

Trends in the Diffusion of Misinformation on Social Media

Hunt Allcott, Matthew Gentzkow, and Chuan Yu

Online Appendix

I Actions Against Misinformation

Both Facebook and Twitter have taken steps to reduce the circulation of misinformation on their platforms. Online Appendix Table 1 lists twelve such announcements by Facebook since the 2016 election, while Online Appendix Table 2 lists five by Twitter. All announcements are taken from the platforms’ official websites.¹ Broadly, the platforms have taken three types of actions to limit misinformation. First, they have limited its supply, by blocking ads from pages that repeatedly share false content and removing accounts that violate community standards. Second, they have introduced features such as “disputed” flags or “related articles” that provide corrective information related to a false story. Third, they have changed their algorithms to de-prioritize false stories in favor of news from trustworthy publications and posts from friends and family. In addition to their actions targeting misinformation, Facebook has taken a broader range of steps to reduce the circulation of clickbait and sensationalistic news (even if not false), and these actions may also affect the circulation of content from our list of fake news sites.

II Data

We combine five lists of fake news sites created by fact-checking organizations or research studies to form our sample of fake news sites. The union of these lists contains 673 unique sites. Among them, 103 have no data available from BuzzSumo. Thus, our final list includes 570 unique sites. Online Appendix Table 3 presents the 50 largest sites in the list in terms of total Facebook engagements plus Twitter shares from January 2015 to July 2018.

We also collect three other categories of sites: major news sites, small news sites, and business and culture sites covering arts, business, health, recreation, and sports. Online Appendix Table 4 presents these lists.

III Robustness Checks

As discussed in the manuscript, a key concern is that our list of fake news sites may suffer from sample selection bias. To mitigate this concern, we consider alternative sets of fake news sites as robustness checks.

¹Facebook: <https://newsroom.fb.com/news/>; Twitter: https://blog.twitter.com/official/en_us.html.

III.A Lists of Fake News Sites

Our five different lists each have different inclusion criteria for defining a fake news site, and one might disagree with a particular list’s approach. We thus carry out two sets of robustness checks. First, in Online Appendix Figure 1, we focus on sites that are identified as fake news sites by at least two or three lists instead of one, leaving 116 and 19 sites, respectively. Second, Online Appendix Figure 2 replicates the results using sites assembled from any four out of the five original lists. By doing this, we exclude sites that are only identified by one particular list. The downward trend in the ratio of Facebook engagements to Twitter shares since the beginning of 2017 is invariant to including only sites identified on multiple lists and to excluding any single list.

III.B Time Coverage

It is possible that the original lists of fake news sites primarily include sites that were popular on Facebook before the 2016 election, and this sample selection combined with the rapid entry and exit of small sites focused on false stories could generate a spurious downward trend in both the absolute number of Facebook engagements of fake news and the ratio of Facebook engagements to Twitter shares. In Online Appendix Figure 3, we look at sites that were active during different periods. In Panel A, B, and C, respectively, we focus on sites that started active operation after November 2016, sites that were still in active operation as of the end of the sample in July 2018, and sites that were in active operation from August 2015 to July 2018. (Active operation is defined to be a global traffic rank reported by Alexa of at least one million.) The downward trend in the ratio of Facebook engagements to Twitter shares since the beginning of 2017 remains consistent across these samples.

III.C Number of Interactions

Interactions on social media vary substantially across sites in our list. A natural concern might be that the sums of Facebook engagements and Twitter shares could be driven by a small number of outliers. In Online Appendix Figure 4 Panel A, we exclude the five largest sites in terms of total Facebook engagements plus Twitter shares in our sample period. The trend survives the exclusion of potential outliers. In Panel B and C, we divide all sites into deciles and look at sites in the first decile and sites in the bottom nine deciles separately. The downward trend in the Facebook/Twitter ratio is observed for both large and small sites.

III.D Likelihood to Publish Misinformation

Grinberg et al. (2018) provide three lists of sites which they deem to have different likelihoods to publish misinformation. “Black” domains are reported to publish entirely fabricated stories. The black list is constructed from pre-existing lists of fake news sites constructed by academic work and professional fact-checkers such as PolitiFact, FactCheck, and BuzzFeed. “Red” and “orange” domains are identified by Snopes as sources of false stories or questionable claims

and classified by their levels of perceived likelihood to publish misinformation: stories from red domains have an “extremely high” likelihood of containing misinformation, and stories from orange domains a “high” likelihood. In Online Appendix Figure 5, we look at these lists separately. There are some differences across these lists. The downward trend of Facebook engagements appears only for black and red domains but not for orange domains. The point in time when the Facebook/Twitter ratio begins to decline is also different. For black domains, the ratio drops sharply in mid-2016 and all of 2017. For red and orange domains, however, the decline primarily occurs in 2016. These patterns would be consistent with black and to a lesser extent red domains being the primary target of the changes that Facebook made to its platform following the election.

III.E Sites Focusing on Political News

False stories are often political in nature, and it is possible that the Twitter user base is more consistently politically engaged than the (much larger) Facebook user base. If Facebook users’ interest in political stories is cyclical, rising with major presidential elections and falling after, this could generate a drop in diffusion of false stories on Facebook after the 2016 election that might not be mirrored on Twitter. Thus, the declining Facebook/Twitter ratio beginning in 2017 could be generated by changes in demand for false stories, not changes in supply or efforts by Facebook.

If this explanation is true, one would also expect to see a decline in the diffusion of articles from major political websites on Facebook, but not on Twitter. To test this, Online Appendix Figure 6 presents Facebook engagements, Twitter shares, and their ratio for a list of ten (non-fake) political sites of five types: (i) sites mostly focusing on political news (Politico and The Hill); (ii) major parties and politicians (donaldjtrump.com, hillaryclinton.com, democrats.org, and gop.com); (iii) think tanks (Brookings and AEI); (iv) CSPAN; and (v) a mainstream political blog (Real Clear Politics). There is a decline in the Facebook/Twitter ratio for these sites, but it mainly occurs in late-2015, well before the election.

