efforts to identify relevant control variables for this ensuing tweet-level Lasso analysis next collapsed our sample to daily manipulative tweet counts covering the January 1, 2011, to December 27, 2020, time period. These daily tweet counts were then paired with the day-level aggregations of all variables described further above. Daily frequency counts of Iranian state-based manipulative tweets where then analyzed in relation to these variables via a Lasso model with a Poisson link function. Our Lasso in this case allows for variable selection and regularization in the context of a count-based dependent variable, thereby providing a means of identifying an optimal subset of control variables for use in our ultimate tweet-level STMs under a distinct operationalization of our dependent and independent variables.

Our Lasso results appear in Table 1. We also estimate non-penalized Poisson models that include (i) all Lasso covariates and (ii) the final selected covariates from our Lasso step in the Supplementary Appendix (Table A.1) to evaluate the levels of multicollinearity within each specification via Variance Inflation Factors (VIFs). While our final specification retains a relatively higher share of covariates with VIF scores exhibiting relatively moderate to low VIF scores in our full specification, it does also retain several covariates with relatively high VIF scores. Returning to our main Lasso results, predictors denoted with dashed lines are those that were regularized to zero or that failed to obtain a sufficient significance threshold to be retained. In the latter respect, we set our significance threshold at $p < 0.05$, though we do retain one predictor that falls slightly below this threshold. Turning to Table 1, we find that US and Iranian election periods each reliably increase the daily frequency of Iranian manipulative tweets. On the other hand, higher rates of daily (Iranian or global) political events are each reliably associated with decreased daily Iranian tweet counts, whereas higher rates of repression and protest in Iran see increases in daily Iranian manipulative tweets. Turning to our measures of country-specific daily Covid-19 rates, we find different effects

https://doi.org/10.1017/dap.2024.96 Published online by Cambridge University Press

Table 1. *Lasso results for daily manipulative Tweet frequency*

US election	1.327***	(0.014)
Iran election	0.143***	(0.028)
Israel election	—	
Ln precipitation	—	
Ln Iranian events	−0.008*	(0.004)
Ln Iranian repression events	0.020**	(0.008)
Ln global events	−0.047***	(0.004)
Ln Iranian protest events	0.022**	(0.009)
Ln global Covid cases	—	
Ln Israel Covid cases	0.035***	(0.004)
Ln US Covid cases	−0.108***	(0.004)
Ln Iran Covid cases	−0.009**	(0.004)
Ln ether	−0.224***	(0.012)
Ln bitcoin	—	
Ln silver	0.361***	(0.028)
Ln platinum	—	
Ln gold	—	
Ln VTI	—	
Ln VOO	—	
Temperature	10.007***	(0.001)
Ln Jordan Covid cases	−0.047***	(0.004)
Ln Russia Covid cases	−0.052***	(0.005)
Ln Saudi Arabia Covid cases	0.209***	(0.007)
Ln Syria Covid cases	0.018***	(0.004)
Ln GB Covid cases	—	
Ln Yemen Covid cases	0.032***	(0.005)
Ln Lebanon	0.073***	(0.005)
Ln Palestine	−0.052***	(0.004)
Ln oil price 1	—	
Ln oil price	20.522***	(0.024)
Ln gas price 1	—	
Ln gas price 2	−0.044***	(0.010)
Ln rice price	—	
Ln soy price	—	
Ln corn price	−0.384***	(0.028)
Ln sugar price	−0.623***	(0.028)
Ln barley price	—	
Ln cotton price	—	
Month counter	0.079***	(0.002)
Intercept	−2.792***	(0.241)
Observations	3618	
AIC	7102	
λ	2.1047	

*** p<.01; ** p<.05; * p<.10

depending on the Covid-19 afflicted country under consideration. Covid-19 rates in Israel, Saudi Arabia, Syria, Yemen, and Lebanon each reliably increase the daily frequency of Iranian manipulative tweets. By contrast, increased daily Covid-19 rates in the United States, Iran, Russia, and Jordan

https://doi.org/10.1017/dap.2024.96 Published online by Cambridge University Press

are each reliably associated with fewer daily Iranian manipulative tweets. Daily prices of ether, silver, oil, gas, corn, and sugar are also reliably associated with daily Iranian manipulative tweets, although daily prices of a number of other commodities and investment assets such as bitcoin, gold, cotton, and rice are not. Finally, we also find (unsurprisingly) that a month counter is a reliable positive predictor of daily manipulative tweet frequency in Table 1—implying that this particular Iranian manipulation campaign was increasing in volume over time.

In sum, the Lasso model identifies a wide range of reliable predictors of daily Iranian manipulative tweets. These include Covid-19 cases in a range of allied or rival countries, although not global Covid-19 cases. Our retained predictors also include measures of elections in the United States and Iran, although not Israel. Temperatures in Iran are retained as a reliable predictor, although not precipitation rates. Finally, 6 of our 17 economic- or commodity-based measures are likewise retained, along with our time (month) counter. We accordingly retain all reliably identified predictors from this Lasso model as controls within our STM analyses (alongside our primary independent variable measure of Iranian Covid-19 cases, which we can note is also retained within our Lasso model as a reliable predictor). The latter analyses consider the same English language sample of Iranian manipulative tweets as outlined above, albeit now at the individual tweet level, rather than at the daily tweet count level. We now turn to describing this tweet-level analysis.

4.2. STM estimation

We use an STM to investigate how, and whether, Iranian Covid-19 cases influence strategies of Iranian state-based manipulated media. This STM treats our tweets as an admixture of underlying thematic clusters that themselves are comprised of words. It then recovers these word clusters as "topics" (i.e., probability-indexed word vectors). This STM approach extends prior topic modeling approaches, including latent Dirichlet allocation (LDA; Blei et al., 2003), in several notable manners. Of most relevance for our analysis, the STM allows for the inclusion of covariates when estimating one's topics. This allows us to not only extract a set of latent topics for our manipulative tweets, but also to recover estimates of how our external (tweet-level) covariates affect the relative prevalence of said topics. In addition, the STM allows for correlations among one's estimated topics during estimation. Given the potential for strategic coordination across the Twitter accounts' topics, this is likely relevant to our tactical investigations into categories of Iranian state-linked manipulation.

