

Political Catchphrases

Approaches via data analysis and online media

Emyn Dean - Survey May 2023



The Open
University

This report was commissioned for the workshop

Political Catchphrases in Our Time

Organised by the Language, Literature and Politics Research Group
of the Open University (UK)

2nd – 3rd June 2023

About the author

Emyn “Em” Dean has a MSc in Computer Science from Queen Mary, University of London and is working towards a PhD in Computer Science at the Open University, funded by the UKRI.

One of their first jobs was as a politics and immigration reporter for *Hola America News*, a dual-language (Spanish and English) newspaper delivering to Latin American communities in the Midwest US. During this time, they covered the 2008 US Presidential Election in Iowa, which sparked their interest in digital media as a form of political campaigning.

Em currently acts as a policy author and advisor to a major UK political party on LGBT+ issues and matters pertaining to public debate, media, and journalism, in which capacity they have informed national law and guidance for the governments of Wales and Scotland, and authored briefings for political leaders and public elected officials. They are also (to a lesser extent) active in the politics of the United States, Luxembourg, and the European Union.

Em’s academic interests include public debate in online media, interactive design of social media, theoretical robotics, social data science, digital anthropology, and digital ethics. They have presented their work at the Anthropology, AI and the Future of Human Society Conference (2022) and the ACM Web Conference (2023).

They can be contacted at em.dean@open.ac.uk.

Contents

1. Introduction	4
2. Limitations	4
3. Research questions	5
4. Methodology	5
5. Context: definitions and terms	5
5.1 Key definitions	5
5.2 Technical terms and slang	6
5.3 Is it “catchword” or “catchphrase”? Or something else?	6
5.3.1 Slogans	7
5.4 “Political” catchphrases – by politicians, about politics, or of political origin?	7
5.5 Terms in other major languages	9
6. Context: Social media platforms and communities to consider	9
6.1 Where to start?	9
6.2 Who sees political catchphrases on social media?	11
7. Context: Automation and disinformation	16
7.1 Recommendation algorithms	16
7.2 Bot farms	17
7.3 Disinformation campaigns	18
7.4 Mitigation techniques for disinformation, algorithmic bias, and bots	19
8. Political catchphrases on social media and analysed via computational methods, in academic research	19
8.1 Books	19
8.2 Some Relevant Papers: General	20
8.3 Some Relevant Papers: Regional use; Specific elections	21
9. Technical resources and approaches	22
9.1 Data science basics – tools and principles	22
Key tools	22
Key principles	22
9.2 Key datasets for Reddit, Facebook, and Instagram	24
Google datasetsearch	24
Meta tools (Facebook and Instagram, 2018-present)	24
CrowdTangle (Facebook, Instagram, and Reddit)	24
Kaggle (Most platforms, selected periods 2019-present)	24
Twitter and Facebook hashtags	24
TAPS Dataset (Twitter, 2016-2020)	25
Reddit API (2008-present)	25
Database of Stopwords	25
Other tools	25
9.3 Gathering data in private networks such as WhatsApp and Discord	25
9.4 Some relevant computational data science methods	26
Data exploration	26
Data visualization	26
Data cleanup	26
Dimensionality reduction and feature reduction	26
Sentiment analysis	26
Cluster analysis	27
Knowledge graphs	27
9.6 Technical demonstration of basic catchphrase data exploration and analysis	27
9.6.1 Getting the data	27
9.6.2 Understanding the data’s context	28
9.6.2 Initial data exploration in Excel	28
9.6.3 Finding catchphrases in the “message” text	29
9.6.4 Finding catchphrases with an initial visual analysis	30
10. Concluding remarks	31
11. Works Cited	31

1. Introduction

Ten years ago, in July 2013, George Zimmerman, a volunteer community watch officer, was acquitted of the charge of murdering seventeen-year-old Trayvon Martin, a young black man. Martin had been shot and killed by Zimmerman after an altercation in Zimmerman's neighbourhood, where Martin had gone with his father to visit his father's fiancée.

Trayvon Martin had been an active social media user, and had a wide network of followers on Twitter. Following his death, there were intensive calls for justice, which quickly spread beyond his immediate circle. After Zimmerman's acquittal, these calls became still louder, as there was a widespread feeling that Martin's race contributed to mishandling of the case by the police and judicial system. A phrase rapidly spread on social media, expressing this feeling – “#BlackLivesMatter”.

We don't know who first used the term “Black Lives Matter”, but it was probably not a politician, and it was not, at that time, a catchphrase. However, the strength of feeling from racial justice advocates, combined with the architecture of Twitter's hashtag system, whereby a simple phrase or word could rapidly travel across the globe, has meant that “Black Lives Matter” quickly became a well-known political catchphrase. It is the name of a political action committee in the US, and appears on protest signs across the anglosphere and beyond.

The “Black Lives Matter” phenomenon speaks to the power of social media communication to popularize and globalize political catchphrases, with remarkable alacrity. Moreover, in the last two decades, social media has moved from being largely the preserve of young people in the anglosphere, to being a global phenomenon, used by millions. In some countries, social media now challenges major newspapers and television as one of the largest sources of news, and is a leading expenditure for political advertisers. This suggests that political catchphrase use on social media will only increase in the years to come, and will be a focus of study for linguists, social scientists, and computer scientists.

This presents an opportunity and a challenge for social data science. While much of social media is readily accessible to researchers, some is not. And while there are established data science methodologies for handling political data, techniques applicable to political catchphrases, such as natural language processing and knowledge graphs, are still in a nascent form.