Online Appendix Table 1: Facebook’s Actions to Fight Against Fake News

Date	Actions
Dec 15, 2016	Announced four updates to address hoaxes and fake news: make reporting easier for users; flag stories as “Disputed” with fact-checking organizations and warn people before they share; incorporate signals of misleading articles into rankings; and disrupt financial incentives for spammers. ²
Apr 6, 2017	Described three areas where it is working to fight the spread of false news: disrupt economic incentives; build new products to curb the spread of false news; and help people make more informed decisions. ³
Apr 25, 2017	Tested “Related Articles”, an improved feature that presents users a cluster of additional articles on the same topic when they come across popular links, including potential fake news articles, to provide people easier access to additional information, including articles by third-party fact checkers. ⁴
Aug 8, 2017	Announced it would address cloaking so people see more authentic posts. ⁵
Aug 28, 2017	Announced it would block ads from pages that repeatedly share false news. ⁶
Dec 20, 2017	Announced two changes to fight against false news: replace “Disputed” flags with “Related Articles” to give people more context; and start an initiative to better understand how people decide whether information is accurate. ⁷
Jan 11, 2018	Prioritized posts from friends and family over public content. ⁸
Jan 19, 2018	Prioritized news from publications rated as trustworthy by the community. ⁹
Jan 29, 2018	Prioritized news relevant to people’s local community. ¹⁰
May 23, 2018	Described three parts of their strategies to stop misinformation: remove accounts and content that violate community standards or ad policies; reduce the distribution of false news and inauthentic content; and inform people by giving them more context on the posts they see. ¹¹
June 14, 2018	Detailed how its fact-checking program works. ¹²
June 21, 2018	Announced five updates to fight false news: expand fact-checking programs to new countries; test fact-checking on photos and videos; use new techniques in fact-checking including identifying duplicates and using “Claim Review”; take action against repeat offenders; and improve measurement and transparency by partnering with academics. ¹³

2.Addressing Hoaxes and Fake News.

3.Working to Stop Misinformation and False News.

4.New Test With Related Articles.

5.Addressing Cloaking So People See More Authentic Posts.

6.Blocking Ads From Pages that Repeatedly Share False News.

7.Replacing Disputed Flags With Related Articles.

8.Bringing People Closer Together.

9.Helping Ensure News on Facebook Is From Trusted Sources.

10.More Local News on Facebook.

11.Hard Questions: What’s Facebook’s Strategy for Stopping False News?

12.Hard Questions: How Is Facebook’s Fact-Checking Program Working?

13.Increasing Our Efforts to Fight False News.

Online Appendix Table 2: Twitter’s Actions to Fight Against Fake News

Date	Actions
June 14, 2017	Described the phenomenon of fake news and bots and the approaches it used, including surfacing the highest quality and most relevant content and context first, expanding the team and resources, building new tools and processes, and detecting spammy behaviors at source. ¹⁴
June 29, 2017	(Not officially announced) Tested a feature that would let users flag tweets that contain misleading, false, or harmful information. ¹⁵
Sept 28, 2017	Shared information on its knowledge about how malicious bots and misinformation networks on Twitter may have been used in the 2016 U.S. Presidential elections and its work to fight both malicious bots and misinformation. ¹⁶
Oct 24, 2017	Announced steps to dramatically increase the transparency for all ads. ¹⁷
July 11, 2018	Announced it removed fake accounts. ¹⁸

14.Our Approach to Bots & Misinformation.

15.Twitter is looking for ways to let users flag fake news, offensive content.

16.Update: Russian Interference in 2016 US Election, Bots, & Misinformation.

17.New Transparency For Ads on Twitter.

18.Confidence in follower counts.

Online Appendix Table 3: 50 Largest Fake News Sites

Site	Source							Created	Still	Last
	G-B	G-R	G-O	PF	BF	GNR	FC	Post-Election	Active	Long
indiatimes.com	0	0	0	1	0	0	0	0	1	1
dailywire.com	0	0	1	0	0	1	0	0	1	0
ijr.com	0	0	0	0	0	1	0	0	1	0
dailycaller.com	0	0	1	0	0	0	0	0	1	1
occupydemocrats.com	1	0	0	0	0	0	0	1	1	0
express.co.uk	0	0	1	0	0	0	0	0	1	1
redstatewatcher.com	1	0	0	0	0	1	0	0	1	0
thepoliticalinsider.com	0	0	0	1	0	0	0	0	1	1
thefederalistpapers.org	0	0	1	0	0	0	0	0	1	1
truthfeed.com	0	1	0	0	0	1	0	0	1	0
bipartisanreport.com	0	1	0	0	0	1	0	0	1	0
rightwingnews.com	0	0	1	0	0	0	0	0	1	1
qpolitical.com	0	0	1	0	0	0	0	0	0	0
madworldnews.com	1	0	0	1	0	0	0	0	1	1
yournewswire.com	1	0	0	1	1	0	1	0	1	1
uschronicle.com	0	0	1	0	0	0	0	0	1	0
louderwithcrowder.com	0	1	0	0	0	0	0	0	1	1
jewsnews.co.il	0	0	0	1	0	0	0	0	1	1
100percentfedup.com	0	1	0	0	0	0	0	0	1	1
angrypatriotmovement.com	1	0	0	1	0	1	0	0	1	0
anohq.com	0	1	0	0	0	0	0	0	1	1
inquisitr.com	0	0	1	0	0	0	0	0	1	1
yesimright.com	1	0	0	0	0	1	0	0	0	0
worldtruth.tv	0	1	0	1	0	0	0	0	1	1
collective-evolution.com	0	1	0	0	0	0	0	0	1	1
ilovemyfreedom.org	1	0	0	0	0	1	0	0	1	0
tribunist.com	0	0	1	0	0	0	0	0	1	0
clashdaily.com	1	0	0	1	0	0	0	0	1	1
naturalnews.com	0	1	0	0	0	0	0	0	1	1
joeforamerica.com	0	0	1	0	0	0	0	0	1	1
conservativedailyreport.com	1	0	0	1	0	1	0	0	1	0
worldnewsdailyreport.com	1	0	0	1	1	0	0	0	1	1
trueactivist.com	0	0	1	0	0	0	0	0	1	1
americasfreedomfighters.com	0	1	0	0	0	0	0	0	1	1