As this analysis is at the level of individual tweets, the STM treats each manipulative tweet as the unit of analysis. Before estimating any STMs, we processed these tweet texts in standard manners for STM applications. This includes transforming all text to lower case, omitting all numbers, removing English-language stop words, excluding very sparse terms, removing punctuation, and then finally stemming all retained terms. In these contexts, stop words correspond to a common set of 174 English stop words as defined by the tm package in R (Feinerer et al., 2008). This—and the other preprocessing steps outlined above—is consistent with the set of English stop words used by other similar STM applications (e.g., Bagozzi and Berliner, 2018; Kwon et al., 2019.)

The STM, like LDA and other mixed-membership topic models, is flexible in the number of topics that it is able to recover from any given corpus, defined hereafter as k. This flexibility in turn requires that we choose a specific value of k within our STM application. For interpretability, we focus on estimating a relatively small number of topics to represent the overarching themes underlying our tweets. We then leverage a variety of model fit diagnostics to determine a defensible k within this range. More specifically, we consider four common diagnostics (i.e., model residuals, held-out likelihood, lower-bound, and semantic coherence). These are common criteria for STM applications, including ones considering Covid-19-related text and/or social media data (Abramova et al., 2022; Chung et al., 2022; Rodriguez and Storer, 2020) in evaluating a plausible range of topics lying between 10 and 60. This range of topic number draws upon the guidance offered in existing STM research, for example, Roberts et al. (2014a) emphasize that "[t]here is no right answer to the appropriate number of topics. More topics will give more fine-grained representations of the data at the potential cost of being less precisely estimated. […] For

https://doi.org/10.1017/dap.2024.96 Published online by Cambridge University Press

small corpora (a few hundred to a few thousand) 5–20 topics is a good place to start." These diagnostics appear in Figure 1 and suggest that an optimal topic number lies in the 20-topic range. This is especially the case for hold-out likelihood, residuals, and lower bound, which each exhibit diminishing returns for $k > 20$, whereas semantic coherence is inconclusive. We therefore assign k to be 20 in our STM analysis.

For our $k = 20$ STM, we treat our preprocessed manipulative tweets as our "documents" and then add our Lasso-selected variables as prevalence covariates. To safeguard against instability in our final set of topic estimates, we follow extant research (e.g., Bagozzi and Berliner, 2018; Roberts et al., 2014b) by estimating a array of 50 ($k = 20$) STMs for our manipulative tweet corpus, each with distinct starting parameter values. We then chose an optimal STM from these 50 estimated STMs with the aid of semantic coherence and exclusivity scores.

For each of our STM-derived topics, we report the 20 most associated words in Table 2. Therein, each topic's associated top 20 words are ranked in terms of their frequency-exclusivity (FREX) assignment from left to right (i.e., with most highly associated words appearing first, followed by less associated words). Table 2 also includes our own subjective labels for each of the estimated topics on the right-hand side of the plot. To assign these labels, we not only read the top words depicted here but also reviewed the top 100+ tweets that were estimated to be associated with each estimated topic by our final STM. Reading these tweets for relevant topics provided a much clearer picture of what theme or themes a particular topic encompassed than did initial assessments of the top words alone. Our Supplementary Appendix contains full rationales for each topic interpretation.

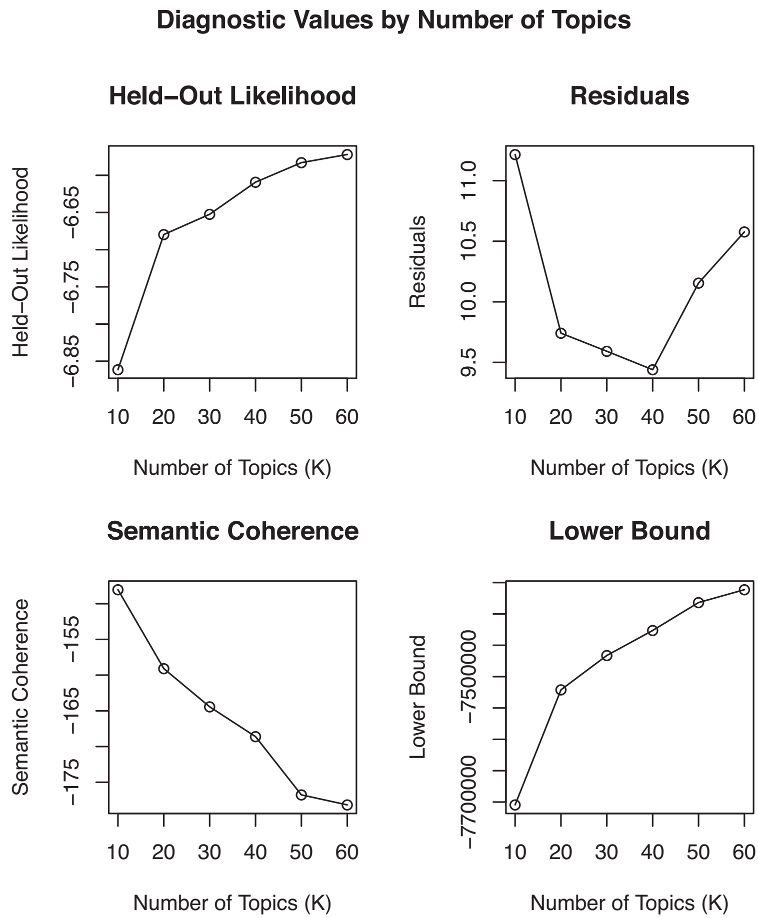

Figure 1. *Topic number selection diagnostics.*

https://doi.org/10.1017/dap.2024.96 Published online by Cambridge University Press