This report attempts to take stock of the progress to date, illustrate the context of this inquiry, and present some ideas and opportunities to researchers who may have interest in this topic.

2. Limitations

This survey does not attempt a comprehensive literature review of political catchphrase research. This is a substantial and accelerating field of study. I have listed a number of papers and resources of interest, particularly with a focus on the digital sphere post-2016. But even this is likely not exhaustive.

I have attempted to make an account of international perspectives in political catchphrases and the study thereof, however, I do not confidently read languages other than English, Spanish, and (sometimes) French. Due to the localization of many political phenomena, it may be expected that some key research exists in virtually every modern language, as well as significant political catchphrases themselves in (perhaps most notably) Arabic, Persian languages, and Chinese.

Finally, and perhaps most importantly, it is said that “a week is a long time in politics”. A week is, similarly, a long time in social media. It is likely that many of the observations in this survey will become out of date in a few years, or in some cases, months.

3. Research questions

RQ1. What are the key datasets, publications, and projects relevant to the spread or diffusion of contemporary political catchphrases?

RQ2. What methods can we use to elicit and analyze data from online resources to further develop this project?

4. Methodology

In answering *RQ1*, I seek to provide:

- An overview of key definitions and terms,
- contextual information on social media and the research space,
- a sampling of literature on the subject of political catchphrases, and
- a sampling of useful datasets.

In answering *RQ2*, I seek to provide:

- an overview of some key technical methods common in data science and social data science,
- a demonstration of how researchers may retrieve data from social media, and
- a demonstration of how researchers might handle a sample dataset.

In longer textual sections, there are “key takeaways” bullet points for people who wish to examine this documentation more briefly.

5. Context: definitions and terms

5.1 Key definitions

catchphrase

We use the term “catchphrase” as per *Webster’s New American Dictionary*, to refer to “a word or expression that is used repeatedly and conveniently to represent or characterize a person, group, idea, or point of view” [1]. For a further discussion of this definition, see Section 5.3 below.

catchword

The word “catchword” is not used extensively in this report as it is generally considered to be fully included by the word “catchphrase”. For a further discussion of this definition, see Section 5.3 below.

political catchphrase

The phrase “political catchphrase” is used herein to refer to the above, where the person, group, idea, or point of view is commonly described as “political”, e.g., a politician, a political party, an ideological movement, etc. At times, it is used in papers and databases cited herein in a narrower sense, to refer only to those catchphrases used by or originating with individuals, e.g., politicians or political leaders.

social media

“Social media”, in its most basic form, is any digital media which is designed to, or effects, relationships between users by leveraging user-generated content, thus developing a social network in the online space.

5.2 Technical terms and slang

algorithm refers to programmatic processes used in social media (and the internet in general) to retrieve content based on user inputs or other user data. For example, a search engine uses an algorithm generated from the text you enter a search box to return content you might be looking for. This is often used as a byword for a “**recommendation algorithm**”. A recommendation algorithm uses data about a user to show that user customized content, thus keeping them more engaged with a specific website or app than they would be if they only saw “generic” content.

group refers to a community of individuals organized in a location within a larger social media environment, such as a Facebook group, Discord server, or sub-Reddit.

GDPR and **E-Privacy** are regulations and laws in the European Union governing the use of data by and of EU citizens. These currently also apply in the UK.

VOIP refers to “voice over internet protocol”, a system for communication via voice to remote users via the internet. For example, Skype is a VOIP program. Some social media sites, most notably Discord, have VOIP features.

hashtag activism or **hashtivism** refers to the use of **hashtags** (words or phrases preceded with a “#” on some social media platforms, which generate searchable links to other content using that hashtag) in social activism. Successful hashtags may be considered as a type of catchphrase.

platform is a marketing term for a website or other digital tool which delivers content in a multimodal and flexible way, such as via mobile phone apps, multiple websites, etc. Large social media websites are often referred to as “platforms”.

NLP, or **natural language processing**, refers to computational attempts – typically involving machine learning – to process conventional human languages, such as Chinese and English.

sentiment analysis is a set of NLP methods to detect the tone, or “sentiment” of a lexical item. For example, a basic sentiment analysis would consider the sentence “Great job!” to be “positive”. Sentiment analysers vary in their level of detail and ability to handle sarcasm, double negatives, etc. It is commonly deployed in social value analysis of online communication, and by algorithmic detectors looking for hate speech and other undesired content online.

5.3 Is it “catchword” or “catchphrase”? Or something else?

Key takeaways:

- In English, the word “catchphrase” can generally be considered inclusive of the word “catchword”.
- In English, “catchphrase” and “catchword” have meanings which are distinct from and narrower than that of “slogan”. However, the terms “slogan” and “catchphrase” may be regarded as near-synonyms for many inquiries.

We should note at the outset that contemporary definitions of these terms conflate the traditional meanings of “word” and “phrase”. Traditionally, “phrase” would refer to two or more words. However, the *Oxford English Dictionary* considers the primary relevant definition of “catchword” to be “a

frequently used **word or phrase**, esp. one associated with a particular group or fashionable at a particular time; a topical slogan; a **buzzword**.” (my emphases) [2]. *Merriam-Webster’s New American Dictionary*, more bluntly, simply points us from “catchword” to “catchphrase” [1].