Site	Source							Created	Still	Last
	G-B	G-R	G-O	PF	BF	GNR	FC	Post-Election	Active	Long
conservative101.com	0	0	0	1	0	0	0	0	1	0
usanewsflash.com	1	0	0	1	0	1	0	0	0	0
babylonbee.com	0	0	0	1	0	0	0	0	1	0
firstpost.com	0	0	0	1	0	0	0	0	1	1
zerohedge.com	0	0	1	0	0	0	0	0	1	1
teaparty.org	0	0	0	1	0	0	0	0	1	1
palmerreport.com	0	0	1	0	0	0	0	1	1	0
judicialwatch.org	0	1	0	0	0	0	0	0	1	1
disclose.tv	1	0	0	1	0	0	0	0	1	1
conservativepost.com	0	1	0	0	0	0	0	0	1	1
thegatewaypundit.com	0	1	0	1	0	0	0	0	1	1
infowars.com	0	1	0	0	0	0	0	0	1	1
dailysnark.com	0	0	0	1	0	0	0	0	1	1
postcard.news	0	0	0	1	0	0	0	0	1	0
higherperspectives.com	0	0	0	1	0	0	0	0	1	0
tmn.today	0	0	1	0	0	1	0	0	1	0
...										
Total	382	61	47	325	223	92	61	308	287	82

Notes: This table lists 50 largest fake news sites in terms of total Facebook engagements plus Twitter shares from January 2015 to June 2018. The complete list can be found [here](#). Column 2-8 lists the fake news sites identified by five sources described above, where a value of 1 indicates the site appears in the corresponding source and 0 not. **G-B**, **G-R**, and **G-O** represent the black domains, red domains, and orange domains in Grinberg et al. (2018). **PF** represents PolitiFact. **BF** represents BuzzFeed. **GNR** represents Guess et al. (2018). **FC** represents FactCheck. The last three columns list sites that started active operation after the election in November 2016, sites that were in active operation in July 2018, and sites that were in active operation during the whole sample period from August 2015 to July 2018. A site is defined as being in active operation if it is tracked by Alexa with a global rank higher than one million in terms of total traffic.

Online Appendix Table 4: Lists of Sites in Each Category

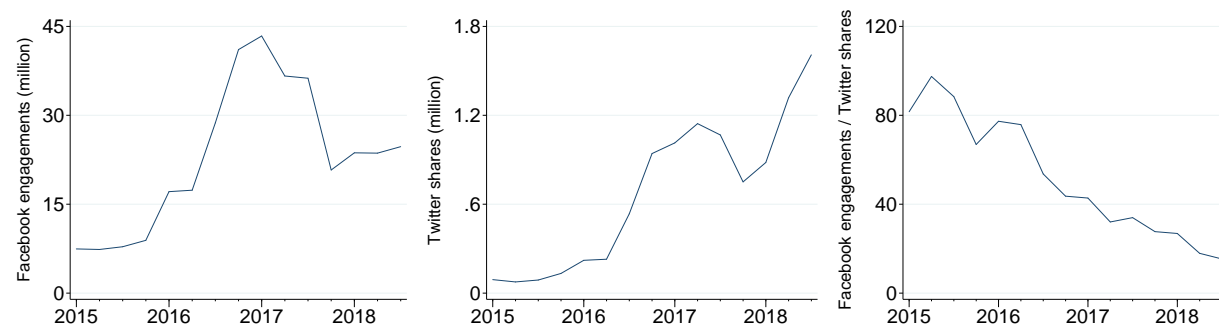
Category	Site		
<i>Major News Sites</i>	cnn.com	nytimes.com	theguardian.com
	washingtonpost.com	foxnews.com	huffingtonpost.com
	usatoday.com	wsj.com	cnbc.com
	reuters.com	time.com	nypost.com
	usnews.com	cbsnews.com	chron.com
	thehill.com	nbcnews.com	theatlantic.com
	latimes.com	abcnews.go.com	thedailybeast.com
	sfgate.com	newsweek.com	chicagotribune.com
	economist.com	theroot.com	voanews.com
	nj.com	miamiherald.com	mercurynews.com
	bostonglobe.com	seattletimes.com	oregonlive.com
	washingtontimes.com	azcentral.com	ajc.com
	philly.com	sacbee.com	
<i>Small News Sites</i>	aspentimes.com	bakersfield.com	bendbulletin.com
	bnd.com	broadcastingcable.com	charlestoncitypaper.com
	chicagomaroon.com	collegian.psu.edu	columbian.com
	dailynebraskan.com	dailynexus.com	dailynorthwestern.com
	dailypress.com	dailyprogress.com	dailytexanonline.com
	dailytrojan.com	dcourier.com	delcotimes.com
	durangoherald.com	fair.org	fredericksburg.com
	globegazette.com	greenvilleonline.com	greenwichtime.com
	havasunews.com	hcn.org	heraldnet.com
	heraldsun.com	heraldtimesonline.com	ibj.com
	independent.com	islandpacket.com	jou.ufl.edu
	journalism.org	journalismjobs.com	journaltimes.com
	kitv.com	knoxnews.com	lacrossetribune.com
	leadertelegram.com	macon.com	myrtlebeachonline.com
	naplesnews.com	nashvillescene.com	news.cornell.edu
	news.usc.edu	newseum.org	news-journalonline.com
	news-leader.com	newstimes.com	nwfdailynews.com
	pjstar.com	presstelegram.com	rapidcityjournal.com
	readingeagle.com	redandblack.com	rgj.com
	sacurrent.com	santacruzsentinel.com	santafenewmexican.com
	sgvtribune.com	signalscv.com	siouxcityjournal.com
	standard.net	stanforddaily.com	steynonline.com
	studlife.com	tallahassee.com	theday.com
	theeagle.com	theledger.com	timesleader.com