Table 2. *Top 20 words per STM topics, based upon FREX*

Topic	Labels	Top 20 words
1	Sanctions	iran, state, unit, regim, iranian, sanction, polici, interest, russia, foreign, nuclear, implement, meet, depart, financi, must, korea, european, punish, secretari
2	Aggression	let, pleas, friend, fuck, true, hear, guy, away, thought, pass, what, stori, listen, bill, mrsrab-bitresist, there, penc, hero, mcspocki, wall
3	political polarization	america, lie, corrupt, media, obama, twitter, fake, actual, traitor, billionair, blame, deceiv, massacr, want, ruin, team, fals, made, purpos, interfer
4	Israel–Palestine	palestin, palestinian, isra, palestinewillbefre, zionist, gaza, occup, occupi, climatechang, land, west, freepalestin, climat, grouppalestin, climatecrisi, jerusalem, palestineresist, badrianaila, centuri, annex
5	Commodity prices	silver, price, usd, physic, percent, demand, high, billion, sign, shortag, industri, tech, reopen, potenti, appl, forecast, stage, best, quick, ofstrength
6	Election lead-up	senat, novemb, forget, cut, longer, suprem, work, everyth, realdonaldtrump, goal, forev, full, begin, itsjefftiedrich, complet, definit, tramp, boycott, sure, certain
7	Racial tensions	black, still, middl, poor, environ, gun, african, africa, heard, color, wake, anoth, breath, grow, other, mcconnel, robert, around, beat, jobless
8	Gender and politics	realli, women, everi, that, rape, worri, encourag, prefer, rapest, pedophil, approv, voter, amaz, tara, raed, hariss, abil, grant, waduh, hot
9	Religion and taxes	love, god, life, tax, rais, imam, reduc, heart, ali, beauti, choic, thepromisedsaviour, prophet, arbaeen, cost, hussain, livelikeali, savior, muhammad, whole
10	Yemen	yemen, war, nation, saudi, yemeni, poll, situat, yemencantwait, saudiarabia, arabia, lead, danger, food, conflict, clinton, humanitarian, fox, exact, hillari, civil
11	International markets	china, euro, forex, india, wait, futur, ban, sell, isnt, crash, join, safe, stock, replac, way, nokia, bitcoin, cripto, though, market
12	Encouragements	can, now, come, happi, save, week, send, hard, trust, expect, perfect, step, easili, tonight, whatev, help, short, explain, mean, hurt
13	Economic anxiety	white, cant, doesnt, tweet, read, someon, respect, find, ignor, question, wont, john, kid, told, idea, less, proud, incom, mayb, tell
14	Negativity	person, hes, keep, your, hate, ever, didnt, someth, enough, play, speech, wonder, miss, ive, idiot, shameonaungsansuukyi, pelosi, upset, gonna, happen
15	Political violence	protest, attack, polic, iraq, blacklivesmatt, syria, iraqi, syrian, violenc, racism, base, troop, portland, georg, shot, iraqprotest, feder, portlandprotest, demonstr, isi
16	COVID	coronavirus, even, corona, virus, report, video, health, mask, humanrightsviol, pandem, wear, test, level, outbreak, busi, top, post, medic, record, suggest
17	Temporal context	amp, just, day, rememb, start, feel, might, ago, obvious, remind, trend, manipul, scream, shit, doj, els, fond, cftc, soonyou, suboz
18	Anti-democrat	joe, job, harri, manag, hugo, chavez, rich, berni, low, fair, hunter, energi, share, biden, sander, dem, win, result, school, joebiden
19	Election aftermath	thank, defeat, restartlead, retweet, admit, neither, yes, xauusd, turmoil, plummet, trumppenc, educ, candid, paid, polit, comment, votebluetosaveamerica, sevilwar, blackcat, patriot
20	Mass dissemination	right, wish, soldier, morn, general, sad, sdrmedco, photo, eye, pic, treat, event, girl, healthi, cri, nice, may, charg, absolut, equal

https://doi.org/10.1017/dap.2024.96 Published online by Cambridge University Press

The frequency plot in Figure 2 highlights each topic's relative frequency across our entire tweet corpus, after classifying each tweet according to its most dominant topic based upon the STM's word-indexed topic probabilities. We find reasonable coverage across all of our topics, with all topics aside from Topic 7 exhibiting over 2000 associated tweets and our two largest topics (Topics 4 and 15) exhibiting over 12,000 associated tweets a piece. It is worth further noting here that the three topics of focus below all lie within the top 12 most frequent topics according to Figure 2, with one such topic (Topic 16) falling within the top five most common topics across our corpus.

4.3. Focused topic analysis

Of particular interest to our present analysis are Topics 5 (commodity prices), 11 (international markets), and 16 (Covid-19). These three topics exhibit the closest relevance to the overarching content themes of interest that were highlighted further above—namely themes of international finances/investment and Covid-19. Accordingly, we next expand upon the rationale for the subjectively assigned label associated with Topics 5, 11, and 16, based upon our own qualitative review of these associated top tweets and FREX words. Our Appendix provides further elaboration and interpretation of these three topics. After providing detail on Topics 5, 11, and 16 below, we directly assess the extent to which an increase in daily (ln) Iran Covid-19 rates affects attention to each of these topics at the individual tweet level.

Topic 5: Commodity prices. Silver, price, and US dollars (USD) are three of the most important words in this topic. This topic pertains to commodities prices with a particular focus on the change in prices as a result of high demand for rare earth minerals in new technologies. For example, one tweet that weighs heavily on Topic 5 is: "@Tickeron Silver have critical role in high tech such as 5 g mobile solar panel and electronics military high tech … buy physical silver is the best investment at this years investment demand for silver is sky rocketing price of silver to 100 usd soon current price of silver is 17 usd." This, and our broader reading of this topic's content, leads to a natural label of commodity prices for this topic.