This appears to be a contemporary development. William Safire, in his *Dictionary of Catchwords, Slogans, and Political Usage* (1968), notes that a single word would be a “catchword”, while a series of words using a catchword should be regarded as a “catchphrase” [3]. However, trends captured by the Google Books Ngram Viewer show a steep rise in the use of the term “catchphrase” from 1980 to the present, while there has been a slow decline in the use of the term “catchword” over the same period. This suggests a gradual assimilation of “catchword” by “catchphrase”. Accordingly, we can presume that, at least in popular contemporary English use, the term “catchphrase” includes any and all relevant meanings of the term “catchword”.

5.3.1 Slogans

An adjacent field of interest is the study of “slogans”. In the English language, slogans differ from catchphrases in some crucial elements. “Slogans” are more commonly noted in English inquiry - The query “political slogans” returns 18,500 results (Google estimate) on Google Scholar, whereas “political catchword” returns only 587 and “political catchphrase” 508.

Webster defines slogan as “a word or phrase used to express a characteristic position or stand or a goal to be achieved” and “a brief attention-getting phrase used in advertising or promotion”. Both these definitions lack the element of being “used repeatedly” [4]; for example, a company may have a private slogan as a kind of mission statement, which is not commonly used, such as Google’s “Don’t be evil”. We could therefore observe that, while some or most catchphrases began their life as slogans, not all slogans are – or have become – catchphrases.

In the political context, we could consider certain slogans as catchphrases-in-the-making. For instance, various politicians and movements, including then Presidential candidate Joe Biden, candidate for leader of the Liberal Democrats Layla Moran, the UK Conservative Party, and the Reform Party of Singapore, all used the slogan “Build Back Better” as an electoral slogan in the year 2020. The term appealed to a similar formulation, “Building Back Better”, used as a slogan by the United Nations as part of a new disaster relief framework rolled out in the wake of the Sendai nuclear power plant failure. The implication seems to be that recovery from the Covid-19 pandemic might require a similar “rebuilding” effort to that of a natural disaster. In isolation, any of these uses may have been merely a slogan, but with various politicians in different political contexts and parts of the world using the term simultaneously, it has clearly “caught on” and may be regarded as a “catchphrase”.

While this report concentrates on catchphrases – that is, a disseminated or popularized slogan – academic inquirers may wish to include “slogans” in place of, or in addition to, queries for “catchphrases” when conducting keyword-based research. Moreover, in some languages, the distinction between “slogan” and “catchphrase” may be yet more vague, or absent entirely, as noted below.

5.4 “Political” catchphrases – by politicians, about politics, or of political origin?

Key takeaways:

- Depending on the context, “political catchphrases” may be said to originate with politicians or political parties, or more broadly, be characteristic of a political community to ideology.

- Social media catchphrases may originate with politicians or political parties, or from social movements, or may arise seemingly spontaneously from popular activity on social media itself.
- Popular and academic literature may use the term narrowly or broadly, depending on the context.
- Many catchphrases may coincide merely with a single, ephemeral political event; others have a wider social impact.

Catchphrases may be said to be “political” when they are associated with a political leader or figurehead. However, a broader definition may include catchphrases which are not proper to an individual, but are more characteristic of a political movement or community, such as a party, or a campaign, thus appealing to the “group, idea, or point of view” section of *Webster’s* definition given above.

In this regard, we should be alert to the fact that political catchphrases, in their popular and disseminated form, may also evolve beyond the intended meaning of their political author. For example, the term “fake news” was a political catchphrase employed by (then US Presidential candidate) Donald Trump to refer to news which, in his view, made him the object of unfair criticism. Subsequently, it became a key phrase of political actors worldwide, many of whom – like Donald Trump – had reasons to cast doubt on the veracity of public media. However, it has since evolved to describe misinformation and disinformation, which describes not only media regarding a candidate or candidates, but a wide range of public information. In this case, we see a catchphrase evolving from being *characteristic of a person*, to being *characteristic of a group*, to being within common parlance. While a study of this particular catchphrase could focus on its use as a political catchphrase in a limited context, it seems to have acquired a “life of its own” which may be insensible to ignore.[5]

Papers using the term “political catchphrase” make use of both the narrowest sense of the term (the catchphrase of a political figure), the broader sense of the term (the catchphrase of a political group, idea, event, or point of view), and to signal a popularized term with origins as a political catchphrase (such as “fake news”). Indeed, several papers using the term which rank highly (as of May 2023) on Google and EBSCO searches for “political slogan” and “political catchphrase” searches consider terms which are “not just a political catchphrase”, for example, to trace the histories of such terms as “welfare state”, “sustainable development”, and similar [6], [7].

However, in the political sphere, many catchphrases, by nature, will be ephemeral. We may presume, given the increasing prominence of social media as a news source, that there will be many political catchphrases on social media in the future, but we do not know what those will be. We do not know, even, which of the catchphrases of the 2010s may continue to endure and spread. For every “Black Lives Matter”, a catchphrase which originated on social media itself, there are perhaps a dozen or more failed political slogans which will be less remembered. Phrases like “Build Back Better”, “Get Brexit Done”, and “Make America Great Again” were significant for an election or for a short term in office, and in this context, *were* political catchphrases. Are they still, today? This is difficult to say. The users of these catchphrases – Joe Biden, Boris Johnson, and Donald Trump, respectively – remain politically active, but we do not yet know what, if any, the legacy of these catchphrases will be. Moreover, there have been slogans such as “We Go Higher¹”, “Brilliant Boris²” and (a personal

¹ Used by Hillary Clinton’s 2016 campaign at rallies and events.