Category	Site		
<i>Small News Sites</i>	ubm.com	vcstar.com	wacotrib.com
	wfcourier.com	wvgazetteemail.com	yakimaherald.com
<i>Arts</i>	imdb.com	ign.com	rottentomatoes.com
	ultimate-guitar.com	npr.org	vice.com
	tmz.com	pitchfork.com	wired.com
<i>Business</i>	forbes.com	shutterstock.com	businessinsider.com
	finance.yahoo.com	bloomberg.com	eventbrite.com
	fortune.com	adweek.com	
<i>Health</i>	webmd.com	psychologytoday.com	who.int
	apa.org	bmj.com	mercola.com
	menshealth.com	self.com	nejm.org
<i>Recreation</i>	9gag.com	jalopnik.com	timeout.com
	lonelyplanet.com	caranddriver.com	hollywoodreporter.com
	nationalgeographic.com	rd.com	topix.com
<i>Sports</i>	espn.com	cricbuzz.com	nba.com
	espnricinfo.com	sports.yahoo.com	bleacherreport.com
	nhl.com	cbssports.com	nfl.com
	iplt20.com	skysports.com	deadspin.com
	nbcsports.com	wwe.com	si.com
	sbnation.com	formula1.com	rivals.com
	foxsports.com		

Notes: This table lists sites in the comparison groups. *Major News Sites* include 38 sites selected from the top 100 sites in Alexa’s News category. *Small News Sites* include 78 sites selected from the sites ranking 401-500 in the News category. *Business and Culture Sites* include 54 sites selected from the top 50 sites in each of the Arts, Business, Health, Recreation, and Sports categories. For each group, we omit from our sample government websites, databases, sites that do not mainly produce news or similar content, international sites whose audiences are primarily outside the U.S., and sites that are included in our list of fake news sites.

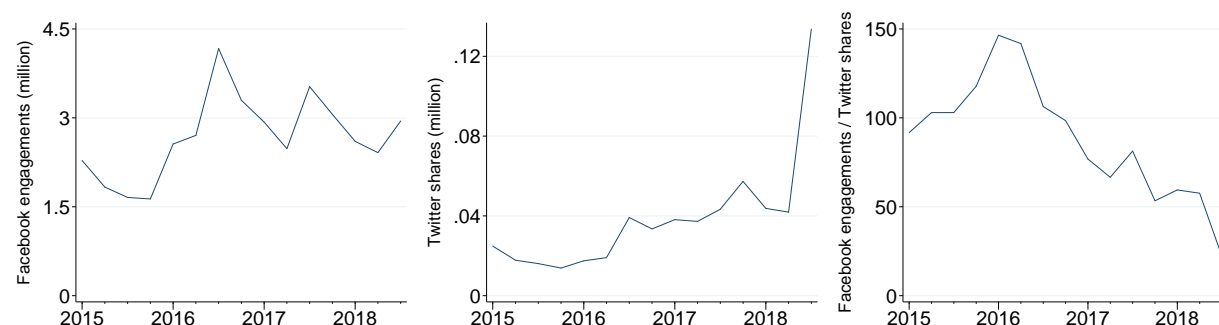
Online Appendix Figure 1: Robustness Checks of Fake News Sites - Multiple Lists

Panel A: Sites Identified by At Least Two Lists



Number of sites: 116

Panel B: Sites Identified by At Least Three Lists

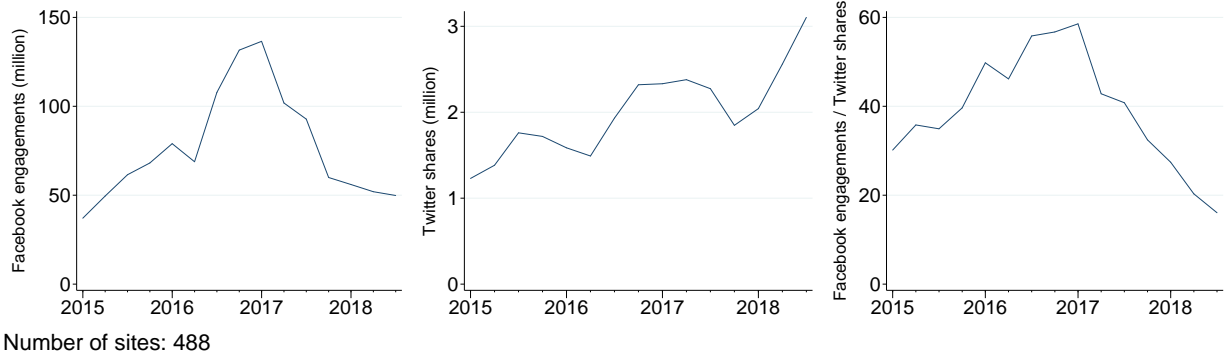


Number of sites: 19

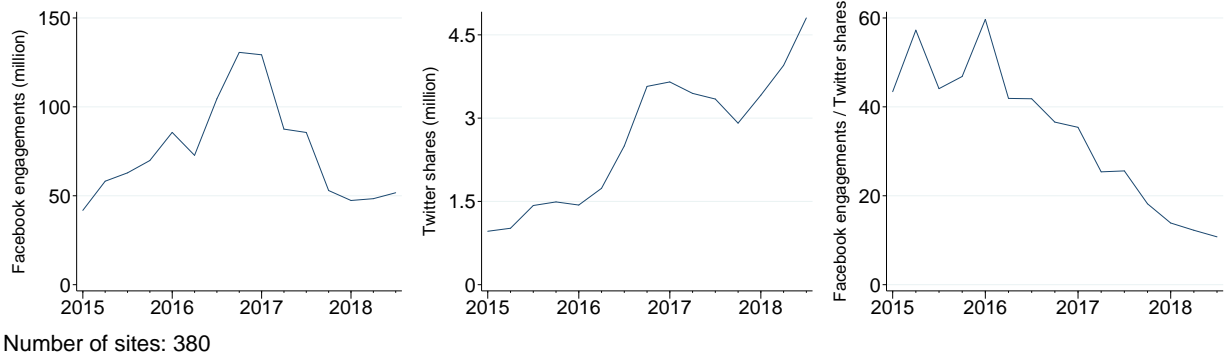
Notes: This figure plots robustness checks for the sample of fake news sites by looking at sites only identified by multiple lists. Each panel plots monthly Facebook engagements, Twitter shares, and the ratio of Facebook engagements to Twitter shares averaged by quarter. Panel A includes sites identified by at least two lists out of five. Panel B includes sites identified by at least three lists. Grinberg et al.'s (2018) provide three types of domains. The black domains derive from lists that we already use (with the exception of nine sites, as PolitiFact and FactCheck updated their lists at some point). We avoid double-counting black domains when we count the number of lists that identify a fake news site.