Topic 11: International markets. This topic is similar to Topic 5 in that it focuses on finances. However, in contrast to Topic 5 which focuses largely on commodities, Topic 11 has a much more international focus with China, Euro, forex (foreign exchange), and India marking some of the most important words. This topic also covers Bitcoin and crypto markets. We therefore label this topic as

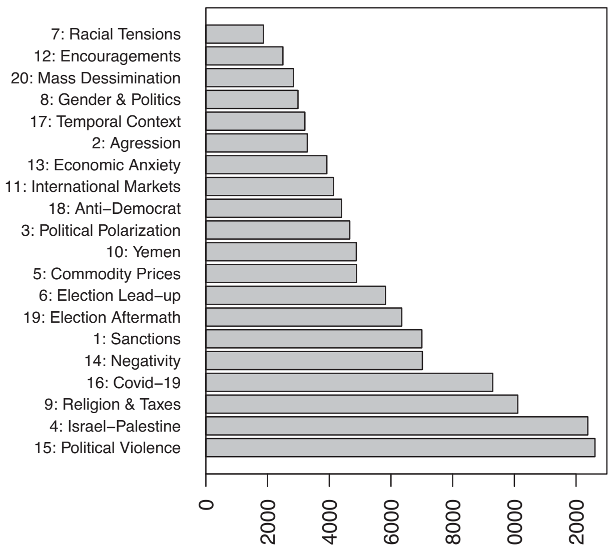

Figure 2. *Distribution of topics across Tweet Corpus, according to each Tweet's most dominant topic.*

https://doi.org/10.1017/dap.2024.96 Published online by Cambridge University Press

"International Markets." An illustrative tweet is: "@ScotlandNT @dundeeunitedfc Euro is oldman eus money look at italy and Spain and Greece euro it isn't safe haven see brexit europe union end soon europe dont have oil gas mine army and young people europe dont have good furure#trump #gold #stock market #china forex signal eurusd #euro."

Topic 16: Covid-19. Nearly every top word in this topic can be easily connected to Covid-19. Even some that may not be as easily connected (humanrightsviol or wear) are likely about mask mandates are the resistance of some to adherence to these mandates. Thus, we label this topic as "Covid-19." Many top tweets from this topic are fairly informational, such as: "Singapore reports record 386 new Covid-19 cases and 9th death #COVID19 #coronavirus https://t.co/3Sw8oVloYK."

Figure 3 presents the estimated effect of a substantively meaningful increase in ln Iran Covid-19 cases (specifically, a change from one standard deviation below the sample mean of Ln Iran Covid-19 cases to one standard deviation above the sample mean of Ln Iran Covid-19 cases) on the prevalence of our three primary topics of interest: Topics 5 (commodity prices), 11 (international markets), and 16 (Covid-19). These substantive effects are plotted with 95% confidence intervals, where a positive (negative) value on the x-axis implies than an increase in ln Iran Covid-19 cases is associated with an increase (decrease) in attention to a particular topic within our sample of Iranian state-linked manipulative tweets.

We find in this case that a reasonable increase in the log of Iran's domestic Covid-19 cases is associated with a reliable decrease in a manipulative tweet's proportion of attention dedicated toward our identified Covid-19 topic. At the same time, a comparable increase in Iran's (logged) domestic Covid-19 case rate is reliably associated with increases in a given manipulative tweet's attention toward themes relating to international markets and commodity prices. Together, these results provide evidence for Iran's use of manipulation tactics in a diversionary manner when faced with increased domestic challenges: as domestic Covid-19 cases climb, Iran reduces its efforts to direct attention toward the pandemic via manipulation, instead favoring an intensification of manipulation that redirects attention toward international investing and commodity prices.

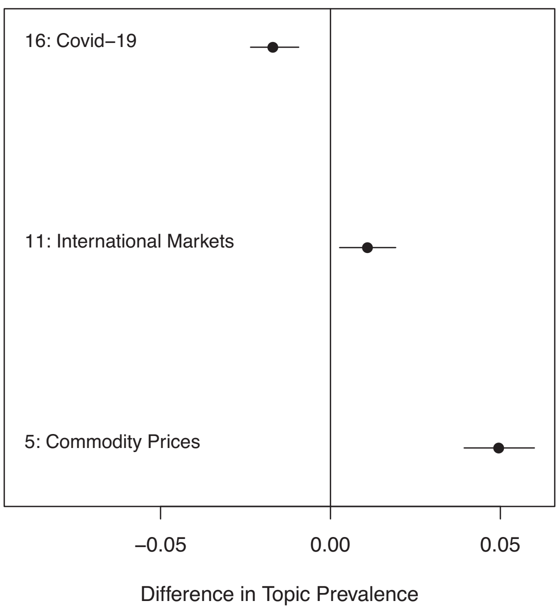

Figure 3. *Estimated effect of change in Ln Iran Covid19 cases on topics of interest.*

https://doi.org/10.1017/dap.2024.96 Published online by Cambridge University Press

4.4. Sensitivity assessment

In the Supplementary Appendix, we reevaluate the robustness of our analysis in several manners. To demonstrate that our findings are not dependent upon our main STM specification's over- or under-controlling for possible confounds, we reestimate our primary STM specification and subsequent focused topic analysis in two manners. First, we reestimate our 20-topic STM when including *ln Iran Covid-19 cases* as the only prevalence covariate. After estimating this STM and re-identifying our three focused topics of interest, we find in Supplementary Figure A1 that the estimated effects of *ln Iran Covid-19 cases* mirrors those reported in Figure 3 in both direction and magnitude—albeit with more precise 95% confidence intervals. Next, we re-estimate our 20-topic STM when including all covariates initially included in our Lasso variable selection step, rather than only those variables retained by the Lasso. In Supplementary Figure A2, we find that *ln Iran Covid-19 cases* exhibits comparable relationships (in reliability, direction, and magnitude) with our three topics of interest when compared to those reported in Figure 3. We then re-estimate our primary 20-topic STM specification after adding the interaction between *ln Iran Covid-19 cases* and *ln Iran protest events.* As demonstrated in Supplementary Figure A3, we find some evidence to suggest that *ln Iran Covid-19 cases* may exert a larger effect (in our anticipated directions) on tweets discussing international markets and Covid-19 when *ln Iran protest events* are high as opposed to low. However, this was not the case for commodity prices.