² Used by fans of Boris Johnson during the 2019 UK General Election.

favourite) “Skills Wallets³”, which are, perhaps, likely to be forgotten, despite being statistically and sociopolitically significant for a brief period in time.

We may distinguish, therefore, in some cases, between “political catchphrases significant for a particular period” and “political catchphrases which become a cultural staple”, and in some cases, perhaps, a third category of “slogans, many of which may be failed catchphrases”. This is particularly relevant for the fast-paced nature of digital spaces.

5.5 Terms in other major languages

Key takeaways:

- All major languages have words or phrases translatable to “political catchphrase”, however some of these are more tightly defined than others.
- Several major languages conflate “catchphrase” and “slogan”.

Chinese has the word 口号 (Simplified: 口号) for “catchphrase”, which can also mean “slogan”.

Arabic, Persian, and Urdu have the word شِعار (in Urdu, شعار) to express “catchphrase”; this word may also signify “password” or “slogan”.

Portuguese has the word *bordão* to express “catchphrase”.

Spanish does not have a precise term for “catchphrase” as distinct from “phrase”, nor does it have a universally accepted term for “political catchphrase”. Phrases used include *frase política* for “political (catch)phrase” and *eslogan político* – using an English loanword – for “political slogan”. However, *frase política* is uncommon in Spanish academic literature. For a further treatment of the subject, Manuel Garrido Lora, Juan Rey, and Marina Ramos-Serrano have written at some length about the defining of such terms in Spanish [11].

French has the phrases *petite phrase* (lit. “little phrase”) and *phrase fétiche* (lit. “lucky phrase”) to express “catchphrase”; in literature on political catchphrases and on marketing catchphrases, *petite phrase* appears to be preferred, while *phrase fétiche* is preferred in personal contexts (e.g., a favoured phrase for an individual or a fictional character). The English loanword *slogan* is also used.

German uses the words *Schlagwort* and *Slogan* for “catchphrase”, both of which can be conflated with the English word “slogan”.

6. Context: Social media platforms and communities to consider

6.1 Where to start?

Key takeaways:

- While the largest platforms globally are Facebook, YouTube, WhatsApp, and Instagram, smaller platforms may be of interest in specific regions or for examination for political catchphrases specifically. Researchers may wish to consider carefully the platforms and communities most likely to contain political discussion and/or advertising in their region or language community of interest.

³ A slogan promoted by Jo Swinson’s campaign during the 2019 UK General Election, largely spreading via criticism and ridicule.

- Certain communities may exist which, while not being overly political, may engage in political catchphrase use.

As of January 2023, data compiled by Meltwater suggests that Facebook remains the social media website/platform with the largest number of active users, at nearly 3 billion. This is closely followed by YouTube at 2.5 billion, WhatsApp and Instagram (both of which are owned by Facebook) at 2 billion each, WeChat at 1.3 billion, and TikTok at 1 billion [8].

It is important to note that even social media platforms, and in some cases individual communities, have populations which are sizeable enough to have a political impact, and/or which are political by description or in nature. For example, “r/politics”, a community for US political discussion on the discussion platform “Reddit”, has over 8 million users alone (as of May 2023). This community has been described as left-leaning (in the American political context) [9].

Moreover, sizeable communities exist which, while not overtly political by description, may engage in political, or “parapolitical⁴”, activity, due to a certain demographic interest. For example, large, loosely organized social media groups exist in K-Pop (Korean popular music) fandom, and while not expressly political, these may be populated by younger people with a strong interest in music, and accordingly intersect strongly with left-wing political interests in some regions. This became evident when members of the K-Pop community on TikTok and Twitter engaged in spontaneous political activism by registering for large numbers of free tickets for a campaign rally for President Donald Trump during the 2020 US Presidential Election, with apparently no intention to actually attend. Subsequently, a large number of seats were left empty. While the campaign officially blamed physical protests blocking access to the venue, journalists and opposition politicians estimates that the K-Pop community’s campaign had some impact. [10]

In specific regions, certain platforms may have an outsized following for news and politics. Researchers concentrating on a specific region may wish to focus on platforms of interest in that region among news consumers and/or people with political interests. In many countries, data pointing to this may be available from market research or regulatory sources. For example, in the UK, press regulator Ofcom releases an annual report called “News Consumption in the UK”. According to their 2022 data, 32% of UK adults use Facebook for news, slightly exceeding the proportion using the state broadcaster’s BBC News television channel (see *Fig. 1* below). Moreover, Facebook, Twitter and Instagram all rank higher than any major UK newspaper at 16% or higher⁵, however YouTube reaches only 8%. Young users may also prefer use of social media for news and politics; according to the 2022 Ofcom report, Instagram is in fact the largest source of news among people aged 16-24.[11]

⁴ I use this word to describe political activity which is not formally party political and which does not necessarily engage in electoral politics, but is political for other reasons, for example, known for espousing a political position on a certain issue, and/or aspiring to political influence, e.g. by lobbying of politicians. In the UK, an organization such as the Royal Society for the Protection of Birds may be described as parapolitical; it is not necessarily political in nature (it is concerned with the conservation of bird species), but it regularly engages in political communication, particularly regarding the environment and Brexit. Similarly, the American Civil Liberties Union is an organization which offers legal assistance and lobbying on issues of personal liberty and freedom; it has espoused positions which may be described as both left wing and right wing in the American political context, and is not formally party political. Moreover, they do not endorse parties or politicians. However, they are regularly engaged in political campaigns.

