

## RESEARCH ARTICLE

# Computational timeline reconstruction of the stories surrounding Trump: Story turbulence, narrative control, and collective chronopathy

Peter Sheridan Dodds<sup>1,2\*</sup>, Joshua R. Minot<sup>1</sup>, Michael V. Arnold<sup>1</sup>, Thayer Alshaabi<sup>1</sup>, Jane Lydia Adams<sup>1</sup>, Andrew J. Reagan<sup>3</sup>, Christopher M. Danforth<sup>1,4</sup>

**1** Computational Story Lab, Vermont Complex Systems Center, MassMutual Center of Excellence for Complex Systems and Data Science, Vermont Advanced Computing Core, University of Vermont, Burlington, VT, United States of America, **2** Department of Computer Science, University of Vermont, Burlington, VT, United States of America, **3** Department of Mathematics & Statistics, MassMutual Data Science, Boston, MA, United States of America, **4** Department of Mathematics & Statistics, University of Vermont, Burlington, VT, United States of America

\* [peter.dodds@uvm.edu](mailto:peter.dodds@uvm.edu)



## OPEN ACCESS

**Citation:** Dodds PS, Minot JR, Arnold MV, Alshaabi T, Adams JL, Reagan AJ, et al. (2021) Computational timeline reconstruction of the stories surrounding Trump: Story turbulence, narrative control, and collective chronopathy. *PLoS ONE* 16(12): e0260592. <https://doi.org/10.1371/journal.pone.0260592>

**Editor:** Alexandre Bovet, University of Oxford, UNITED KINGDOM

**Received:** February 3, 2021

**Accepted:** November 12, 2021

**Published:** December 8, 2021

**Copyright:** © 2021 Dodds et al. This is an open access article distributed under the terms of the [Creative Commons Attribution License](https://creativecommons.org/licenses/by/4.0/), which permits unrestricted use, distribution, and reproduction in any medium, provided the original author and source are credited.

**Data Availability Statement:** Twitter's Terms of Service prohibit sharing the raw data. A similar size sample from their API will reproduce the results reported here. In order to reproduce the results of our study, researchers can query Twitter's API for English language messages matching the word "trump" starting on January 1 2015. The documentation for access to this API can be found here: <https://developer.twitter.com/en/docs/twitter-api/early-access> In addition, there are several software packages enabling this sort of query

## Abstract

Measuring the specific kind, temporal ordering, diversity, and turnover rate of stories surrounding any given subject is essential to developing a complete reckoning of that subject's historical impact. Here, we use Twitter as a distributed news and opinion aggregation source to identify and track the dynamics of the dominant day-scale stories around Donald Trump, the 45th President of the United States. Working with a data set comprising around 20 billion 1-grams, we first compare each day's 1-gram and 2-gram usage frequencies to those of a year before, to create day- and week-scale timelines for Trump stories for 2016–2021. We measure Trump's narrative control, the extent to which stories have been about Trump or put forward by Trump. We then quantify story turbulence and collective chronopathy—the rate at which a population's stories for a subject seem to change over time. We show that 2017 was the most turbulent overall year for Trump. In 2020, story generation slowed dramatically during the first two major waves of the COVID-19 pandemic, with rapid turnover returning first with the Black Lives Matter protests following George Floyd's murder and then later by events leading up to and following the 2020 US presidential election, including the storming of the US Capitol six days into 2021. Trump story turnover for 2 months during the COVID-19 pandemic was on par with that of 3 days in September 2017. Our methods may be applied to any well-discussed phenomenon, and have potential to enable the computational aspects of journalism, history, and biography.

## 1 Introduction

What happened in the world last week? What about a year ago? As individuals, it can be difficult for us to freely recall and order in time—let alone make sense of—events that have occurred at scopes running from personal and day-to-day to global and historic [1–10]. One

including the "Focalevents" package from Ryan Gallagher which can be found on GitHub: <https://github.com/ryanjgallagher/focalevents> Given that our results are statistical estimates of word frequencies found in a 10% random subset of "trump" messages, drawn from the Decahose stream provided by Twitter, qualitatively similar results will be found using any sufficiently large collection of "trump" messages. Messages authored by President Trump can be found at the Trump Twitter Archive can be used as a source for the former President's messages at <https://www.thetrumparchive.com>.

**Funding:** This project was supported by a gift to the Vermont Complex Systems Center from MassMutual, Google Open Source, and NSF awards #2117345. The funders had no role in study design, data collection and analysis, decision to publish, or preparation of the manuscript.

**Competing interests:** The authors are grateful for the computing resources provided by the Vermont Advanced Computing Core which was supported in part by NSF award No. OAC-1827314, financial support from the Massachusetts Mutual Life Insurance Company and Google Open Source under the Open-Source Complex Ecosystems And Networks (OCEAN) project. This does not alter our adherence to PLOS ONE policies on sharing data and materials.

emblematic challenge for remembering story timelines is presented by the 45th US president Donald J. Trump, our interest here. Stories revolving around Trump have been abundant and diverse in nature. Consider, for example, being able to remember and then order stories involving: North Korea, Charlottesville, kneeling in the National Football League, Confederate statues, family separation, Stormy Daniels, Space Force, and the possible purchase of Greenland.

Added to these problems of memory is that people's perception of the passing of time is subjective and complicated [11–18]. Days can seem like months ("this week dragged on forever") or might seem to be over in a flash ("time flies"). Story-wise, periods of time can also range from being narratively simple ("it was the only story in town") to complicated and hard to retell ("everything happened all at once"). At the population scale, major news stories may similarly arrive at slow and fast paces, and may be coherent or disconnected. As one example, within the space of around 15 minutes after 9 pm US Eastern Standard Time on March 11, 2020, Tom Hanks and Rita Wilson announced that they had tested positive for COVID-19, the National Basketball Association put its season on hold indefinitely due to the COVID-19 pandemic, and Trump gave an Oval Office Address during which the Dow Jones Industrial Average futures dropped. And to help illustrate the potential disconnection of co-occurring stories within the realm of US politics, at the same time as the above events were unfolding, former US vice presidential candidate Sarah Palin was appearing on the popular Fox TV show "The Masked Singer" performing Sir Mix-A-Lot's "Baby Got Back" in a bear costume.

Here, in order to quantify story turbulence around Trump—and the collective experience of story turbulence around Trump—we develop a data-driven, computational approach to constructing a timeline of stories surrounding any given subject, with high resolution in both time and narrative (see [Data and Methods](#), Sec. 2).

For data, we use Twitter as a vast, noisy, and distributed news and opinion aggregation service [19–23]. Beyond the centrality of Twitter to Trump's communications [24–29], a key benefit of using Twitter as "text as data" [30–34] is that popularity of story is encoded and recorded through social amplification by retweets [35]. We show that Twitter is an effective source for our treatment though our methods may be applied broadly to any temporally ordered, text-rich data sources.

We define, create, and explore week-scale timelines of the most 'narratively dominant' 1-grams and 2-grams in tweets containing the word Trump (Sec. 3.1). We supply day-scale timelines as part of the paper's Online Appendices ([compstorylab.org/trumpstoryturbulence/](http://compstorylab.org/trumpstoryturbulence/)).

Having a computationally determined timeline of  $n$ -grams then allows us to operationalize and measure a range of features of story dynamics.

First, and in a way particular to Trump, we quantify narrative control: The extent to which  $n$ -grams being used in Trump-matching tweets are due to retweets of Trump's own tweets (Sec. 3.2). We show Trump's narrative control varies from effectively zero (e.g., 'coronavirus') to high ('Crooked Hillary').

Second, we compute, plot, and investigate the normalized usage frequency time series for  $n$ -grams that are narratively dominant in Trump-matching tweets for three or more days (Sec. 3.3). Along with their temporal ordering, these day-scale time series provide a rich representation of the shapes of Trump-related stories including shocks, decays, and resurgence. We also incorporate our narrative control measure into these time series visualizations.

Third, we measure story turbulence at the scale of months by comparing Zipf distributions for 1-gram usage frequencies between individual days across logarithmically increasing time scales (Sec. 3.4). We are able to quantify and show, for example, that story turbulence for Trump was highest in the second half of 2017 and lowest during the lockdown period of the COVID-19 pandemic in the US.

Finally, we are then able to realize a numeric measurement of ‘collective chronopathy’ which we define to be how time seems to be passing (‘time-feel’) by the rate of story turnover (Sec. 3.5). We quantify how the past can seem more distant or closer to the present as we move through time, and that it may do so nonlinearly as a function of how far back we look in time. The course of the COVID-19 pandemic in the US rendered, for example, July 2020 as being closer to April 2020 than June.

## 2 Data and methods

We draw on a collection of around 10% of all tweets starting in 2008. We take all English language tweets [35, 36] matching the word ‘Trump’ from 2015/01/01 on. We ignore case and accept matches of ‘Trump’ at any location of a tweet (e.g., ‘@realDonaldTrump’ matches). We break these Trump-matching tweets into 1-grams and 2-grams, and create Zipf distributions at the day scale per Coordinated Universal Time (UTC). In previous work, we have assessed the popularity of major US political figures on Twitter, finding that the median usage rank of the word ‘trump’ across all of Twitter is less than 200, tantamount to that of basic English function words (e.g., ‘say’) [28]. Consequently, our resulting data set is considerable containing around 20 billion 1-grams.

Our main collection of Trump-matching tweets thereby includes tweets about Trump and tweets by Trump. Retweets and quote retweets are naturally accounted for as they are individually recorded in our database. We further filter  $n$ -grams for simple latin character words including hashtags and handles.