Online Appendix Figure 2: Robustness Checks of Fake News Sites - Excluding Lists

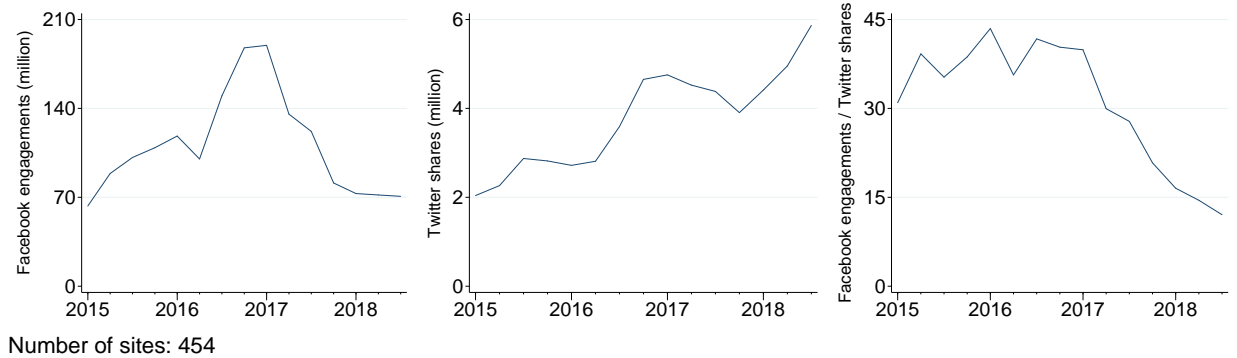
Panel A: Excluding Sites Only Identified by Grinberg et al. (2018)



Panel B: Excluding Sites Only Identified by PolitiFact

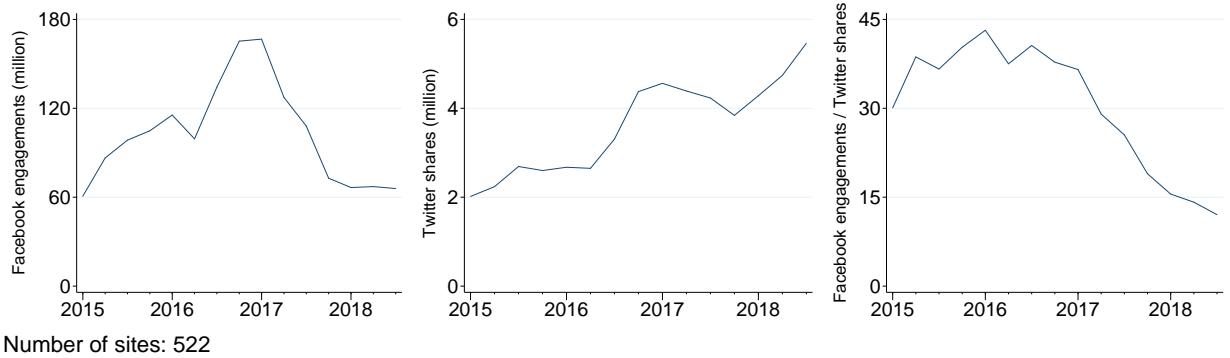


Panel C: Excluding Sites Only Identified by BuzzFeed

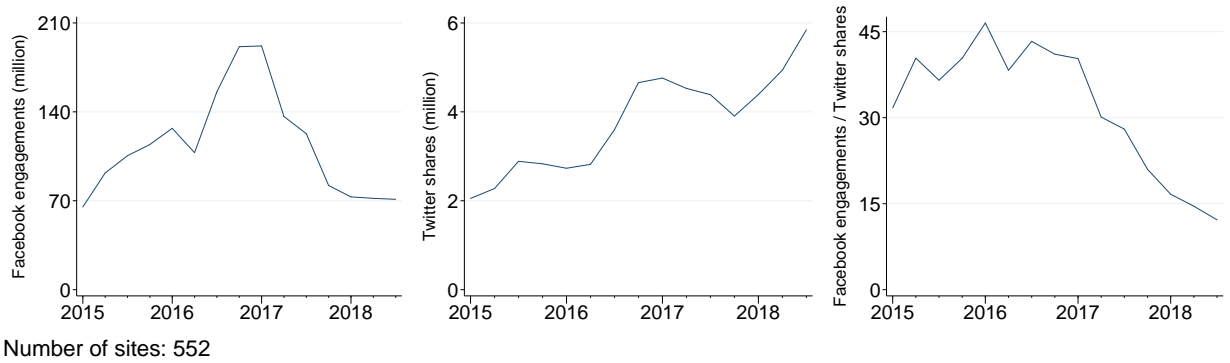


Online Appendix Figure 2: Robustness Checks of Fake News Sites - Excluding Lists (*continued*)

Panel D: Excluding Sites Only Identified by Guess et al. (2018)



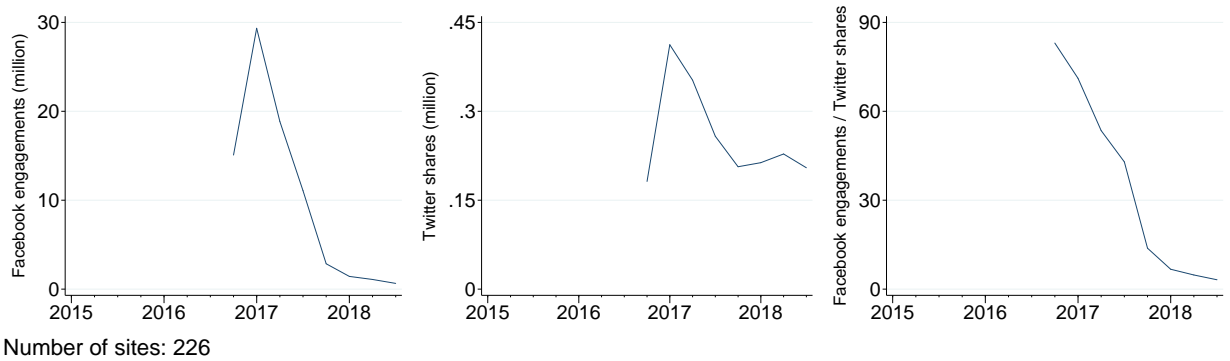
Panel E: Excluding Sites Only Identified by FactCheck