As a final means of assessing the empirical support for our primary STM findings, we return to our main STM and use its prevalence covariates to estimate the temporal change in predicted topic probability for each of our 20 topics in relation to changes in our month counter (i.e., time counter) covariate. We then plot these predicted trends over time for the year 2020. 2020 is the first full year of the Covid-19 pandemic and also the only full Covid-19 year that we have complete coverage on across our associated covariates and STM. The plotted trends are presented in Figure 4 where for each specific topic's predicted prevalence over time (in black), we also include the other associated 19 topics' predicted prevalence time series plots (in grey) for scale. Even at this higher level of aggregation, we find evidence in support of our primary theoretical contentions and STM findings. Herein, we observe pronounced increases in topical attention to Topics 5 (commodity prices) and Topic 11 (international markets) in June of 2020; as well as a second

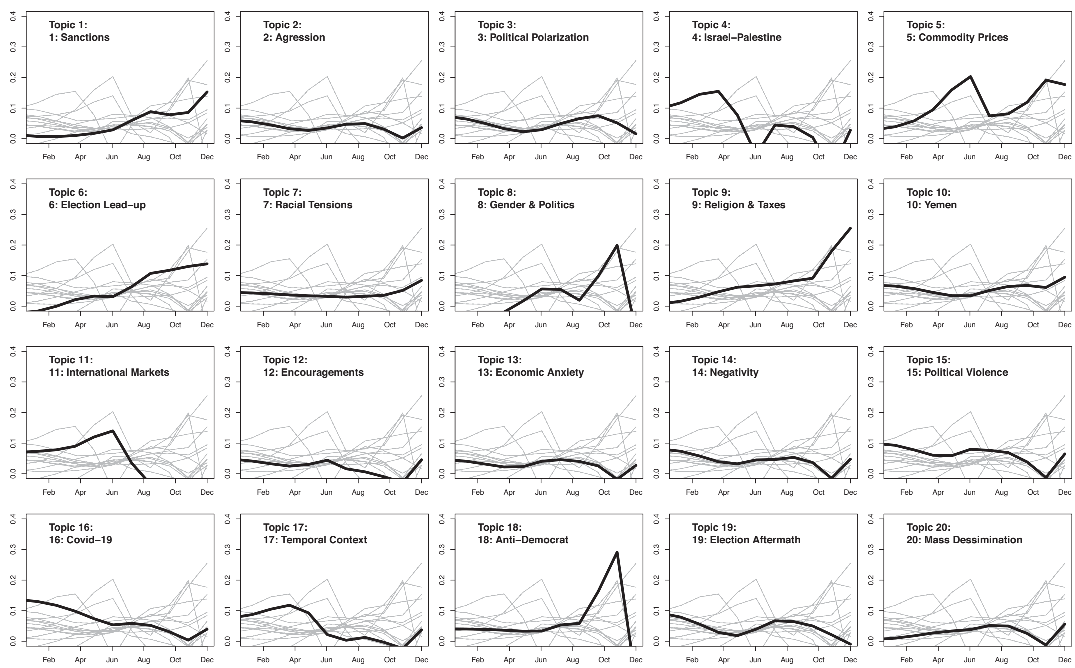

***Figure 4.** Predicted monthly topic prevalence for 2020.*

spike in topical attention to Topic 5 (commodity prices) in November 2020. At the same time, we see a general decline in topical attention to Topic 16 (Covid-19) throughout the 2020 period, with pronounced dips in June 2020 and November 2020. These patterns align well with 2020 Covid-19 trends in Iran in several respects. First, Covid-19 rates were generally increasing throughout 2020 for Iran, which is consistent with our finding of a sustained decrease in attention to Topic 16 (Covid-19) in Figure 4. At the same time, Iran also saw its most pronounced surge in new Covid-19 cases during the first year of the pandemic in November 2020 (Motamedi, 2020; Ritchie et al., 2022). The corresponding dip in attention to Topic 16 (Covid-19)—and spike in attention to Topic 5 (commodity prices)—at that same juncture in Figure (Motamedi, 2020; Ritchie et al., 2022) each thereby reinforce our expectations and primary STM findings for these two topics.

Likewise, June 2020 was a major turning point in Iran's domestic Covid-19 crisis. Indeed, it has been noted that leading up to June 2020 the Iranian government apparently had Covid-19 relatively under control in April to May of 2020 "until the beginning of June, 2020, when the media reported a worrying sharp increase in the number of COVID-19 cases that mirrored March's peak levels [and] Iran, seemingly, has been catapulted into a second wave of disease" (Venkatesan, 2020, 784). Our June 2020 findings in Figure 4 capture this shift with its aforementioned spikes in tweets on Topics 5 (commodity prices) and Topic 11 (international markets) and dip in Tweet attention to Topic 16 (Covid-19). Thus, and consistent with our main STM analyses, manipulative tweet content associated with Covid-19 (international markets and commodity prices) appears to decline (increase) during periods of heightened domestic Covid-19 challenges within Iran. These findings notwithstanding, we do observe a number of additional temporal shifts in Figure 4's remaining topic plots. While many of these do not align with the months highlighted above, some do—suggesting that future investigation of Iran's broader manipulated media strategy in relation to Covid-19 is warranted.

5. Discussion and conclusion

The ubiquity of manipulated media produced by bad actors is an alarming feature of today's internet. But state-linked production of manipulated media is even more pernicious because of the potential scale of operations, and the ability of states to make credible threats of large-scale political disruption. Understanding as much as possible about state-linked manipulated media campaigns, therefore, is of paramount importance for the healthy functioning of democratic societies around the world.

We investigate one aspect of a large-scale manipulated media operation directed by Iran, and we find interesting associations: as Iranian Covid-19 cases increased, the tweets the operation produced that were related to Covid-19 decreased. And similarly, when Iran's Covid-19 cases increased, Iran's production of economic-related tweets increased. While these are certainly not declarations of causality, these associations are plausibly causal. It would not be far-fetched to postulate that Iran may have attempted to use economic-related messaging to divert global attention away from its internal Covid-19 outbreak. And similarly, it would not be outlandish to postulate that Iran would decrease their Covid-19-related media production when they were most vulnerable to negative Covid-19 messaging—namely, when their own internal Covid-19 case rates were high. In other words, we have begun to build a path to understanding what drove Iran's manipulated media messaging strategy.