To quantify the degree to which Trump might be in control of a story, we make a second database using the subset of  $n$ -grams found in retweets of Trump’s tweets (we exclude any quote tweet matter). We then generate day-scale Zipf distributions again, in the same format as for all Trump-matching tweets.

We perform two main analyses of these time series of Zipf distributions, treating 1-gram and 2-gram distributions separately. First, we determine which  $n$ -grams are most ‘narratively dominant’ by comparing each day’s Zipf distribution with the Zipf distribution of the same day one year prior, using our allotaxonomic instrument of rank-turbulence divergence (RTD) [37]. A full derivation and exploration of the benefits of RTD relative to other information theoretic measures is beyond the scope of the present work. However, we briefly describe the formulation and impetus for RTD here to offer context.

Component size distributions for complex systems are typically conducted using comparisons such as Jensen-Shannon divergence. Most of these methods lack transparency and adjustability, and our RTD measure offers a principled approach to make visible component contributions through a tunable instrument for comparing any two ranked lists, here for words and phrases on different dates. Many additional details can be found in Dodds et. al. 2020 [37], where we explore the performance of RTD for a series of distinct settings including species abundance, baby name popularity, market capitalization, and performance in sports among others.

As mentioned previously, we begin by rank ordering a system  $\Omega$ ’s types from largest to smallest size according to some measure in the manner of Zipf [48]. For the present study, we identify the top 10,000 1-grams and 2-grams found in messages containing ‘trump’. We indicate the rank of type  $\tau$  as  $r_\tau$ , and the ordered set of all types and their ranks as  $R_\Omega$ . For all  $n$ -grams with the same count, we assign the mean of the sequence of ranks these types would occupy otherwise.

Given two systems,  $\Omega_1$  and  $\Omega_2$ , both comprised of  $n$ -grams (e.g., 2 word phrases found in tweets containing ‘trump’ on two separate dates), we express rank-turbulence divergence

between these systems as  $D_{RT}^{(\alpha)}(\Omega_1 \parallel \Omega_2)$ , where  $\alpha$  is a single tunable parameter with  $0 \leq \alpha < \infty$ . Given a pair ranked lists,  $R_1$  and  $R_2$ , we will more directly write  $D_{RT}^{(\alpha)}(R_1 \parallel R_2)$  for their comparison, where

$$D_{RT}^{(\alpha)}(R_1 \parallel R_2) = \sum_{\tau \in R_{1,2;\alpha}} D_{RT,\tau}^{(\alpha)}(R_1 \parallel R_2). \quad (1)$$

Finally, we sort  $n$ -grams by descending contribution,  $D_{RT,\tau}^{(\alpha)}(R_1 \parallel R_2)$ , indicating this ordering by the set  $R_{1,2;\alpha}$ .

Having established the methodology, first we compare the top 10,000  $n$ -grams and set the RTD parameter  $\alpha$  at 1/4, a reasonable fit for Twitter data (sensitivity experiments [37]). Tuning away from 1/4 gives similar overall results as does the use of different kinds of probability-based divergences such as Jensen-Shannon divergence. We use a rank-based divergence because of the plain-spoken interpretability and general statistical robustness that ranks confer. As an example, in S1 Fig, we provide and explain an RTD-based allotaxonograph to determine the narratively dominant 1-grams in Trump-related tweets for the date of the Capitol insurrection, 2021/01/06.

We find a year to be both a stable time gap with six months to two years also producing similar results. We then create day- and week-scale timelines of keyword  $n$ -grams, filtering out hashtags and user handles.

Second, we use RTD to quantify the turbulence between Zipf distributions for any pair of dates. To systematically measure change in story over time, we compare each date's Zipf distribution to that of an approximately logarithmically increasing sequence of days  $\delta$  before.

We use the one instrument of rank-turbulence divergence throughout our paper for two reasons: (1) Analytic coherence (each section builds out from the previous ones), and (2) Facilitation of quantification. While we would expect sophisticated topic modeling approaches would also help elucidate stories, our goals are more expansive regarding the experience of time.

## 3 Analysis and discussion

### 3.1 Computational timeline generation for dominant Trump stories

In Figs 1 and 2, we show computational timelines of the most narratively dominant 1-grams and 2-grams in Trump-matching tweets for each week running from the start of 2016 into 2020. We coarse-grain from days to weeks by finding  $n$ -grams with the highest overall rank-turbulence divergence (RTD) sum for each week.

Our computational timelines provide  $n$ -grams as keyword hooks, and immediately give a rich overall view of the major stories surrounding Trump. Broadly, the early major chapters run through the Republican nomination process, the election, and inauguration. Reflected in names of individuals and entities, places, and processes, the timelines then move through a range of US and world events (North Korea, Charlottesville, Parkland, Iran, George Floyd's murder, Portland); US policy and systems (travel ban, Space Force, southern border wall, Supreme Court); natural disasters (hurricanes in 2017, COVID-19 in 2020); scandals (Russia, Stormy Daniels, Mueller, impeachment, Taliban bounties for American soldiers), and the 2020 US election and aftermath (death of Ruth Bader Ginsburg, debate with Biden, Trump's contraction of COVID-19, claims of fraud, a focus on Georgia, the storming of the US Capitol).

The week-scale timelines for 1-grams and 2-grams variously directly agree (e.g., 'Epstein' and 'Jeffrey Epstein' in week 27 of 2019, and 'coronavirus' and 'the coronavirus' for 9

Week	2016	2017	2018	2019	2020	2021
1. 01/01–01/07	Hillary 34.7	hacking 28.6	Bannon 2.2	shutdown 0.0	Iran 9.6	Georgia 14.7
2. 01/08–01/14	Cruz 1.0	Meryl 5.0	Mueller 0.0	shutdown 0.0	Soleimani 5.9	Capitol 0.1
3. 01/15–01/21	Cruz 10.7	inauguration 0.6	DACA 6.7	Pelosi 6.8	Parnas 0.0	Capitol 0.0
4. 01/22–01/28	Cruz 10.6	inauguration 3.1	Mueller 0.0	Pelosi 2.6	Ukraine 5.5	insurrection 0.0
5. 01/29–02/04	Cruz 11.2	ban 2.1	Mueller 0.0	border 0.0	impeachment 0.0	Greene 0.0
6. 02/05–02/11	Cruz 5.1	Bannon 0.0	memo 2.3	Whitaker 0.0	Vindman 2.5	insurrection 0.0
7. 02/12–02/18	Cruz 6.9	Flynn 0.0	Mueller 0.0	emergency 0.0	Barr 2.2	Capitol 0.0
8. 02/19–02/25	Rubio 3.8	Sweden 4.9	Parkland 0.3	Jussie 0.0	Bloomberg 6.3	Capitol 0.0
9. 02/26–03/04	Rubio 9.2	Russia 6.4	Mueller 0.0	Cohen 3.7	coronavirus 0.0	Capitol 0.0
10. 03/05–03/11	Cruz 1.0	Russian 4.8	Mueller 0.0	Nadler 13.7	coronavirus 0.0	insurrection 0.0
11. 03/12–03/18	Cruz 5.7	tax 1.8	Mueller 2.2	emergency 1.6	coronavirus 0.0	Biden 0.0
12. 03/19–03/25	Arizona 16.8	Nunes 0.0	Mueller 2.2	Barr 0.0	coronavirus 0.0	Biden 0.0
13. 03/26–04/01	women 8.3	Russia 9.9	Stormy 0.0	Schiff 5.2	coronavirus 0.5	Capitol 0.0
14. 04/02–04/08	Cruz 1.5	Russia 2.8	Mueller 0.0	returns 0.0	coronavirus 0.0	Matt 0.0
15. 04/09–04/15	Cruz 1.7	Syria 0.4	Mueller 2.0	Barr 2.4	coronavirus 0.0	Capitol 0.0
16. 04/16–04/22	Cruz 10.5	Russia 0.5	Mueller 0.1	Barr 0.1	coronavirus 0.0	Capitol 0.0
17. 04/23–04/29	Cruz 3.0	days 0.1	Kanye 8.0	Biden 6.0	coronavirus 0.0	audit 0.0
18. 04/30–05/06	Indiana 11.5	Trumpcare 0.0	Mueller 0.0	Barr 0.0	coronavirus 0.0	Cheney 0.0
19. 05/07–05/13	Ryan 2.5	Comey 2.8	Iran 6.6	Barr 0.0	coronavirus 0.0	Cheney 0.0
20. 05/14–05/20	Bernie 25.3	Comey 1.0	ZTE 4.5	Barr 0.0	coronavirus 0.0	Cheney 0.0
21. 05/21–05/27	Clinton 9.5	budget 0.0	Korea 18.2	Barr 0.0	pandemic 0.0	Weisselberg 0.0
22. 05/28–06/03	Hillary 11.9	Kathy 4.4	Roseanne 4.0	USS 3.0	Minneapolis 32.1	reinstated 0.0
23. 06/04–06/10	Clinton 11.1	Comey 0.8	pardon 0.0	Mexico 27.6	police 4.2	McGahn 0.0
24. 06/11–06/17	Orlando 12.4	Mueller 0.0	Kim 4.1	foreign 2.0	Tulsa 4.5	DOJ 0.0
25. 06/18–06/24	Hillary 23.9	Trumpcare 0.0	children 1.0	Iran 12.9	Tulsa 2.1	
26. 06/25–07/01	Clinton 13.0	Russia 5.8	Justice 8.3	Moon 29.9	bounties 0.0	
27. 07/02–07/08	Crooked 80.6	CNN 0.7	toddlers 0.0	parade 0.0	Rushmore 2.3	
28. 07/09–07/15	Crooked 71.5	Russian 1.2	NATO 13.0	Epstein 0.0	coronavirus 0.0	
29. 07/16–07/22	Pence 2.9	Mueller 0.0	Helsinki 3.1	racist 0.8	coronavirus 0.0	
30. 07/23–07/29	DNC 6.1	Scouts 0.0	Cohen 0.0	Baltimore 13.6	Portland 11.8	
31. 07/30–08/05	Khan 6.5	Mueller 0.0	LeBron 0.7	Baltimore 9.4	pandemic 0.0	
32. 08/06–08/12	Crooked 55.2	Korea 5.8	Omarosa 0.4	Paso 7.6	USPS 0.0	
33. 08/13–08/19	Manafort 0.0	Charlottesville 1.5	Omarosa 9.5	Greenland 6.9	USPS 0.0	
34. 08/20–08/26	Clinton 7.6	Charlottesville 3.8	Cohen 2.7	Greenland 8.0	Biden 6.6	
35. 08/27–09/02	Crooked 57.4	Harvey 0.0	Ohr 14.0	Dorian 12.2	Kenosha 9.5	
36. 09/03–09/09	Bondi 0.0	DACA 2.4	Kavanaugh 2.1	Dorian 12.6	Atlantic 4.8	
37. 09/10–09/16	deplorable 0.0	ESPN 2.7	Puerto 7.5	flavored 0.0	Woodward 2.6	
38. 09/17–09/23	Clinton 6.5	Kim 4.9	Kavanaugh 1.7	Ukraine 4.5	coronavirus 0.0	
39. 09/24–09/30	debate 4.9	Puerto 4.7	Kavanaugh 9.5	Ukraine 6.8	ballots 0.7	
40. 10/01–10/07	Pence 4.9	Puerto 2.1	Kavanaugh 6.8	Ukraine 5.1	Covid 1.4	
41. 10/08–10/14	sexual 0.3	Puerto 1.8	Kavanaugh 4.3	Kurds 8.2	COVID 1.4	
42. 10/15–10/21	rigged 10.1	Puerto 0.2	Saudi 5.3	Kurds 3.7	Biden 8.2	
43. 10/22–10/28	star 0.0	Mueller 0.0	caravan 0.0	impeachment 0.0	Biden 9.2	
44. 10/29–11/04	FBI 5.9	Mueller 0.0	caravan 0.0	impeachment 0.0	Biden 10.0	
45. 11/05–11/11	Clinton 0.9	Gillespie 12.0	Whitaker 6.2	Ukraine 6.2	votes 3.4	
46. 11/12–11/18	Bannon 0.0	sexual 1.7	caravan 0.0	Ukraine 5.2	Dominion 23.2	
47. 11/19–11/25	Hamilton 12.4	LaVar 21.3	Saudi 1.6	Ukraine 3.5	Sidney 0.1	
48. 11/26–12/02	recount 0.0	Moore 0.0	Moscow 0.1	impeachment 3.1	votes 24.1	
49. 12/03–12/09	Taiwan 7.8	Mueller 0.0	Cohen 2.1	impeachment 0.0	Georgia 20.2	
50. 12/10–12/16	Russia 2.9	Mueller 0.0	Cohen 6.9	impeachment 0.0	vaccine 11.1	
51. 12/17–12/23	inauguration 11.8	Mueller 0.0	wall 9.8	impeachment 1.4	vaccine 15.4	
52. 12/24–12/31	inauguration 3.2	Mueller 0.0	wall 20.4	impeachment 7.6	Election 60.2	