⁵ This data excludes membership magazines and newsletters; certain union periodicals and interest groups may have further reach, but this is not captured by Ofcom.

Top 20 news sources – trend data

% of all adults 16+ using each source for news nowadays

	2018	2019	2020	2022*
TV channel				
BBC One	62%	58%	56%	53%
ITV/ITV WALES/UTV/STV	41%	40%	41%	35% ▼
Newspaper (print + website/app)				
Facebook	33%	35%	34%	32%
BBC News Channel	26%	23%	21%	24%
Sky News Channel	24%	23%	25%	23%
Radio station				
BBC website/app**	23%	25%	23%	23%
Social media				
Twitter	14%	16%	17%	17%
Channel 4	18%	17%	18%	17%
Instagram	9%	13%	14%	16%
Daily Mail/Mail on Sunday	18%	18%	17%	15%
WhatsApp	10%	14%	13%	14%
Google (search engine)	17%	19%	15%	12% ▼
BBC Two	14%	11%	11%	11%
BBC Radio 2	12%	12%	12%	11%
The Guardian/Observer	11%	11%	10%	10%
BBC Radio 4	10%	9%	9%	9%
Channel 5	10%	10%	8%	8%
BBC Radio 1	9%	9%	9%	8%
YouTube website/app	5%	6%	6%	8% ▲
Sky News website/app	6%	7%	8%	7%
Other website/app				

Source: Ofcom News Consumption Survey 2022 – COMBINED F2F & ONLINE sample

Question: D2a-D8a. Thinking specifically about <platform>, which of the following do you use for news nowadays?

Base: All Adults 16+ - 2022 W2*=2792, 2020=4576, 2019=4691, 2018=4618

*2022 W1, and 2021, data not shown because face-to-face fieldwork was not possible during Covid-19 pandemic. **Includes Welsh language version
Green/red triangles indicate statistically significant differences between 2022 and 2020 (at 99% confidence level)

Figure 1: Ofcom (UK) data on the top 20 sources of news for UK consumers in 2022. Note Instagram, Twitter, Facebook use compared to major newspapers. (C) Ofcom, 2022 [11]

6.2 Who sees political catchphrases on social media?

Key takeaways:

- In the anglosphere, social media users may be presumed to be younger, left-leaning, and somewhat more female than the general population. Facebook, however, is older and more ideologically balanced in the US and UK.
- In the global south, demographics may vary considerably. For example, in Southeast Asia, nearly 3 in 4 social media users are male.
- Careful consideration of the demographics of inquiry on a regional or per-platform basis should be critical to the contextualization of focused research.

In the anglosphere, social media users tend to be younger than average, with above average rates of participation by women [12]. Perhaps in consequence, social media users are often left-leaning (see Fig. 2 below) [13]. On Facebook, however, the population has been suggested to be more conservative-leaning; this may correlate with Facebook's older-than-average user base [13][12]. On the other hand, a large-scale experiment on political social media reach by Huszár *et al.* has suggested that right-leaning content drives more engagement (that is, user interactivity such as reading and disseminating) than left-leaning content (See Fig. 3 below) [14].

YouGov, an online pollster, frequently collects data on internet usage across the world (see Fig. 4 below). This data illustrates that for a number of countries, the internet *is* social media. In Poland, for example, the European country with the highest social media news consumption rate in a 2022 YouGov survey [15], much of the internet infrastructure relies on 3G and 4G mobile networks, with comparatively little penetration by conventional wired broadband infrastructure such as DSL

(internet over phone line) and cable. Hypothetically, this may have resulted in a larger number of users relying on mobile phone connectivity for internet, and consequently, on social media. This hypothesis has been substantiated in similar environments; famously, over twice as many people in India have a smartphone (about 50% in 2020) than have running water (about 20% in 2020).

In Indonesia, the leading country in the YouGov survey for using social media for news, over 68% of the population have a smartphone, compared to less than 5% who have a broadband internet subscription [16][15][17]. Moreover, in many countries with high social media penetration, mobile phone packages come with free or discounted data usage for social media networks; in consequence, a mobile network user can use the Facebook app without expending their monthly data allocation.

It is also important to note that demographic trends internationally can also vary sharply on a national or regional basis. For example, in South Asia, market analysts estimate that as many as 3 in 4 social media users are male, in contrast to the female-majority global north [8].

Certain social media platforms may be preferred for political interest at a certain time. While comprehensive data on this phenomenon is difficult to acquire, we can take clues from observations regarding key political figures and their followings. Data compiled from various sources (see *Table 1* below) suggest that, on individual platforms, the most followed political accounts vary somewhat, suggesting different platform preferences for political interests on a national or regional basis. They may also reflect changing preferences in social media over time.