This work must be caveated in two major ways. First, we are only able to analyze tweets of accounts that have been preidentified by Twitter as Iranian manipulated media. It is plausible, if not likely, that there exists (or existed) a significant number of undetected accounts acting on directions issued by Iran. Second, once Twitter identified the Iranian operation, the accounts were shut down. Because of this, we are only able to analyze the Iranian operation during the period of operations. It is possible (and likely) that Iran's strategy for future manipulated media production will be different, as they have now had time to adjust. Nevertheless, an understanding of their past behavior will very likely be important in developing a predictive profile of future actions they (and others like them) may take in the space.

The implications of this work are clear and point stakeholders toward important actions to counter manipulated media production by bad state actors. For democratic governments targeted by this

https://doi.org/10.1017/dap.2024.96 Published online by Cambridge University Press

manipulated media, this illuminates tactics used in the past. With this knowledge, government agencies may be better equipped to identify entities behind malicious accounts, their whereabouts, and operations—which could be used to disrupt and deter future manipulated media production.

For policymakers involved in foreign policy and diplomacy, we recommend the following actionable steps:

1. Develop monitoring systems to detect shifts in thematic focus of state-linked manipulated media campaigns in real-time.
2. Create rapid response teams to counter disinformation with accurate, timely information.
3. Invest in public education campaigns to increase media literacy and awareness of disinformation tactics.
4. Establish international cooperation frameworks to combat state-linked manipulated media across borders.

For social media companies, this research adds detail on tactics that could be used for speedier detection and suspension of manipulated media operations. We recommend that these companies:

1. Implement content analysis tools to detect shifts in thematic focus over time.
2. Develop artificial intelligence-driven systems to identify patterns indicative of state-linked campaigns.
3. Increase transparency by regularly publishing reports on detected state-linked manipulation attempts.
4. Collaborate with academic researchers to refine detection methods and understanding of manipulated media tactics.

Finally, this work opens the door for further academic research on state-linked manipulated media production. We illustrate a unique dataset and methodology that can and should be utilized for analysis of other bad actors as well as to uncover more tactics utilized by these bad actors. For example, future research could leverage such data to delve into the principal–agent dynamics involving state-linked social media manipulation strategies—drawing upon the extant international relations scholarship on similar principal-agent dynamics involving illicit state-sponsored actors (e.g., Byman and Kreps, 2010; Salehyan et al., 2014). For such analyses or otherwise, further research might also include non-English tweets (Spanish and Farsi, in particular, for this dataset), patterns in the time-of-day tweets are published, and potentially even considerations of the propagation success (measured in terms of retweets and growth of follower counts) of manipulated media. Understanding the actions, motivations, and strategies of these bad state actors is crucially important to protecting our democratic institutions and discourse—the future of democratic societies depend on it.

Supplementary material. The supplementary material for this article can be found at http://doi.org/10.1017/dap.2024.96.

Data availability statement. The data and replication code for this article and its Supplementary Appendix can be found at the following Harvard Dataverse entry: https://doi.org/10.7910/DVN/MJHKNI.

Author contribution. B.E.B., K.B., R.G., and C.P. contributed equally to: conceptualization, data curation, methodology, formal analysis, visualization, writing—original draft, and writing—review and editing.

Ethical standard. The research meets all ethical guidelines, including adherence to the legal requirements of the study country.

References

A3M (2024) Covid-19 Pandemic - Iran. A3M Global Monitoring GmbH. Available at https://global-monitoring.com/gm/page/events/epidemic-0001932.BlC72bvwdP6P.html?lang=en.

Abramova O, **Batzel K and Modesti D** (2022) Collective response to the health crisis among German twitter users: A structural topic modeling approach. *International Journal of Information Management Data Insights 2*(2), 100126.

Alrababa'h A and Blaydes L (2021) Authoritarian media and diversionary threats: Lessons from 30 years of Syrian state discourse. *Political Science Research and Methods 9*(4), 693–708.

Amarasinghe A (2022) Diverting domestic turmoil. *Journal of Public Economics 208*, 104608.

Bagozzi BE and Berliner D (2018) The politics of scrutiny in human rights monitoring: Evidence from structural topic models of U.S. State Department Human Rights Reports. *Political Science Research and Methods 6*(4), 662–677.

Bagozzi BE, Berliner D and Welch RM (2021) The diversity of repression: Measuring state repressive repertoires with events data. *Journal of Peace Research 58*(5), 1126–1136.

Bagozzi Benjamin E, Rajni Goel, Brunilda Lugo de Fabritz, Kelly Knickmeier-Cummings and Karthik Balasubramanian (2024) A Framework for Enhancing Social Media Misinformation Detection with Topical-Tactics. *Digital Threats: Research and Practice 5*(2), 1–29.

Banerjee AK, Sensoy A, Godell JW and Mahapatra B (2024) Impact of media hype and fake news on commodity futures prices: A deep learning approach over the Covid-19 period. *Finance Research Letters 59*, 104658.

Barceló J, Kubinec R, Cheng C, Rahn TH and Messerschmidt L (2022) Windows of repression: Using Covid-19 policies against political dissidents? *Journal of Peace Research 59*(1), 73–89.

Blei DM, Ng AY and Jordan MI (2003) Latent Dirichlet allocation. *Journal of Machine Learning Research 3*, 993–1022.

Boschee E, Lautenschlager J, O'Brien S, Shellman S, Starz J and Ward M (2015) ICEWS coded event data. Harvard Dataverse, V4. https://doi.org/10.7910/DVN/28075.

Bradshaw S and Howard PN (2019) *The Global Disinformation Order: 2019 Global Inventory of Organised Social Media Manipulation*. Oxford: University of Oxford. Available at https://demtech.oii.ox.ac.uk/wp-content/uploads/sites/12/2019/09/CyberTroop-Report19.pdf.