**Fig 1. Computational timeline of ‘narratively dominant’ 1-grams surrounding Trump-matching tweets at the week scale.** The timelines of 1-grams and 2-grams (Fig 2) give an overall sense of story turbulence through turnover and recurrence. For each week, we show the 1-gram with the highest sum of rank-turbulence divergence (RTD) contributions [37] relative to the year before, based on comparisons of day-scale Zipf distributions for 1-grams in Trump-matching tweets. See S1 Fig for a full example of 2021/01/06, the day of the Capitol insurrection. Each 1-gram must have been the most narratively dominant for at least 1 day of the week (i.e., highest RTD contribution). The numbers and pale pink bars represent Trump’s narrative control as measured by the percentage of an *n*-gram appearing in retweets of Trump’s own tweets during a given week. The limits of 0 and 100 for

narrative control thus correspond to a 1-gram being never tweeted by Trump and a 1-gram only appearing in retweets of Trump. After Trump’s account was suspended by Twitter following the Capitol insurrection, Trump’s narrative control necessarily falls to 0. To align weeks across years, we assign the final 8 days to Week 52, and for each leap year we include February 29 as an 8th day in week 9. See the paper’s Online Appendices ([compstorylab.org/trumpstoryturbulence/](http://compstorylab.org/trumpstoryturbulence/)) for the analogous visualization at the day scale as well as a visualization of the daily top 10 most narratively dominant 1-grams. In constructing this table and the table for 2-grams in Fig 2, we excluded hashtags, a small set of function word *n*-grams, and expected-to-be surprising *n*-grams such as ‘of the’ and, in 2017, ‘President Trump’.