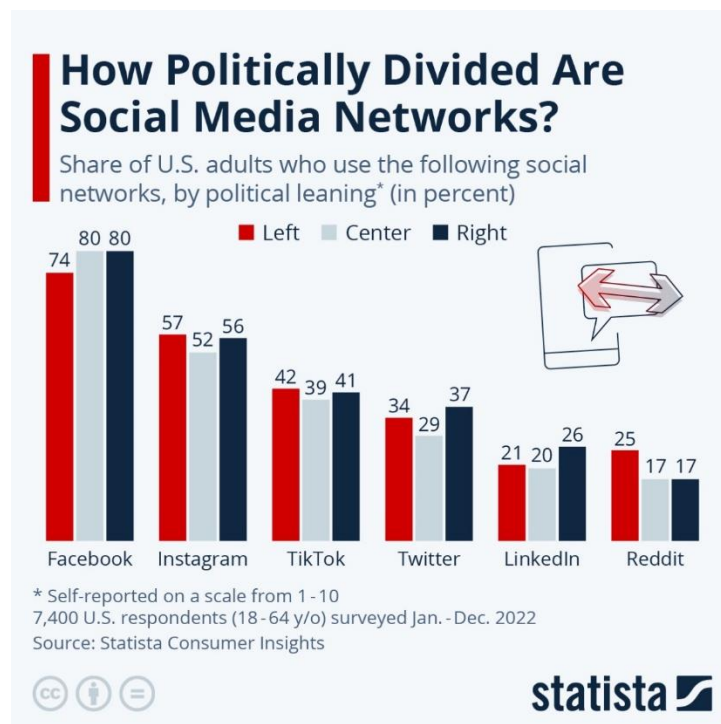


Figure 2: Political left-right leanings of social media users in the US. Data compiled by Statista, 2021-2022. (CC) Statista, 2022 [13]

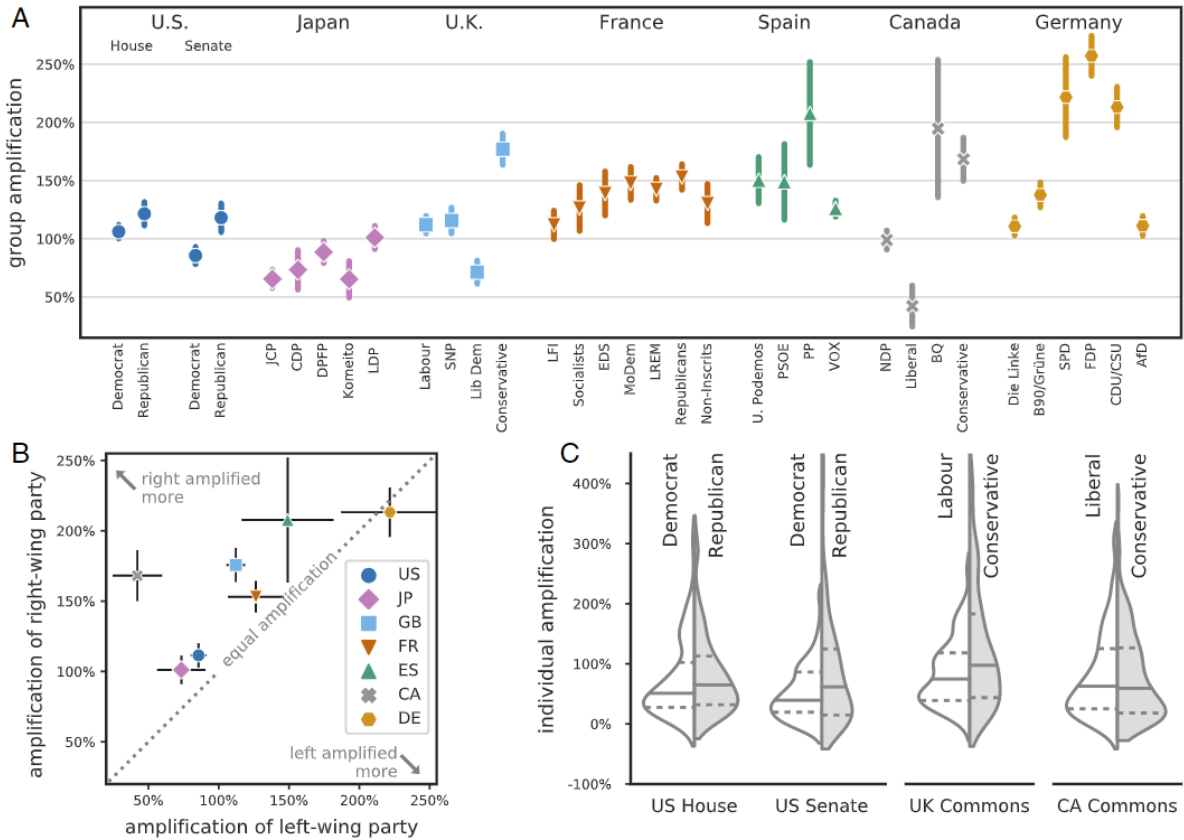


Figure 3: amplification (spread) of political content by political party (left-right positioning is determined by the Chapel Hill Expert Study). From Huzsar *et al.*, "Algorithmic amplifications of political content on social media". [14]