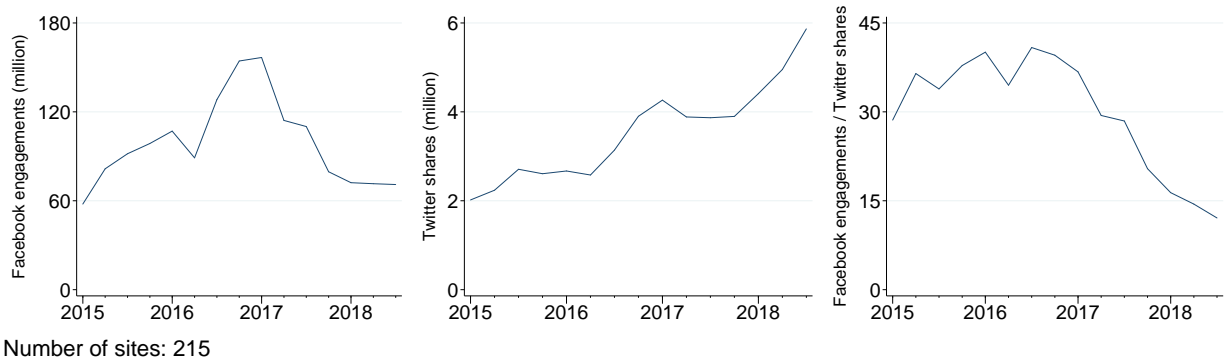
Notes: This figure plots robustness checks for the sample of fake news sites by excluding sites only identified by a particular list. Each panel plots monthly Facebook engagements, Twitter shares, and the ratio of Facebook engagements to Twitter shares averaged by quarter. Panel A excludes sites only identified by Grinberg et al. (2018). Panel B excludes sites only identified by PolitiFact. Panel C excludes sites only identified by BuzzFeed. Panel D excludes sites only identified by Guess et al. (2018). Panel E excludes sites only identified by FactCheck.

Online Appendix Figure 3: Robustness Checks of Fake News Sites - Time Coverage

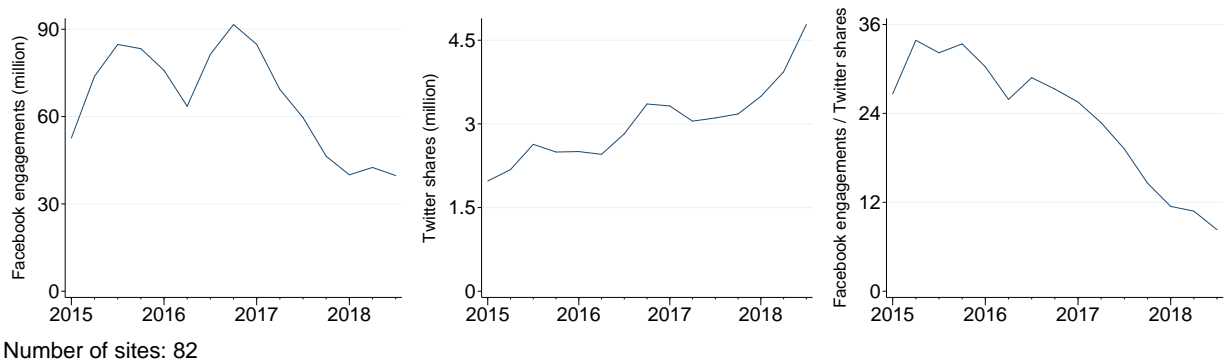
Panel A: Sites that Started Active Operation after November 2016



Panel B: Sites that were in Active Operation in July 2018



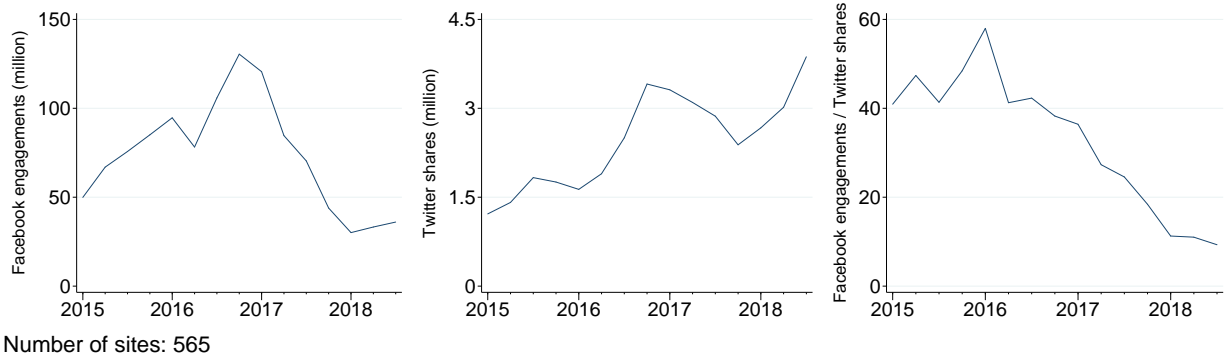
Panel C: Sites that were in Active Operation during August 2015 to July 2018



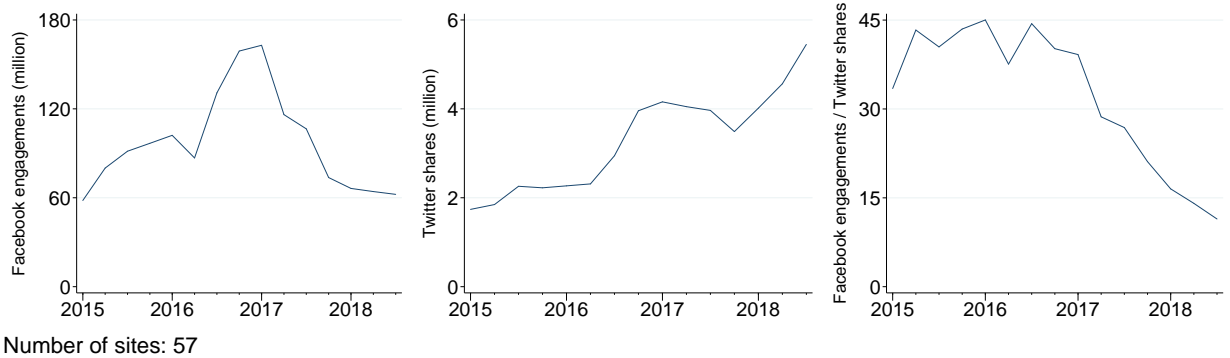
Notes: This figure plots robustness checks for the sample of fake news sites by looking at sites active in different period. Each panel plots monthly Facebook engagements, Twitter shares, and the ratio of Facebook engagements to Twitter shares averaged by quarter. Panel A includes sites that started active operation after the election in November 2016. Panel B includes sites that were still in active operation in July 2018. Panel C includes sites that were in active operation during August 2015 to July 2018. A site is defined as being in active operation if it is tracked by Alexa with a global rank higher than one million in terms of total traffic.