<https://doi.org/10.1371/journal.pone.0260592.g001>

Week	2016	2017	2018	2019	2020	2021
1. 01/01–01/07	Hillary Clinton 32.7	plant in 85.1	Steve Bannon 5.7	the government 0.0	a war 6.6	in Georgia 20.2
2. 01/08–01/14	Trump rally 0.0	Meryl Streep 6.6	shithole countries 0.0	the border 1.0	impeachment trial 0.0	the Capitol 0.0
3. 01/15–01/21	Ted Cruz 26.0	Trump’s inauguration 0.0	the government 1.4	Cohen to 0.0	impeachment trial 0.0	the Capitol 0.0
4. 01/22–01/28	Megyn Kelly 4.9	executive order 0.0	the FBI 5.6	the government 0.0	impeachment trial 0.0	the Capitol 0.0
5. 01/29–02/04	Ted Cruz 19.7	travel ban 1.6	the FBI 9.4	Ralph Northam 26.0	impeachment trial 0.0	the Capitol 0.0
6. 02/05–02/11	New Hampshire 19.5	travel ban 1.1	military parade 0.0	El Paso 4.7	Alexander Vindman 0.0	the Capitol 0.0
7. 02/12–02/18	Ted Cruz 15.7	Michael Flynn 0.0	school shooting 3.1	national emergency 0.0	Roger Stone 4.0	the Capitol 0.0
8. 02/19–02/25	Ted Cruz 30.1	Trump administration 0.0	the NRA 0.0	Jussie Smollett 0.0	Bernie Sanders 13.6	the Capitol 0.0
9. 02/26–03/04	vote for 4.4	to Russia 22.0	Hope Hicks 0.0	Michael Cohen 5.3	the coronavirus 0.0	the Capitol 0.0
10. 03/05–03/11	Ted Cruz 2.4	travel ban 0.0	Stormy Daniels 0.0	Tim Apple 0.0	the coronavirus 0.0	voted for 0.0
11. 03/12–03/18	Trump is 0.1	Meals on 0.0	Stormy Daniels 0.0	New Zealand 17.9	the coronavirus 0.0	Lara Trump 0.0
12. 03/19–03/25	Lyn’ Ted 66.2	health care 0.0	Cambridge Analytica 0.0	Mueller report 0.0	the coronavirus 0.0	the border 0.0
13. 03/26–04/01	Trump is 0.0	Freedom Caucus 20.8	Stormy Daniels 0.0	Mueller report 0.0	the coronavirus 0.0	Matt Gaetz 0.0
14. 04/02–04/08	Ted Cruz 3.9	Susan Rice 0.3	National Guard 0.0	tax returns 0.0	the coronavirus 0.0	Matt Gaetz 0.0
15. 04/09–04/15	New York 19.3	in Syria 0.2	Michael Cohen 0.0	sanctuary cities 5.3	the coronavirus 0.0	Matt Gaetz 0.0
16. 04/16–04/22	Ted Cruz 28.1	turnout for 0.0	Michael Cohen 2.4	Mueller report 0.0	the coronavirus 0.0	Maxine Waters 0.0
17. 04/23–04/29	Trump rally 0.0	tax plan 0.0	the Korean 0.0	Mueller report 0.0	the coronavirus 0.0	Liz Cheney 0.0
18. 04/30–05/06	Ted Cruz 5.5	health care 0.0	Stormy Daniels 0.0	Mueller report 0.0	treated worse 0.0	Liz Cheney 0.0
19. 05/07–05/13	Paul Ryan 2.0	James Comey 6.7	the Iran 9.0	tax returns 0.0	tested positive 0.0	Liz Cheney 0.0
20. 05/14–05/20	Hillary Clinton 26.5	Saudi Arabia 12.5	are animals 0.0	Lindsey Graham 0.0	the pandemic 0.0	Kevin McCarthy 0.0
21. 05/21–05/27	Hillary Clinton 24.8	Saudi Arabia 8.2	the FBI 23.3	Nancy Pelosi 12.5	a mask 6.3	the January 0.0
22. 05/28–06/03	Trump University 3.4	Kathy Griffin 5.7	Samantha Bee 4.4	John McCain 0.0	photo op 0.0	Memorial Day 0.0
23. 06/04–06/10	Hillary Clinton 18.6	James Comey 0.2	Justin Trudeau 8.5	with Mexico 39.2	Left Democrats 75.1	Jean Carroll 0.0
24. 06/11–06/17	Trump is 0.0	obstruction of 12.6	their parents 0.0	the FBI 8.5	in Tulsa 7.4	Trump DOJ 0.0
25. 06/18–06/24	Hillary Clinton 20.6	Karen Handel 16.6	their parents 3.4	need soap 0.0	in Tulsa 2.2	
26. 06/25–07/01	Hillary Clinton 20.5	Fake News 37.6	Supreme Court 3.7	Jean Carroll 0.0	American soldiers 0.0	
27. 07/02–07/08	Crooked Hillary 82.8	North Korea 28.6	Trump administration 0.0	Jeffrey Epstein 0.0	Mount Rushmore 3.9	
28. 07/09–07/15	Crooked Hillary 73.3	Trump Jr 0.0	Supreme Court 7.9	Jeffrey Epstein 0.0	Roger Stone 0.0	
29. 07/16–07/22	Mike Pence 6.8	Secret Service 0.0	in Helsinki 1.7	a racist 0.0	in Portland 0.0	
30. 07/23–07/29	Crooked Hillary 79.6	Boy Scouts 0.0	Walk of 0.0	Elijah Cummings 27.2	in Portland 8.9	
31. 07/30–08/05	Khizr Khan 0.0	Maxine Waters 0.0	enemy of 22.2	El Paso 11.1	the election 3.4	
32. 08/06–08/12	Hillary Clinton 10.5	North Korea 5.7	Space Force 11.1	El Paso 7.7	Social Security 0.0	
33. 08/13–08/19	Trump campaign 0.0	white supremacists 0.0	security clearance 0.0	New Hampshire 26.5	the USPS 0.0	
34. 08/20–08/26	Hillary Clinton 19.1	Joe Arpaio 3.5	Michael Cohen 4.3	Prime Minister 28.7	Joe Biden 5.9	
35. 08/27–09/02	Crooked Hillary 61.8	Hurricane Harvey 0.1	John McCain 0.2	Hurricane Dorian 9.6	Joe Biden 2.7	
36. 09/03–09/09	in Detroit 0.0	to end 0.0	Brett Kavanaugh 7.6	the Taliban 3.0	Joe Biden 3.4	
37. 09/10–09/16	tax returns 0.0	white supremacist 0.0	Puerto Rico 8.4	Dan Bishop 37.7	Joe Biden 13.3	
38. 09/17–09/23	Trump Jr 0.0	North Korea 12.8	Blasey Ford 0.0	a foreign 6.4	Supreme Court 7.3	
39. 09/24–09/30	Hillary Clinton 7.5	Puerto Rico 5.2	Brett Kavanaugh 15.7	Impeachment inquiry 0.0	Supreme Court 5.7	
40. 10/01–10/07	Mike Pence 8.9	Puerto Rico 2.6	Supreme Court 6.9	Adam Schiff 13.3	Walter Reed 5.7	
41. 10/08–10/14	sexual assault 0.0	Puerto Rico 2.2	Kanye West 0.0	the Kurds 11.3	Biden is 26.5	
42. 10/15–10/21	Hillary Clinton 19.9	families of 0.0	Saudi Arabia 6.6	the Kurds 3.8	Joe Biden 12.1	
43. 10/22–10/28	Hillary Clinton 11.7	Myeshia Johnson 0.0	the bombs 0.0	World Series 0.0	Joe Biden 10.1	
44. 10/29–11/04	Hillary Clinton 6.5	Twitter employee 0.0	birthright citizenship 0.0	the impeachment 0.0	Joe Biden 12.6	
45. 11/05–11/11	Trump wins 0.0	mental health 0.0	Jim Acosta 0.0	pro quo 8.1	the election 2.2	
46. 11/12–11/18	Steve Bannon 0.0	ban on 0.0	president who 0.0	impeachment inquiry 0.0	the election 7.5	
47. 11/19–11/25	Mike Pence 24.3	Roy Moore 0.0	Saudi Arabia 2.5	quid pro 1.3	the election 6.7	
48. 11/26–12/02	popular vote 17.4	Native American 0.1	Trump Tower 2.5	Hong Kong 0.0	voter fraud 32.2	
49. 12/03–12/09	Air Force 18.2	Roy Moore 3.5	campaign finance 0.0	to impeach 7.7	in Georgia 12.9	
50. 12/10–12/16	of State 7.6	of sexual 0.0	Michael Cohen 7.8	articles of 0.0	the election 9.0	
51. 12/17–12/23	Electoral College 5.8	tax bill 0.0	the wall 13.7	Christianity Today 8.1	election fraud 13.9	
52. 12/24–12/31	Trump next 0.0	the FBI 0.1	Border Security 70.6	the Senate 29.1	on January 16.7	

**Fig 2. Computational timeline of narratively dominant 2-grams surrounding Trump at the week scale.** The timeline’s construction and the plot’s details are analogous to that of Fig 2. The *n*-grams in both timelines match (‘coronavirus’ and ‘the coronavirus’), expand on each other (‘bounties’ and ‘American troops’), or point to different stories (‘Mueller’ and ‘Stormy Daniels’). See the paper’s Online Appendices ([compstorylab.org/trumpstoryturbulence/](http://compstorylab.org/trumpstoryturbulence/)) for a day-scale visualization of the top 10 most narratively dominant 2-grams in Trump-matching tweets.

<https://doi.org/10.1371/journal.pone.0260592.g002>

consecutive weeks in 2020); make connections (e.g., ‘Crooked’, ‘Hillary’, and ‘Crooked Hillary’ in 2016); or point to different stories (e.g., ‘Syria’ and ‘Trump Foundation’ in week 51 of 2018).

The  $n$ -gram timelines also provide an overall qualitative sense of story turbulence. In the lead up to the 2016 election, the stories around Trump largely concern his opponents, particularly Ted Cruz and Hillary Clinton. Near the election, we see an increase in story turbulence with ‘sexual assault’, ‘rigged’, and ‘FBI’. During Trump’s presidency, the timelines give a sense of stories becoming more enduring over time. After a tumultuous first nine months of 2017, we see the first stretch of four or more weeks being ruled by the same dominant story: ‘Puerto’/‘Puerto Rico’ and ‘Mueller’. In 2018, ‘Mueller’ then dominates for 12 of 17 weeks and then later ‘Kavanaugh’ leads for 5 of 6 weeks. In 2019, ‘Barr’ is number one for 4 of 5 weeks early on. Later, ‘Ukraine’ and ‘impeachment’ reflect the main story of the last four months of 2019, broken up only by ‘Kurds’. And in 2020, COVID-19 is the main story for 13 consecutive weeks for 1-grams starting at the end of February, pushed down by the murder of George Floyd and subsequent protests, before returning again as the major story of July. The remainder of the timeline largely reflects the 2020 election and its aftermath of claims of fraud by Trump, leading to the Capitol insurrection in 2021.

Our intention with Figs 1 and 2 is to give one page summaries of over five years of stories around Trump. Below these dominant  $n$ -grams at the week scale are  $n$ -grams at the day scale representing many other stories. For example, Table 1 shows the the top 12 most narratively dominant 1-grams for three consecutive days in late June of 2020 where the story of Russia paying Taliban soldiers bounties for killing US soldiers rose abruptly to prominence. We see that by June 27, references to the COVID-19 pandemic had dropped down the list (‘coronavirus’ would return to the top 3 on July 1), as had indicators of other stories (‘Biden’, ‘lobster’, and ‘statues’ point to Trump’s 2020 challenger, a call by Trump to subsidize the US lobster industry, and the taking down of Confederate statues in the aftermath of George Floyd’s murder). The Russian bounties story would fall away within a week, with the next non-pandemic story appearing being the arrest of Ghislaine Maxwell on July 2, 2020. Maxwell was charged for sex trafficking of underage girls in connection with convicted sex offender Jeffrey Epstein.

As part of the Online Appendices, we provide a range of visualizations and data files for day-scale computational timelines for the most narratively dominant 1-grams and 2-grams.

**Table 1. Top 12 most narratively dominant 1-grams for Trump-matching tweets for an example three day period.** The sharp rise of the Russian bounties for Taliban soldiers to kill US troops, along with the fall of other stories, gives an example of a microscopic detail of story turbulence we aim to quantify macroscopically.

Rank	2020/06/25	2020/06/26	2020/06/27
1	coronavirus	pandemic	bounties
2	Biden	bounties	bounty
3	pandemic	coronavirus	soldiers
4	testing	Biden	militants
5	lobster	militants	Russia
6	Matter	hiring	kill
7	statues	cases	briefed
8	virus	testing	Afghanistan
9	cases	virus	intelligence
10	ad	Statues	Taliban
11	fishing	bounty	troops
12	ChinaVirus	Care	pandemic

<https://doi.org/10.1371/journal.pone.0260592.t001>

### 3.2 Narrative control

Because we are working with Twitter, some  $n$ -grams may become narratively dominant due to high overall use by many distinct individual users because of one highly retweeted individual's tweet, and all degrees in between. Trump is the evident potential source of many retweets in our data set, and we introduce the concept of 'narrative control' as part of our analysis. In Figs 1 and 2, the underlying pale pink bars and accompanying numbers indicate our measure of Trump's narrative control,  $C_{\tau,t}$ . We quantify  $C_{\tau,t}$  for any given  $n$ -gram  $\tau$  on Twitter for some time period  $t$  as the percentage of the occurrences of  $\tau$  due to retweets of Trump tweets:

$$C_{\tau,t} = 100f_{\tau,t}^{\text{RT}}/f_{\tau,t}.$$

We see that Trump's week-scale levels of narrative control vary greatly over time. A few example highs, ordered by their date of first becoming narratively dominant, are 'Crooked Hillary' (82.6%), 'Fake News' (37.6%) 'Border Security' (70.6%), 'Minneapolis' (32.1%), and 'Left Democrats' (75.1%).