Broniatowski DA, Kerchner D, Farooq F, Huang X, Jamison AM, Dredze M and Quinn SC (2021) The Covid-19 social media infodemic reflects uncertainty and state-sponsored propaganda. arXiv preprint 2007.09682v1. Available at https://arxiv.org/abs/2007.09682v1.

Byman D and Kreps SE (2010) Agents of destruction? Applying principal-agent analysis to state-sponsored terrorism. *International Studies Perspectives 11*(1), 1–18.

Carter EB (2020) Diversionary cheap talk: Economic conditions and US foreign policy rhetoric, 1945-2010. *International Interactions 46*(2), 163–198.

Chung G, Rodriguez M, Lanier P and Gibbs D (2022) Text-mining open-ended survey responses using structural topic modeling: A practical demonstration to understand parents' coping methods during the Covid-19 pandemic in Singapore. *Journal of Technology in Human Services 40*(4), 296–318.

da Silva FCD, Vieira R and Garcia ACB (2019) Can machines learn to detect fake news? A survey focused on social media. Paper presented at the Hawaii International Conference on System Sciences, January 8–11, Grand Wailea, HI, USA.

Dang A, Moh'd A, Islam A and Milios E (2019) Early detection of rumor veracity in social media. Paper presented at the Hawaii International Conference on System Sciences, January 8–11, Grand Wailea, HI, USA.

Duncombe C (2019) The politics of twitter: Emotions and the power of social media. *International Political Sociology 13*(4), 409–429.

Eady G, Paskhalis T, Zilinsky J, Bonneau R, Nagler J and Tucker JA (2023) Exposure to the Russian internet research agency foreign influence campaign on twitter in the 2016 US election and its relationship to attitudes and voting behavior. *Nature Communications 14*(62). 1–11

Enterline AJ (2010) Introduction to CMPS special issue: Diversionary theory. *Conflict Management and Peace Science 27*(5), 411–416.

Faris DM and Rahimi B (2016) *Social Media in Iran: Politics and Society After 2009*. New York: SUNY Press.

Feinerer I, Hornik K and Meye D (2008) Text mining infrastructure in R. *Journal of Statistical Software 25*(5), 1–54.

Goodrich NH (2020) English in Iran. *World Englishes 39*(3), 482–499.

Grasse D, Pavlik M, Matfess H and Curtice TB (2021) Opportunistic repression: Civilian targeting by the state in response to Covid-19. *International Security 46*(2), 130–165.

Gray M (2010) *Conspiracy Theories in the Arab World: Sources and Politics*. New York: Routledge.

Hashemi LM, Wilson SL and Petrarca CS (2022) Five hundred days of Farsi Twitter: An overview of what Farsi Twitter looks like, what we know about it, and why it matters. *Journal of Quantitative Description: Digital Media* 2, 1–29.

Himelein-Wachowiak M, Giorgi S, Devoto A, Rahman M, Ungar L, Schwartz HA, Epstein DH, Leggio L and Curtis B (2021) Bots and misinformation spread on social media: Implications for Covid-19. *Journal of Medical Internet Research 23*(5), e26933.

Jafari H (2020) Infographic: Twitter usage statistics in Iran. TechRasa. Available at http://techrasa.com/2017/08/02/infographic-twitter-usage-statistics-iran/.

Kanat KB (2011) Leadership Style and Diversionary Theory of Foreign Policy: The Use of Diversionary Strategies by Middle Eastern Leaders during and in the Immediate Aftermath of the Gulf War. PhD dissertation, Department of Political Science, Syracuse University, NY, USA.

Kanat KB (2014) Diversionary foreign policy in authoritarian states: The use of multiple diversionary strategies by Saddam Hussein during the Gulf war. *Journal of Strategic Security 7*(1), 16–32.

Kermani H and Adham M (2021) Mapping Persian Twitter: Networks and mechanism of political communication in Iranian 2017 presidential election. *Big Data & Society 8*(1), 1–16.