Online Appendix Figure 4: Robustness Checks of Fake News Sites - Number of Interactions

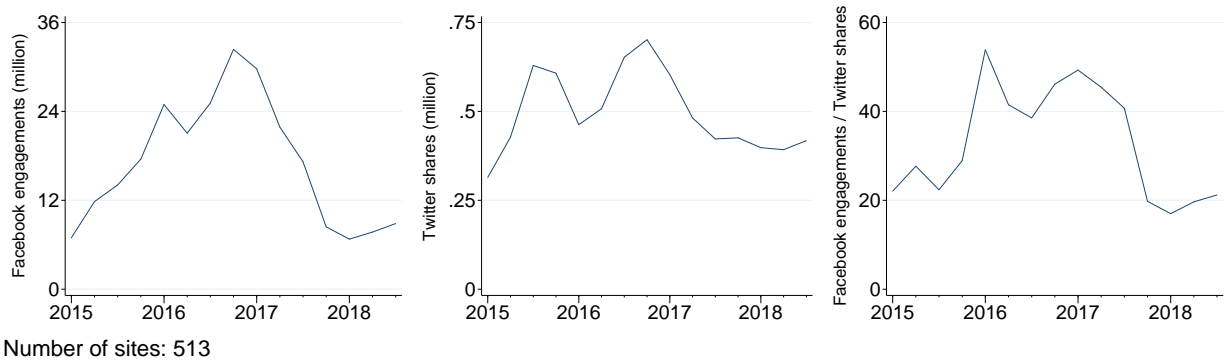
Panel A: Excluding Top Five Fake News Sites