By contrast Trump's narrative control has been low to non-existent for narratively dominant  $n$ -grams representing many stories. For 'Mueller' in 2017 and 2018,  $C_{\tau,t}$  ranges from 0% to 2.2% (see however Sec. 3.3 below). The names 'Stormy Daniels', 'Jeffrey Epstein', and 'Ghislaine Maxwell' all register  $C_{\tau,t} = 0$ , as does 'the bombs' in reference to pipe bombs mailed to leading Democrats and journalists in October, 2018.

Strikingly, Trump's narrative control for 'coronavirus' has been effectively 0. Of course, and as for all  $n$ -grams, Trump may have used other terms to refer to them. In the case of 'coronavirus', he has used 'corona virus' (with optional capitalizations), 'invisible enemy', and a derogatory term. But our measurement here gets at the degree to which Trump has engaged with a specific  $n$ -gram being used on Twitter in connection with him.

Transitions in narrative control also stand out. In 2020, the most striking shifts at the week scale are 0% for 'pandemic' to 32.1% for Minneapolis, 0% for 'photo op' to 75.1% for 'Left Democrats', and 0% for 'coronavirus' to 11.8% for 'Portland'. And the start of 2020 with the US assassination of Iranian general Soleimani jumps out amid the  $n$ -grams of the impeachment hearings ('Iran' 9.6%, 'Soleimani' 5.9%, 'a war' 6.6%).

The storming of the Capitol on 2021/01/06 by Trump supporters lead to Twitter permanently banning Trump's Twitter account. In the context of Twitter and by the nature of our measure, Trump's narrative control therefore ended abruptly on 2021/01/08.

### 3.3 Time series for dominant Trump stories

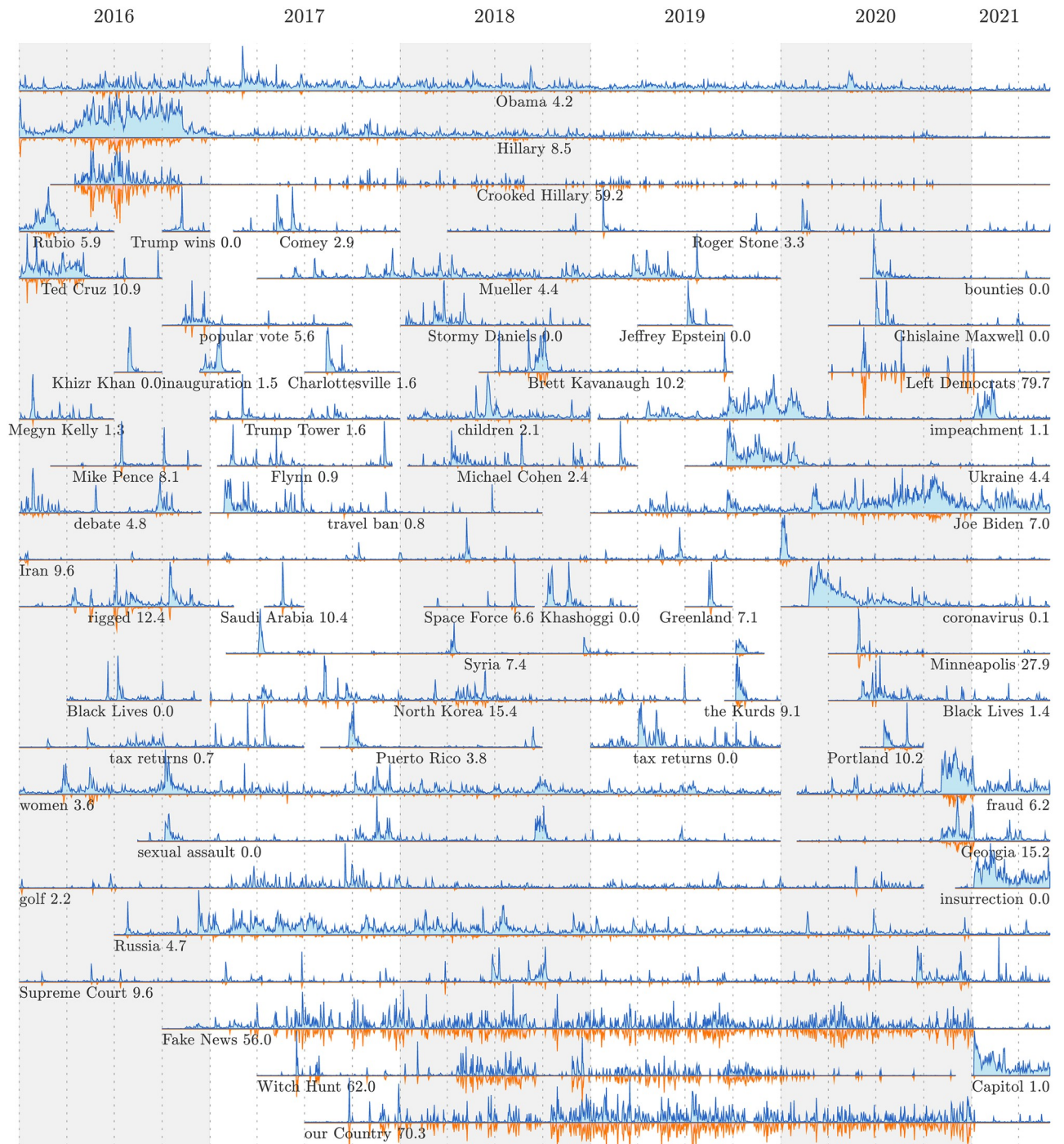
We now take our list of narratively dominant 1-grams and 2-grams and extract day-scale time series of normalized usage rate in Trump-matching tweets. We also generate the corresponding narrative control time series.

In Fig 3, we show a selection of  $n$ -gram time series (blue lines) with Trump's narrative control time series inverted (orange lines with pale pink fill). Below each horizontal axis, we annotate the  $n$ -gram along with the overall narrative control percentage for section of time series visible. We normalize each time series by its maximum.

We see a range of motifs common to sociotechnical time series: spikes ('Saudi Arabia', 'Syria', 'Cambridge Analytica'), shocks with decays ('Puerto Rico', 'Kurds'), sharp drop offs ('Ted Cruz'), episodic bursts ('Roger Stone', 'Syria', 'sexual assault'), and noise ('Russia', 'Obama') [38–40].

While an  $n$ -gram may be ranked as the most narratively dominant for some period of time, its usage rate may fluctuate during that time. A clear example is the time series for 'coronavirus' which begins with an initial spike followed by a shock, and then trends linearly down until the end of June, 2020.





**Fig 3. Example time series of narratively dominant 1-grams and 2-grams for Trump-matching tweets.** Each time series shows the relative usage rate for  $n$ -grams (blue curve), and Trump's level of narrative control (inverted orange curve with pale pink fill). See the Online Appendices ([compstorylab.org/trumpstoryturbulence/](https://compstorylab.org/trumpstoryturbulence/)) for full time series for all  $n$ -grams which have been narratively dominant on at least three days. The background shading and vertical dashed lines indicate years and quarters.

<https://doi.org/10.1371/journal.pone.0260592.g003>

As it is derived from a subset's fraction of a whole, the narrative control time series can at most exactly mirror the overall time series. Four examples of time series with high and enduring narrative control percentages are 'Crooked Hillary', 'Witch Hunt', 'Fake News', and 'our Country' (59.2%, 62.0%, 56.2%, and 70.7%).

More subtly, we see that Trump's narrative control for 'Mueller' increases over time, though remaining modest at 4.4%.

Many of the examples we show exhibit relatively flat narrative control time series. For 2-grams, Trump did not tweet about 'Black Lives' (Matter) in 2016, and only a fraction in 2020 (2.5%). And we see zero narrative control for 'sexual assault' whose time series correlates strongly with 'women'.

Post the 2020 election, we show three time series connected with Trump's claims of a 'rigged election': 'fraud' (6.6%), 'Georgia' (16.3%), and 'Capitol' (2.0%).

In the Online Appendices and as an ancillary file on the arXiv, we include a PDF booklet of time series for all  $n$ -grams which have been narratively dominant in Trump tweets on at least three days. We extend the time series for all  $n$ -grams to run from 2016/01/01 on, and order them by the date they first achieve narrative dominance. Moving through this ordered sequence of 308 time series gives another qualitative experience of how the stories around Trump have unfolded in time, placing them in temporal context.