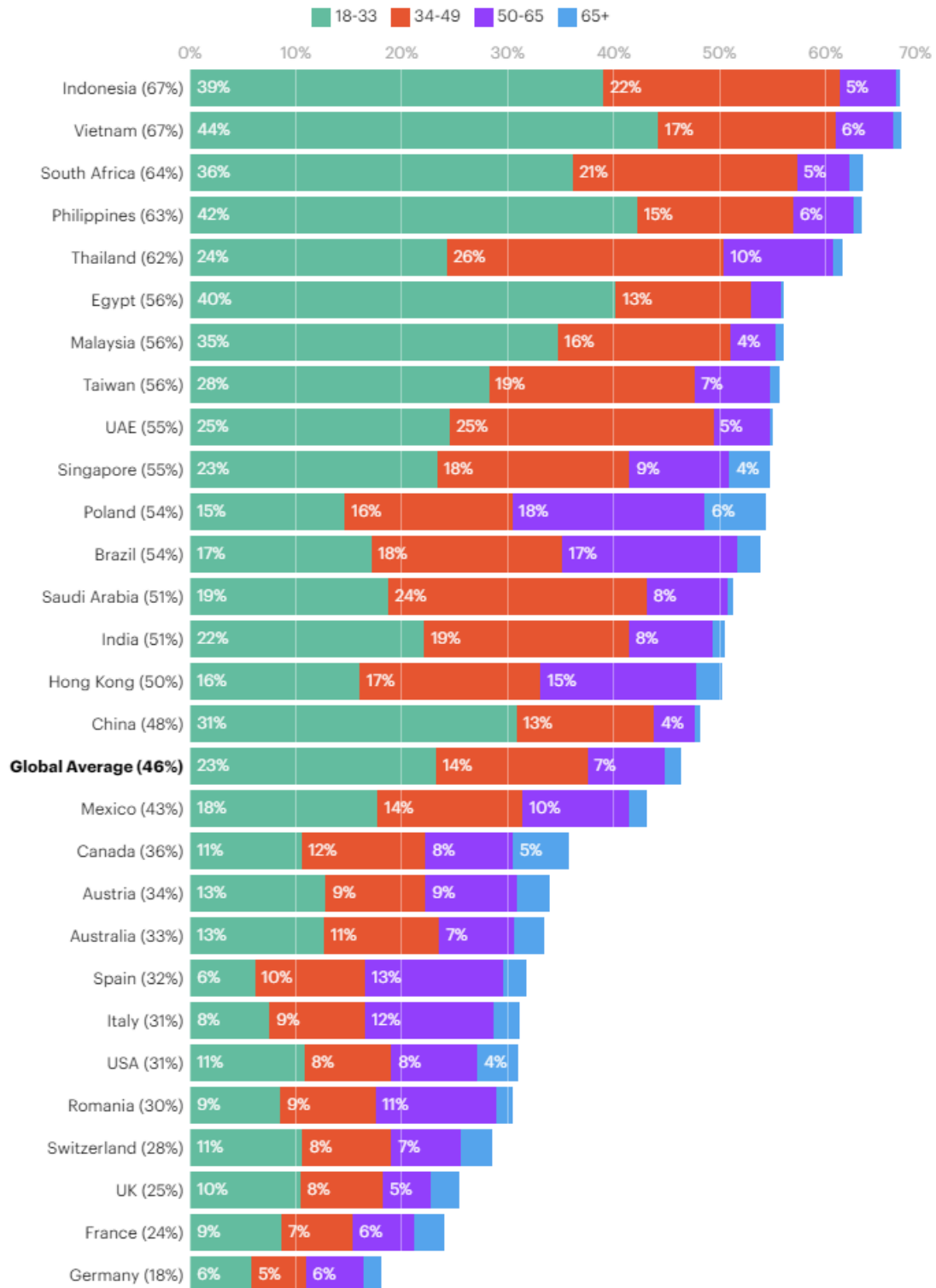
For example, while the largest political account on Twitter and Facebook remains the former US President, Barack Obama, his Instagram following is dwarfed by that of Indian Prime Minister Narendra Modi, and significantly lags that of Indonesian President Joko Widodo. This may (1) reflect the rising popularity of Instagram as a news source in recent years, and/or (2) reflect the social media picture in Indonesia and India respectively, where significantly more users have Instagram accounts than Twitter or Facebook accounts.

However, the use of Twitter as a particularly political platform can be argued; of the top 10 political accounts on Twitter, all of them appear in the top 200 accounts, compared to only three (Modi, Widodo, and Obama) on Instagram. And, uniquely on Instagram, Barack Obama's account has fewer followers than that of his wife, former first lady Michelle Obama.

To summarize, the demographics of political catchphrase exposure may vary considerably on a platform and regional basis, because of the differing usage profiles of social media users. It may be important for researchers to consider carefully the target demographics of their research and to select social media websites for inquiry accordingly. Quantitative analysis of these platforms via market analytics and datapoint observations may inform that contextual research.

Use of social media as a source of news around the world

Which, if any, of the following sources do you use to access news? Please select all that apply.
(Showing % yes for "social network websites")



YouGov

YouGov Global Profiles: April 2022 • Embed

Figure 4: Data by YouGov (compiled November 2022) indicating the percentage of internet users using social media as a source of news, (C) YouGov, 2022-2023. [13]

Name	Description	Rank on Twitter (followers)	Rank on Facebook (likes)	Rank on Instagram (followers)
Amit Shah	Minister of Home Affairs, India	5 th (33.2m)	*	*
Arvind Kejriwal	Chief Minister of Delhi	7 th (27.0m)	8 th (8.0m)	*
Barack Obama	President of the US (2008-2016)	1 st (132m)	1 st (55.1m)	3 rd (35.5m)
BJP4India	Indian political party	10 th (20.3m)	*	*
Donald J. Trump	President of the US (2016-2020)	3 rd (87.3m) †	3 rd (34.1m) †	4 th (23.4m) †
Hamdan bin Mohammed Al Maktoum	Crown Prince of Dubai	*	*	9 th (15.7m)
Hillary Clinton	Former US Presidential Candidate (2008 and 2016)	6 th (31.4m)	4 th (9.7m)	*
Imran Khan	Former President of Pakistan <i>and former professional cricketer</i>	*	5 th (9.7m)	*
Jair Bolsonaro	Former President of Brazil	*	*	6 th (19.3m)
Joe Biden	President of the US	4 th (37.2m)	*	7 th (17.7m)
Joko Widodo	President of Indonesia	*	*	2 nd (52.5m)
Justin Trudeau	Prime Minister of Canada	*	9 th (6.8m)	*
Kamala Harris	Vice President of the US	*	*	8 th (17.2m)
Mitt Romney	Former US Presidential Candidate (2012)	*	6 th (9.6m)	*
Najwa Shihab	Indonesian political journalist <i>and actress</i>	*	*	5 th (23.1m)
Narendra Modi	Prime Minister of India	2 nd (88.0m)	2 nd (44.7m)	1 st (75.2m)
Rahul Ghandi	Former Indian MP	7 th (23.3m)	*	*
Rashtrapati Bhevan (Droupadi Murmu)	Official account for the President of India	9 th (21.9m)	*	*
Recep Tayyip Erdoğan	Prime Minister of Turkey	*	7 th (9.3m)	*
Tehreek-e-Insaf	Pakistani political party	*	10 th (6.6m)	*