Khazraee E (2019) Mapping the political landscape of Persian Twitter: The case of 2013 presidential election. *Big Data & Society 6*(1), 1–15.
Koehler-Derrick G, **Nielsen RA and Romney D** (2022) The supply of conspiracy theories in state-controlled media. Working Paper, Massachusetts Institute of Technology, Boston, MA, USA.
Kwon KH, **Chadha M and Wang F** (2019) Proximity and networked news public: Structural topic modeling of global Twitter conversations about the 2017 Quebec mosque shooting. *International Journal of Communication*, *13*, 24.
Lee EW, **Bao H**, **Wang Y and Lim YT** (2023) From pandemic to plandemic: Examining the amplification and attenuation of Covid-19 misinformation on social media. *Social Science & Medicine*, *328*, 115979.
Lewandowsky S, **Ecker UK**, **Seifert CM**, **Schwarz N and Cook J** (2012) Misinformation and its correction: Continued influence and successful debiasing. *Psychological Science in the Public Interest 13*(3), 106–131.
McLaughlin Mitchell S and Thyne CL (2010) Contentious issues as opportunities for diversionary behavior. *Conflict Management and Peace Science 27*(5), 461–485.
Mitternich NW (2020) Drawback before the wave? Protest decline during the Covid-19 pandemic. Open Science Foundation (OSF). Working Paper. Version 2, Department of Political Science, University College London, UK.
Moghanibashi-Mansourieh A, **Arshi M and Arabgari A** (2023) *Iran's Experience with the COVID-19 Pandemic: Focusing on Vulnerable Populations*. Singapore: Springer Nature Singapore, pp. 537–550.
Motamedi M (2020) Iran imposes nationwide Covid-19 restrictions but no lockdown. Al Jazeera. Available at https://www.aljazeera.com/news/2020/11/21/iran-imposes-nationwide-covid-19-restrictions-but-no-lockdown.
Moy W and Gradon K (2020) Covid-19 effects and Russian disinformation. *Homeland Security Affairs 16*(8), 1–27.
Nemr C and Gangware W (2019*) Weapons of Mass Destruction: Foreign State-Sponsored Disinformation in the Digital Age*. US State Department. Available at https://www.state.gov/wp-content/uploads/2019/05/Weapons-of-Mass-Distraction-Foreign-State-Sponsored-Disinformation-in-the-Digital-Age.pdf.
Oakes A (2012) *Diversionary War: Domestic Unrest and International Conflict*. Redwood City: Stanford University Press.
PBS (2020) Senate panel finds Russia interfered in the 2016 US election. Available at https://www.pbs.org/newshour/politics/senate-panel-finds-russia-interfered-in-the-2016-us-election.
Ritchie H, **Roser M and Rosado P** (2022). Energy - Iran: Coronavirus pandemic country profile. *Our World in Data.* Available at https://ourworldindata.org/energy; https://ourworldindata.org/coronavirus/country/iran.
Roberts ME, **Stewart BM and Tingley D** (2014a) stm-package {stm}. R Documentation. Available at https://cran.r-project.org/web/packages/stm/index.html (accessed 20 June 2015).
Roberts ME, **Stewart BM**, **Tingley D and Airoldi EM** (2013) The structural topic model and applied social science. Paper presented at Advances in Neural Information Processing Systems Workshop on Topic Models: Computation, Application, and Evaluation, Lake Tahoe, NV, USA, pp. 1–2.
Roberts ME, **Stewart BM**, **Tingley D**, **Lucas C**, **Leder-Luis J**, **Gadarian S**, **Albertson B and Rand DG** (2014b) Structural topic models for open ended responses. *American Journal of Political Science 58*(4), 1064–1082.
Rodriguez MY and Storer H (2020) A computational social science perspective on qualitative data exploration: Using topic models for the descriptive analysis of social media data*. *Journal of Technology in Human Services*, *38*(1), 54–86.
Ross AS and Rivers DJ (2018) Discursive deflection: Accusation of "fake news" and the spread of mis-and disinformation in the tweets of president trump. *Social Media+ Society 4*(2), 2056305118776010.
Rozenas A and Stukal D (2019) How autocrats manipulate economic news: Evidence from Russia's state-controlled television. *The Journal of Politics 81*(3), 982–996.
Rulis, Mina. 2024. "The Influences of Misinformation on Incidences of Politically Motivated Violence in Europe." *The International Journal of Press/Politics*, *0*(0), 1–20.
Salehyan I, **Siroky D and Wood RM** (2014) External rebel sponsorship and civilian abuse: A principal-agent analysis of wartime atrocities. *International Organization 68*(3), 633–661.
Sciubba Caniglia C (2020) Signs of a new world order: Italy as the Covid-19 disinformation. *The Harvard Kennedy School Misinformation Review 1*(3), 1–8.
Smith A (1996) Diversionary foreign policy in democratic systems. *International Studies Quarterly 40*(1), 133–153.
Steinert-Threlkeld ZC (2017) Spontaneous collective action: Peripheral mobilization during the Arab spring. *American Political Science Review 111*(2), 379–403.
Stukal D, **Sanovich S**, **Bonneau R and Tucker JA** (2022) Why botter: How pro-government bots fight opposition in Russia. *American Political Science Review 116*(3), 843–857.
Tan SD and Tas O (2021) Social media sentiment in international stock returns and trading activity. *Journal of Behavioral Finance 22*(2), 221–234.
Twitter (2021) Disclosing state-linked information operations we've removed. Available at https://blog.twitter.com/en_us/topics/company/2021/disclosing-state-linked-information-operations-we-ve-removed.
Twitter (2023) Platform manipulation and spam policy. Available at https://help.twitter.com/en/rules-and-policies/platform-manipulation.
Valle-Cruz D, **Fernandez-Cortez V**, **López-Chau A and Sandoval-Almazán R** (2022) Does Twitter affect stock market decisions? Financial sentiment analysis during pandemics: A comparative study of the H1N1 and the Covid-19 periods. *Cognitive Computation 14*, 372–387.
Venkatesan P (2020) Covid-19 in Iran: Round 2. *Lancet Infectious Diseases 20*(7), 784.

https://doi.org/10.1017/dap.2024.96 Published online by Cambridge University Press

Verrall N (2022) Covid-19 disinformation, misinformation and malinformation during the pandemic infodemic: A view from the United Kingdom. In Gill R, Goolsby R (eds), *COVID-19 Disinformation: A Multi-National, Whole of Society Perspective. Advanced Sciences and Technologies for Security Applications*. Cham: Springer, pp. 81–112.

Wu L, **Morstatter F**, **Carley KM and Liu H** (2019) Misinformation in social media: Definition, manipulation, and detection. *ACM SIGKDD Explorations Newsletter 21*(2), 80–90.

Yeh Y-Y and Wu CK (2020) Diversionary behavior for weak states: A case study of Taiwan. *Journal of Asian and African Studies 55*(2), 221–234.

Yini Zhang FC and Lukito J (2023) Network amplification of politicized information and misinformation about Covid-19 by conservative media and partisan influencers on Twitter. *Political Communication 40*(1), 24–47.

Zeitzoff T (2018) Does social media influence conflict? Evidence from the 2012 Gaza conflict. *Journal of Conflict Resolution 62*(1), 29–63.

Zhang Y and Hara T (2020) A probabilistic model for malicious user and rumor detection on social media. Paper presented at the Hawaii International Conference on System Sciences, University of Hawaii at Manoa, HI, USA.

Ziabari K (2023) Tweeting is banned in Iran, but not for the regime's supporters. *New Lines Magazine.* Available at https://newlinesmag.com/spotlight/tweeting-is-banned-in-iran-but-not-for-the-regimes-supporters/.

Cite this article: Bagozzi BE, Balasubramanian K, Goel R and Parker C (2025). State-linked manipulated media in the time of Covid-19: a look at Iran. *Data & Policy*, 7: e19. doi:10.1017/dap.2024.96