### 3.4 Story turbulence

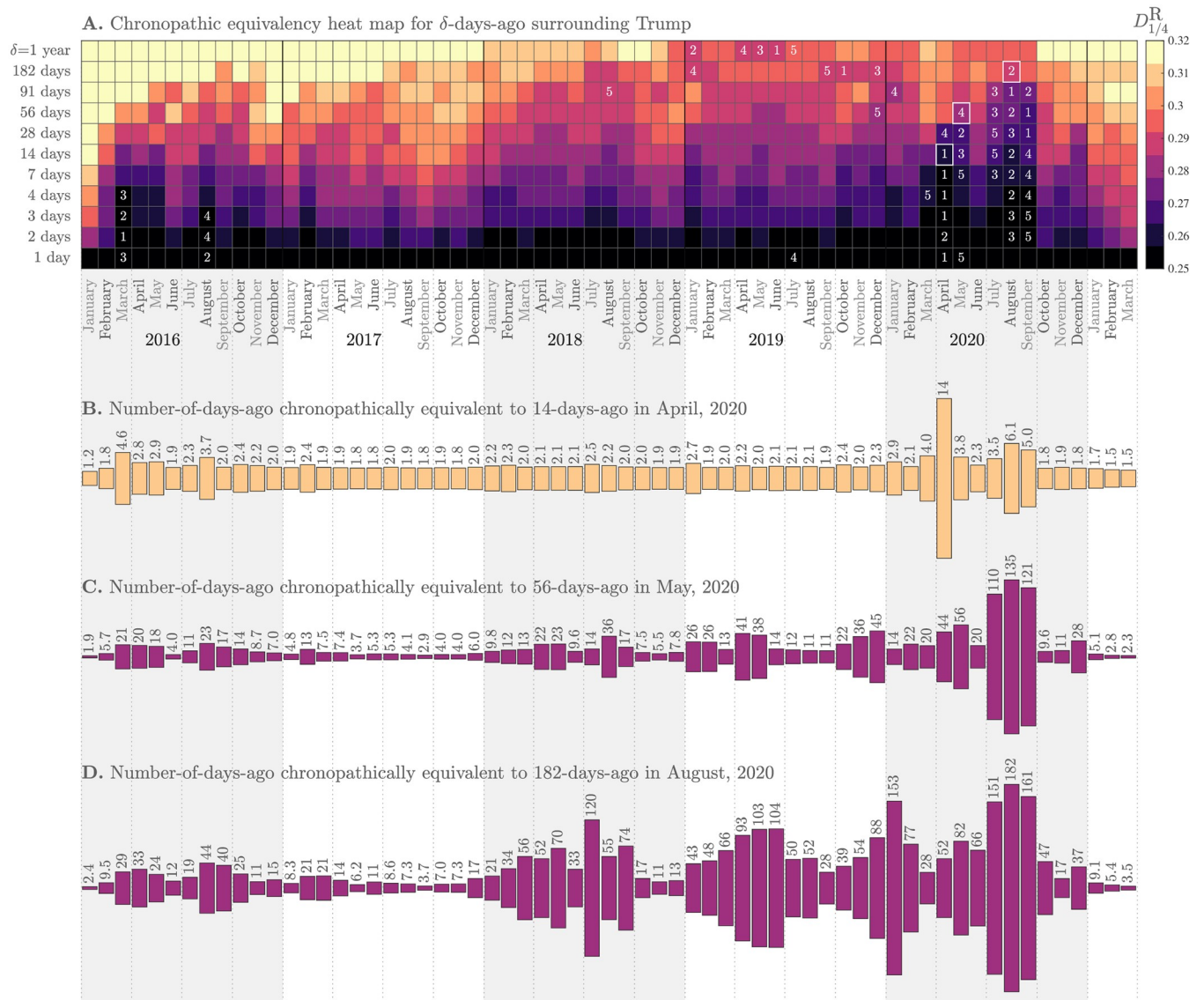
We have so far demonstrated that the  $n$ -grams of Trump-matching tweets track major event[s] and narratives around Trump. We now move to using our comparisons of Zipf distributions for 1-grams via rank-turbulence divergence (RTD) to operationalize and quantify two aspects of story experience: (1) Story turbulence, the rate of story turnover surrounding Trump, and (2) Collective chronopathy, the feeling of how time passes at a population scale. Chronopathy is to be distinguished from the differently defined chronesthesia [11]. Here, we use the -pathy suffix primarily to mean 'feel' (as in 'empathy') though a secondary connotation of sickness proves serviceable too (as in 'sociopathy').

In Fig 4, we present visualizations of the collective chronopathy for Trump-matching tweets. We focus first on the heat map of Fig 4A, the core distillation of our measurement of the passing of time.

For each day in each month, we compare 1-gram Zipf distributions to the 1-gram Zipf distribution for  $\delta$  days before using RTD with  $\delta$  varying approximately logarithmically from 1 day to 1 year (vertical axis of Fig 4A). We generate the heat map using the median value of RTD for each month and each  $\delta$ . We employ a 10 point, dark-to-light perceptually uniform color map for increasing RTD which corresponds to the speeding up of story turnover. For each  $\delta$ , we indicate the five slowest months by annotation with 1 being the slowest.

Overall, we see a general trend of story turnover slowing down—the plot thickens—with several marked exceptions, leading to an extraordinary slowing down across many time scales throughout the COVID-19 pandemic in 2020.

We describe the main features of the heat map, moving from left to right through time. As we would expect, the election year of 2016 shows fast turnover at the longer time scales of  $\delta \sim$  six months to a year. This turnover carries through into 2017 as the shift to being president necessarily leads to different word usage around Trump. For the shorter time scales, January 2016 has fast turnover at all time scales—Trump's narrative was rapidly changing as he was becoming the lead contender in the Republican primaries. Some of the slowest turnovers for 1 day out to 14 days occur in 2016, notably in March and August.



**Fig 4. A. Chronopathic equivalency heat map:** Each cell represents the ‘story distance’ for a given month and a given number of  $\delta$  days before. We measure story distance as the median rank-turbulence divergence (RTD) between the Zipf distributions of 1-grams used in Trump-matching tweets for each day of a month and  $\delta$  days before (we use RTD parameter  $\alpha = 1/4$ ). Lighter colors on the perceptually uniform color map correspond to higher levels of story turnover. Numbers indicate the slowest five months for each value of  $\delta$ . After the story turbulence of the 2016 election year and especially the first year of Trump’s presidency, there has been a general slowing down in story turnover at all time scales (the ‘plot thickens’). By September US of 2020, the COVID-19 pandemic had induced record slowing down of story turnover around Trump at time scales up to 91 days, the story being punctuated by and then combined with the Black Lives Matter protests in response to George Floyd’s murder on May 25, 2020. **B.** Using an example anchor of April 2020 and  $\delta = 14$  days (white square in panel A), a plot of chronopathic equivalent values of  $\delta$  across time. During Trump’s presidency, the same story turnover occurred as fast as every 1.8 days in September 2017 and 1.7 days in October 2020. Because story turbulence is nonlinear, using a different anchor month and  $\delta$  (i.e., selecting a different cell in the heatmap) potentially gives a different chronopathic equivalency plot. **C.** Anchor of 56 days in May 2020. **D.** Anchor of 182 days in August 2020. See the Online Appendices ([compstoriylab.org/trumpstoryturbulence/](https://compstoriylab.org/trumpstoryturbulence/)) for the most recent version of this figure. As for Fig 3, shading and lines give guides for years and quarters.

<https://doi.org/10.1371/journal.pone.0260592.g004>

Story turnover generally increases through 2017 at all time scales  $\delta$ , with September, 2017 being the month with days least connected to all that had come before in the previous year. As suggested by Figs 1 and 2, the first year of Trump’s presidency burned through story, with just some of the dominant narratives concerning Flynn, Comey, Russia, Mueller, North Korea, Charlottesville, DACA, NFL, and three major hurricanes.

In 2018 and 2019, we see some of the slowest turnover at longer time scales. Trump has the consistency of being president in the previous year. Two periods where story turnover for shorter time scales slow down in the second and third years of the Trump presidency are around the government shutdown (January 2019) and the impeachment hearings (last three months of 2019).

In March 2020, the COVID-19 pandemic brings story turnover to a functional halt at shorter time scales. April is especially slow with either the slowest or second slowest turnovers for time scales ranging from 1 day to 14 days. The same time scales for May are all still slow (top 5) but now longer time scales show less story turnover. The murder of George Floyd on May 25 and subsequent Black Lives Matter protests then leads to a sharp transition in the heat map. At the same time, the pandemic had subsided before what would be new surge in August, and was for much of June a secondary story.

Entering July 2020, story turnover slows dramatically for a second time—now slowing at all time scales—as the pandemic once again becomes a major narrative along with that of the Black Lives Matter protests. August has the slowest 91 day story turnover, and September the same for 28 and 56 day time scales. The text around Trump in August 2020 is closer to that of May 2020 than for any other three month comparison.

Story turnover rises sharply again in October and November of 2020, in the lead up to and aftermath of the 2020 US presidential election. The narratives around Trump begin to roil at the end of September which saw the contentious first presidential debate and the death of Justice Ruth Bader Ginsberg. The subsequent nomination and confirmation of Amy Coney Barrett for the Supreme Court, and the direct connection to Trump's contraction of COVID-19, hospitalization, and recovery then led to a thrashing of the narrative timeline. Post election, the lack of a clear immediate winner and then Trump's refusal to concede and claims of fraud made November, December, and January especially turbulent at longer time scales.

### 3.5 Collective chronopathy

We turn finally to determining what we will call chronopathically equivalent time scales. For any given month and time scale  $\delta$ —any cell in the heat map of Fig 4A—we can estimate time scales in other months with corresponding values of RTD.

For three examples, the white-bordered squares in 2020 in Fig 4A mark anchor time scales and months of  $\delta = 14$  days (2 weeks) in April,  $\delta = 56$  days ( $\sim 2$  months) in May, and  $\delta = 182$  days ( $\sim 6$  months) in August. For these three anchors, Fig 4B–4D, show the equivalent values of  $\delta$  across all months.

Fig 4B shows that two weeks in April felt like the longest two weeks across the whole time frame. All other months achieve the same level of story turnover in less than a week, with a maximum of 6.1 days in August 2020. With the first wave of the pandemic unfolding, the stories around Trump became stuck and time dragged, collectively.