Panel B: The First Decile of Fake News Sites

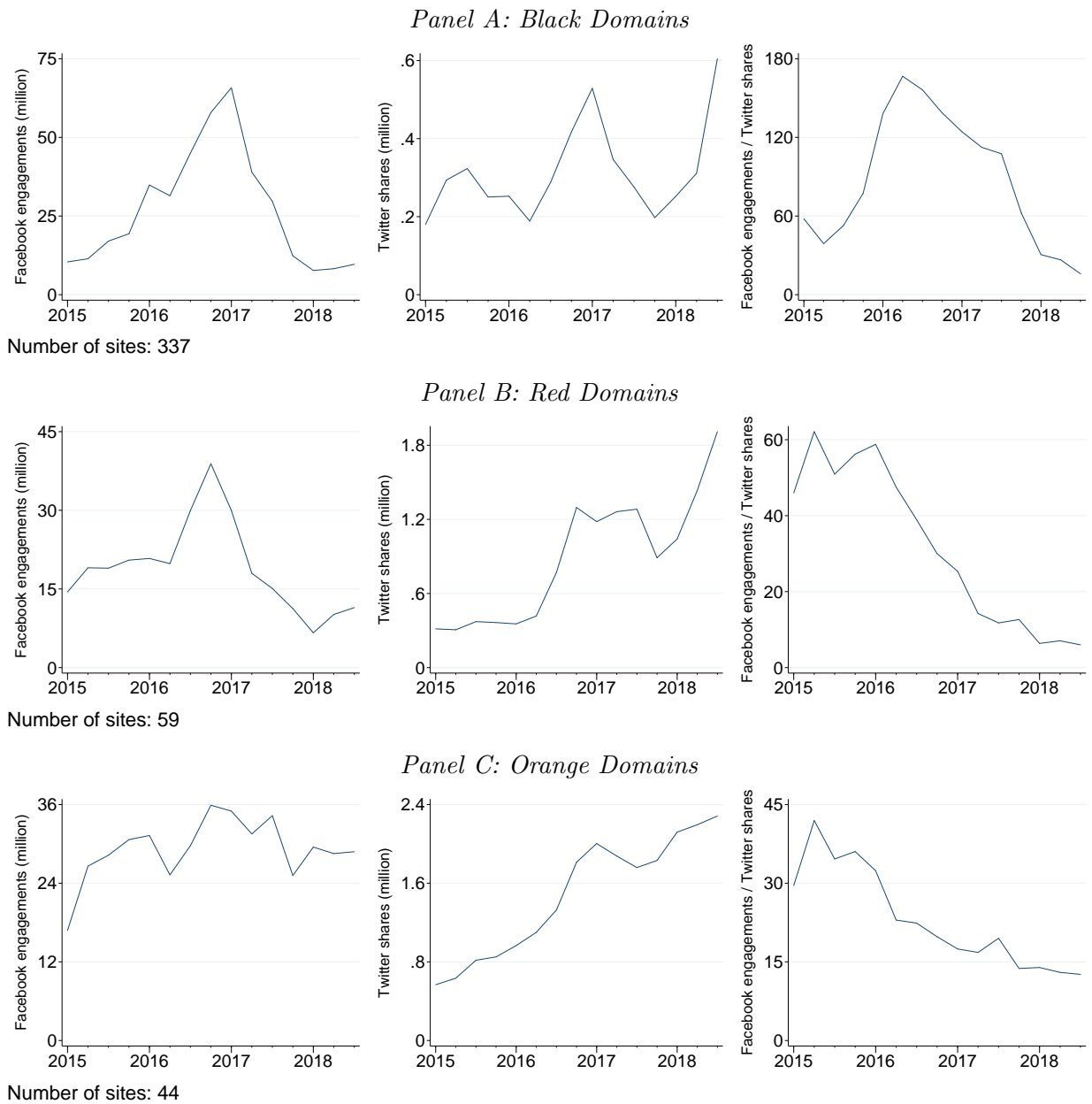


Panel C: The Bottom Nine Deciles of Fake News Sites



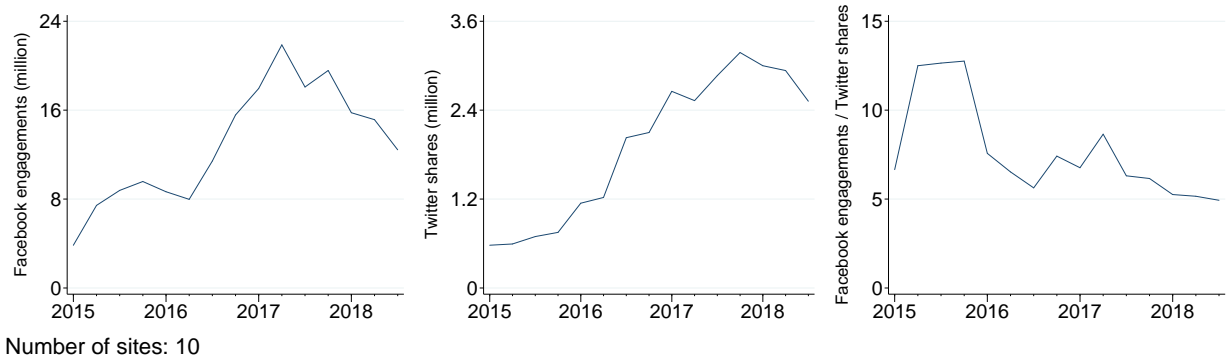
Notes: This figure plots robustness checks for the sample of fake news sites by excluding the largest sites and looking at sites of different sizes. Each panel plots monthly Facebook engagements, Twitter shares, and the ratio of Facebook engagements to Twitter shares averaged by quarter. Panel A excludes five largest sites in terms of total Facebook engagements plus Twitter shares from January 2015 to July 2018. Panel B includes sites in the first decile. Panel C includes sites in the bottom nine deciles, i.e., excludes sites in the first decile. The deciles are also defined in terms of total Facebook engagements plus Twitter shares during the sample period.

Online Appendix Figure 5: Robustness Checks of Fake News Sites - Likelihood to Publish Misinformation



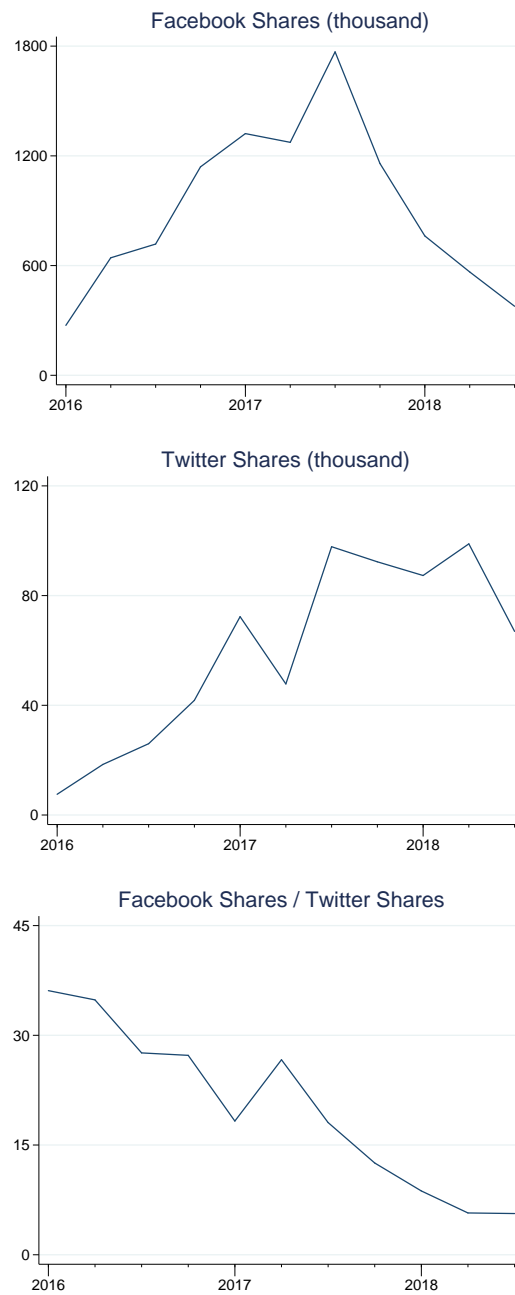
Notes: This figure plots robustness checks for three lists of fake news sites in Grinberg et al. (2018) separately, classified by their likelihood to publish misinformation. Each panel plots monthly Facebook engagements, Twitter shares, and the ratio of Facebook engagements to Twitter shares averaged by quarter. The black domains were reported to have published entirely fabricated stories, taken from pre-existing lists of fake news constructed by the fact-checking and journalistic outlets Politifact, FactCheck, and BuzzFeed, as well as domains used in other academic work. The red and orange domains are identified by Snopes as sources of fake news or questionable claims and classified by the authors by their levels of perceived likelihood to publish misinformation: stories from red domains have an extremely high likelihood of containing misinformation, and stories from orange domains a high likelihood.

Online Appendix Figure 6: Robustness Checks of Sites Focusing on Political News



Notes: This figure plots monthly Facebook engagements, Twitter shares, and the ratio of Facebook engagements to Twitter shares averaged by quarter of sites mostly focusing on political news. The sites include politico.com, thehill.com, brookings.edu, aei.org, c-span.org, realclearpolitics.com, donaldjtrump.com, hillaryclinton.com, democrats.org, and gop.com.

Online Appendix Figure 7: Shares on Facebook and Twitter for Fake News URLs



Notes: This figure shows Facebook shares, Twitter shares, and the ratio of Facebook shares to Twitter shares of a set of 10,240 URLs spreading misinformation. See the Data section in the paper for details on how the set of URLs is constructed. We sum the Facebook shares and Twitter shares of all URLs by month and average by quarter.