Lifting up to the 56-day anchor in May 2020, in Fig 4C, we see the slowdown of July, August, and September 2020 dominate, with time scales doubling. We see some near equivalent time scales in 2018 and 2019, with the impeachment's progression by the end of 2019 producing a 45 day period. The equivalent time scales of around 9.5 for October and November of 2020 point to speed-up factors of  $\sim \times 12$ .

Stepping further out to the level of story turnover in six months relative to August 2020, Fig 4D We see episodic slowdowns in 2018 and 2019, with two main causes in 2019 being the Mueller report and the impeachment hearings. With this anchor, we see November 2020 separate from October with a chronopathically equivalent time scale of 13 days versus 47.

Taken together, Fig 4B–4D make clear, and quantify, the rapidity of story turnover in 2017, especially September of that year.

Story turbulence is complicated as the speed up or slow down of story can vary nonlinearly across different memory time scales. Annual days of the year may seem close to that of a year before (e.g., Thanksgiving) but further away from time periods in between. For Trump stories, our measure of story turnover across 56 days looking back from May 2020 was equivalent to around 3 days in September 2017, a speed up of  $\times 19$ , and 110 days in July 2020, a slowing down by a factor of  $\times 2$ . In choosing 28 days instead of 56 in May 2020, we would find equivalent time scales of 2.1 and 16 in September 2017 and July 2020, with speed-up factors of  $\times 13$  and  $\times 1.7$ . So by this shorter time scale benchmark of about a month in May, story seems to have moved faster in July because of the narrative dominance of protests in June.

## 4 Concluding remarks

We have shown that interpretable, high-level timeline summaries of historical events can be derived from Twitter. Our process, which we have worked to make as rigorous as possible, is necessarily computational: In our approximately 10% sample of all of Twitter, there were from 2015/01/01 through to 2020/08/12 around 20 billion distinct 1-grams in Trump-matching tweets, with 1.5 billion of those being contained in retweets of Trump's tweets.

Because we have shown that the narratively dominant  $n$ -grams our method find are historically sensible, we are then able to defensibly quantify narrative control, story turnover, and collective chronopathy in the context of Trump.

We observe that our focus on  $n$ -grams has resulted in timelines that are largely descriptive of major events and stories through the surfacing of the names of people, places, institutions, processes, and social phenomena. While the collective construction of what matters cannot be said to be objective—what matters socially is what matters socially [41, 42]—we find that our computationally derived timelines are not laden with overt opinion or framing with the notable exception of certain phrases due to Trump himself.

Our study of stories around Trump suggests much future possible research.

Broader explorations of Trump's (and others') narrative control should be possible. We have limited ourselves here to narratively dominant  $n$ -grams and only within the same time frame. We have elsewhere examined "ratioing" of tweets by Obama and Trump [43], exploring in particular the balance of likes to retweets, and these and other measures of interactions could provide a more nuanced way to operationalize narrative control. Studies across different news outlets and social media platforms could explore the extent to which Trump's narratives are driven by external stories and vice versa, as well as their persistence in time, and certain modeling and prediction may be possible.

We have indexed stories by keywords of 1-grams and 2-grams. Connecting these  $n$ -grams to Wikipedia would possibly allow for the automatic generation of timelines augmented by links to (or summaries of) Wikipedia entries. Disconnects between Twitter and Wikipedia (and other texts) would be of interest to explore as well.

Computationally generating a taxonomy of story type would be a natural way to improve upon our work. How have the kinds of stories around Trump changed over time, which ones have persisted, and which ones have been more likely to last only a few days?

We have presented narratively dominant  $n$ -grams for the time scales of weeks, while reserving the top 20 at the day scale for the paper's Online Appendices. While this is reasonable for a study going across now five years, for certain major events, shorter time scales of hours or even minutes may be better suited as the temporal units of analysis.

The relationship between Trump's favorability polls with story turbulence and story kind could also be examined. As a rough observation, we note that the speeding up and then slowing down of story turnover around Trump through 2017 and into 2018 appears to be anti-

correlated with Trump's approval ratings during that same period. Having established a story taxonomy would be important for moving in this direction.

While our specific instrumental focus has involved the use of Twitter to analyze stories surrounding Trump, in doing so we have laid out a general, structured approach to quantifying and exploring story turbulence for any well-defined topic in temporally ordered, large-scale text (e.g., news outlets, online forums, and Reddit). Wherever Zipf distributions can be derived, our methods will allow for the computational construction of  $n$ -gram-anchored story timelines along with measures of story distance, turbulence, and chronopathy. For such extensions, we caution that the popularity of text must be incorporated in some fashion through measures of readership, shares, etc. We also observe that stronger variations in findings would potentially come from choosing different text sources rather than through reasonable adjustments of the methods of analysis we have laid out here.

We believe that our techniques will be of value for computationally augmented journalism and the computational social sciences [44–47]. News reporters and historians in particular are faced with having to process ever more data, text, and media, and we see our work here as contributing to a far broader effort to develop ways to provide powerful computational support for the vital work performed by domain-knowledge experts.

## Supporting information

**S1 Fig. Example allotaxonograph [37] comparing 1-grams in Trump-matching tweets on two dates, separated by a year: 2020/01/06 and 2021/01/06.** Our findings all build out from allotaxonomy: The principled measurement of the difference between the architectures of any two complex systems as represented by Zipf distributions [48]. For 2020/01/06 versus 2021/01/06, 1-grams on the first date reflect the aftermath of the assassination of Iranian military commander Qasem Soleimani by the United States on 2020/01/03, while 1-grams appearing on the second date revolve around the Capitol insurrection by Trump supporters. The rank-rank histogram on the left of the allotaxonograph displays the joint Zipf distribution, naturally laid out in double-logarithmic space. We generate the 1-gram ranking on the right of the allotaxonograph using rank-turbulence divergence (RTD) [37] with tuning parameter  $\alpha = 1/4$ . The contour lines on the histogram provide guides for RTD showing a reasonable fit to the joint Zipf distribution's form. Variations around  $\alpha = 1/4$  will not change the overall findings (i.e., the orderings of contributing 1-grams as well as the overall RTD score). See Ref. [37] for a full explanation of allotaxonomy and RTD. In Sec. 3.1 in the main paper, we use RTD at the year scale to determine narratively dominant 1-grams and 2-grams arising on the second of the two dates being compared. By comparing to a year ago, we are able to generate a background Zipf distribution that will help remove calendrical features as well as generically Twitter- and Trump-related 1-grams (e.g., 'RT' and 'Donald'). While the allotaxonograph shows 1-grams on both dates, we emphasize that our focus is on the second date, in that we are seeking to determine the most important 1-grams of today. For 2020/01/06 versus 2021/01/06, looking at the ranked list on the right, we see that the top five 1-grams for the day of the Capitol insurrection are 'Capitol', 'supporters', 'police', 'building', and 'breached'. The salience of 'BLM' (Black Lives Matter) and 'Antifa' point to the immediate confusion and disinformation surrounding the Capitol insurrection. We provide all year-scale allotaxonographs at the paper's Online Appendices. In Secs. 3.4 and 3.5, we then use RTD at time scales of a day up to a year to quantify story turbulence and collective chronopathy. (PDF)

## Author Contributions

**Conceptualization:** Peter Sheridan Dodds.

**Data curation:** Peter Sheridan Dodds, Joshua R. Minot, Michael V. Arnold, Thayer Alshaabi, Jane Lydia Adams, Andrew J. Reagan, Christopher M. Danforth.

**Formal analysis:** Peter Sheridan Dodds.

**Funding acquisition:** Peter Sheridan Dodds.

**Investigation:** Peter Sheridan Dodds.

**Methodology:** Peter Sheridan Dodds.

**Project administration:** Peter Sheridan Dodds, Christopher M. Danforth.

**Resources:** Peter Sheridan Dodds.

**Software:** Peter Sheridan Dodds, Joshua R. Minot, Michael V. Arnold, Thayer Alshaabi, Jane Lydia Adams, Andrew J. Reagan.

**Supervision:** Peter Sheridan Dodds.

**Validation:** Peter Sheridan Dodds, Michael V. Arnold, Thayer Alshaabi, Andrew J. Reagan.

**Visualization:** Peter Sheridan Dodds, Joshua R. Minot, Jane Lydia Adams, Christopher M. Danforth.

**Writing – original draft:** Peter Sheridan Dodds.

**Writing – review & editing:** Peter Sheridan Dodds, Joshua R. Minot, Michael V. Arnold, Thayer Alshaabi, Jane Lydia Adams, Andrew J. Reagan, Christopher M. Danforth.

## References

1. Gunter B. *Poor reception: Misunderstanding and forgetting broadcast news*. Routledge, 1987.
2. Price V. and Zaller J. Who gets the news? Alternative measures of news reception and their implications for research. *Public opinion quarterly*, 57(2):133–164, 1993. <https://doi.org/10.1086/269363>
3. Neisser U., Fivush R., et al. *The remembering self: Construction and accuracy in the self-narrative*. Cambridge University Press, 1994.
4. Lang A. The limited capacity model of mediated message processing. *Journal of Communication*, 50(1):46–70, 2000. <https://doi.org/10.1111/j.1460-2466.2000.tb02833.x>
5. Tulving E. Episodic memory: From mind to brain. *Annual Review of Psychology*, 53(1):1–25, 2002. <https://doi.org/10.1146/annurev.psych.53.100901.135114> PMID: 11752477
6. Schacter D. L. *Searching for memory: The brain, the mind, and the past*. Basic Books, 2008.
7. Fivush R. The development of autobiographical memory. *Annual Review of Psychology*, 62:559–582, 2011. <https://doi.org/10.1146/annurev.psych.121208.131702> PMID: 20636128
8. Ocasio W., Mauskopf M., and Steele C. W. History, society, and institutions: The role of collective memory in the emergence and evolution of societal logics. *Academy of Management Review*, 41:676–699, 2016. <https://doi.org/10.5465/amr.2014.0183>
9. Garibaldi E. and Garibaldi P. Ordering history through the timeline, 2017. CEPR Discussion Paper No. DP12508.
10. Loftus E. F. Eyewitness science and the legal system. *Annual Review of Law and Social Science*, 14:1–10, 2018. <https://doi.org/10.1146/annurev-lawsocsci-101317-030850>
11. Tulving E. Chronesthesia: Conscious awareness of subjective time. In Stuss D. T. and Knight R. T., editors, *Principles of frontal lobe function*, pages 311–325. Oxford University Press, 2002.
12. Edy J. *Troubled pasts: News and the collective memory of social unrest*. Temple University Press, 2006.
13. Droit-Volet S. and Meck W. H. How emotions colour our perception of time. *Trends in cognitive sciences*, 11:504–513, 2007. <https://doi.org/10.1016/j.tics.2007.09.008> PMID: 18023604

14. Sackett A. M., Meyvis T., Nelson L. D., Converse B. A., and Sackett A. L. You're having fun when time flies: The hedonic consequences of subjective time progression. *Psychological Science*, 21:111–117, 2010. <https://doi.org/10.1177/0956797609354832> PMID: 20424031
15. Phillips I. Perceiving temporal properties. *European Journal of Philosophy*, 18:176–202, 2010. <https://doi.org/10.1111/j.1468-0378.2008.00299.x>
16. Grondin S. Timing and time perception: A review of recent behavioral and neuroscience findings and theoretical directions. *Attention, Perception, & Psychophysics*, 72(3):561–582, 2010. <https://doi.org/10.3758/APP.72.3.561> PMID: 20348562
17. Allman M. J. and Meck W. H. Pathophysiological distortions in time perception and timed performance. *Brain*, 135:656–677, 2012. <https://doi.org/10.1093/brain/awr210> PMID: 21921020
18. Rudd M., Vohs K. D., and Aaker J. Awe expands people's perception of time, alters decision making, and enhances well-being. *Psychological science*, 23:1130–1136, 2012. <https://doi.org/10.1177/0956797612438731> PMID: 22886132
19. Dodds P. S., Harris K. D., Kloumann I. M., Bliss C. A., and Danforth C. M. Temporal patterns of happiness and information in a global social network: Hedonometrics and Twitter. *PLoS ONE*, 6:e26752, 2011. <https://doi.org/10.1371/journal.pone.0026752> PMID: 22163266
20. Gallagher R. J., Stowell E., Parker A. G., and Foucault Welles B. Reclaiming stigmatized narratives: The networked disclosure landscape of #MeToo. *Proceedings of the ACM on Human-Computer Interaction*, 3(CSCW):1–30, 2019. <https://doi.org/10.1145/3359198>
21. Jackson S. J., Bailey M., and Foucault Welles B. *#HashtagActivism: Networks of Race and Gender Justice*. MIT Press, 2020.
22. Tangherini T. R., Shahsavari S., Shahbazi B., Ebrahimzadeh E., and Roychowdhury V. An automated pipeline for the discovery of conspiracy and conspiracy theory narrative frameworks: Bridgegate, PizzaGate and storytelling on the web. *PLoS ONE*, 15(6):e0233879, 2020. <https://doi.org/10.1371/journal.pone.0233879> PMID: 32544200
23. Allen J., Howland B., Mobius M., Rothschild D., and Watts D. J. Evaluating the fake news problem at the scale of the information ecosystem. *Science Advances*, 6:eay3539, 2020. <https://doi.org/10.1126/sciadv.aay3539> PMID: 32284969
24. Lee J. C. and Quealy K. The 282 people, places and things Donald Trump has insulted on Twitter: A complete list. *The New York Times*, 2016.
25. Ott B. L. The age of Twitter: Donald J. Trump and the politics of debasement. *Critical Studies in Media Communication*, 34(1):59–68, 2017. <https://doi.org/10.1080/15295036.2016.1266686>
26. Bovet A., Morone F., and Makse H. A. Validation of Twitter opinion trends with national polling aggregates: Hillary Clinton vs Donald Trump. *Scientific reports*, 8:1–16, 2018. <https://doi.org/10.1038/s41598-018-26951-y>
27. Ott B. L. and Dickinson G. *The Twitter presidency: Donald J. Trump and the politics of White rage*. Routledge, 2019.
28. Dodds P. S., Minot J. R., Arnold M. V., Alshaabi T., Adams J. L., Dewhurst D. R., et al. Fame and Ultra-fame: Measuring and comparing daily levels of 'being talked about' for United States' presidents, their rivals, God, countries, and K-pop, 2019. Available online at <https://arxiv.org/abs/1910.00149>.
29. Ouyang Y. and Waterman R. W. Trump tweets: A text sentiment analysis. In *Trump, Twitter, and the American Democracy*, pages 89–129. Springer, 2020.
30. Monroe B. L., Colaresi M. P., and Quinn K. M. Fightin' words: Lexical feature selection and evaluation for identifying the content of political conflict. *Political Analysis*, 16:372–403, 2008. <https://doi.org/10.1093/pan/mpn018>
31. Grimmer J. and Stewart B. M. Txt as data: The promise and pitfalls of automatic content analysis methods for political texts. *Political analysis*, 21(3):267–297, 2013. <https://doi.org/10.1093/pan/mps028>
32. Gentzkow M., Kelly B., and Taddy M. Text as data. *Journal of Economic Literature*, 57(3):535–74, 2019. <https://doi.org/10.1257/jel.20181020>
33. Riffe D., Lacy S., Fico F., and Watson B. *Analyzing media messages: Using quantitative content analysis in research*. Routledge, 2019.
34. Boyd R. L., Blackburn K. G., and Pennebaker J. W. The narrative arc: Revealing core narrative structures through text analysis. *Science Advances*, 6:eaba2196, 2020. <https://doi.org/10.1126/sciadv.aba2196> PMID: 32821822
35. Alshaabi T., Dewhurst D. R., Minot J. R., Arnold M. V., Adams J. L., Danforth C. M., et al. The growing amplification of social media: Measuring temporal and social contagion dynamics for over 150 languages on Twitter for 2009–2020. *EPJ Data Science*, 10:15, 2020. <https://doi.org/10.1140/epjds/s13688-021-00271-0>



36. Alshaabi T., Adams J. L., Arnold M. V., Minot J. R., Dewhurst D. R., Reagan A. J., et al. Storywrangler: A massive exploratorium for sociolinguistic, cultural, socioeconomic, and political timelines using Twitter. *Science Advances*, 2021. <https://doi.org/10.1126/sciadv.abe6534> PMID: 34272243
37. Dodds P. S., Minot J. R., Arnold M. V., Alshaabi T., Adams J. L., Dewhurst D. R., et al. Allotaxonomy and rank-turbulence divergence: A universal instrument for comparing complex systems, 2020. Available online at <https://arxiv.org/abs/2002.09770>.
38. Crane R. and Sornette D. Robust dynamic classes revealed by measuring the response function of a social system. *Prod. Nat. Acad. Sci.*, 105:15649–15653, 2008. <https://doi.org/10.1073/pnas.0803685105> PMID: 18824681
39. Dewhurst D. R., Alshaabi T., Kiley D., Arnold M. V., Minot J. R., Danforth C. M., and et al. The shocklet transform: A decomposition method for the identification of local, mechanism-driven dynamics in socio-technical time series. *EPJ Data Science*, 9:3, 2020. <https://doi.org/10.1140/epjds/s13688-020-0220-x>
40. De Domenico M. and Altmann E. G. Unraveling the origin of social bursts in collective attention. *Scientific reports*, 10:1–9, 2020. <https://doi.org/10.1038/s41598-020-61523-z> PMID: 32170082
41. Berger P. L. and Luckmann T. *The social construction of reality: A treatise in the sociology of knowledge*. Penguin UK, 1991.
42. Sandu A. *Social construction of reality as communicative action*. Cambridge Scholars Publishing, 2016.
43. Minot J. R., Arnold M. V., Alshaabi T., Danforth C. M., and Dodds P. S. Ratioing the President: An exploration of public engagement with Obama and Trump on Twitter. *PLOS ONE*, 16, 2021. Available online at <https://arxiv.org/abs/2006.03526>.
44. Lazer D., Pentland A. S., Adamic L., Aral S., Barabasi A. L., Brewer D., et al. Life in the network: the coming age of computational social science. *Science (New York, NY)*, 323:721, 2009. <https://doi.org/10.1126/science.1167742>
45. Conte R., Gilbert N., Bonelli G., Cioffi-Revilla C., Deffuant G., Kertesz J., et al. Manifesto of computational social science. *The European Physical Journal Special Topics*, 214:325–346, 2012. <https://doi.org/10.1140/epjst/e2012-01697-8>
46. Flew T., Spurgeon C., Daniel A., and Swift A. The promise of computational journalism. *Journalism Practice*, 6:157–171, 2012. <https://doi.org/10.1080/17512786.2011.616655>
47. Coddington M. Clarifying journalism's quantitative turn: A typology for evaluating data journalism, computational journalism, and computer-assisted reporting. *Digital journalism*, 3:331–348, 2015. <https://doi.org/10.1080/21670811.2014.976400>
48. Zipf G. K. *Human Behaviour and the Principle of Least-Effort*. Addison-Wesley, Cambridge, MA, 1949.