Table 1: List of top 10 political accounts on selected social media platforms. Accounts deemed "political" by user labels (on Facebook) and by me (on Instagram and Twitter). Facebook data aggregated by Trackalytics, captured 21 May 2023 [18]. Twitter and Instagram aggregated by ViralPitch, captured 20 May 2023 [19], [20]. The source data is publicly visible on the social media platforms. The description notes the most recent political office held of each account (if of an individual). If a user has other characteristics potentially contributing to their popularity, this is noted *in italics*. The list omits royalty or religious figures who are not also holders of political offices.

* Not in the top 10 on this platform.

† Donald Trump's accounts were frozen by Twitter, Facebook, and Instagram for about two years. As of May 2023, he has not posted to Twitter since the restoration of his account. However, he has resumed activity on Facebook and Instagram.

7. Context: Automation and disinformation

Key takeaways:

- Algorithmic recommendation, bot farms, and disinformation commonly intersect with the dissemination of political catchphrases.
- Algorithmic recommendation and bots are marketing tools which are not inherently negative, but they can be undesirable in political contexts.
- Mitigation strategies suggest increase of transparency, accountability, and in some cases, new law and regulation.

Two somewhat separate but closely adjacent fields of study are the study of automation and the study of disinformation. With regard to political catchphrases, we may identify three subfields of particular interest: (1) Content recommender algorithms which aid the dissemination of political catchphrases, (2) “Bot farms” which spread content on behalf of political actors, and (3) disinformation which is designed to manipulate political activity through deceit.

7.1 Recommendation algorithms

It is reasonable to assume that recommendation algorithms play a substantial role in the dissemination of political catchphrases, as they are a driver of much of the social media content users see. As a basic example, if one is looking at the “News Feed” on Facebook, or the “For You” tab on Twitter, one is looking at algorithmically enhanced content designed to engage the user more than a simple list of the most recent content. Google search is another example of a sophisticated recommendation algorithm, which most of us use every day.

Algorithmic enhancement usually comes as the result of other user actions. When evaluating the reach of a comment, post, or other item, marketers often distinguish between “organic” and “paid” reach. “Organic” content refers to dissemination which has not, to the best knowledge of the platform, been paid for; i.e., it is not “paid” advertising. Organic reach is preferred by most advertising algorithms. For example, if a user clicks or shares an advertisement, this may result in the advertisement being *cheaper* overall (on a per-impression basis).

The reason for this is that learning algorithms of this type work on a system of *reward* and *punishment*. When a user clicks on an ad, or comments with a political catchphrase, this organic engagement “rewards” the algorithm. It says “show this ad to more people like this person who engaged”. In this way, the recommender can reach users who even the original ad creator wasn’t intending to target.

Sometimes, reach can be purely organic. #BlackLivesMatter is a good example; this was not, to our best knowledge, a paid campaign. It was organically produced, disseminated, and popularized. As soon as people start clicking and using the hashtag, the hashtag is recommended to more and more similar people. The political potency of this is readily apparent – by leveraging algorithmic spread, political actors find demographics of support they may not even have known existed.

In a sense, this resembles a conventional “focus group” where lines and slogans are tested by political actors, usually political parties, to a small, often carefully selected audience of voters or activists. However, the mechanisms of a community united by a political slogan on social media are opaquer than those in the offline world, and the idea that “the algorithm” may know a user better than they know themselves is often unsettling. Academic examination of online political engagement has suggested that social media recommendation algorithms effectively increase political polarization by exposing users to increasingly partisan content [21].

7.2 Bot farms

The word “bot” in this context refers to automated or semi-automated accounts, which are created by, or more commonly, purchased by, political actors, in order to disseminate content or manipulate a political narrative.

The phrase “bot farm” refers to technologies or agencies which create large numbers of bots to manipulate content, often during an election or campaign. In the context of political catchphrases, this may be to help disseminate a political message; for example, bots may be used in order to power the dissemination of a hashtag, in the hope that it will also be seen and used by a number of human actors.

The use of bot farms is not always sinister; for some, it is simply a marketing tool. However, in the political sphere, where transparency, accountability, and democracy are often prized, it can be a means of circumventing conventional protocol or breaching laws and regulation. It can also give the illusion that organic support for a politician is greater than it actually is. For example, in the 2016 US presidential election, it was noted that a number of fake accounts were tweeting the catchphrase “#TrumpWon”, among others, after key milestones in Donald Trump’s ultimately successful campaign. Some of these accounts appeared to belong to people living in “swing states” – areas of the US with outsized influence over the Electoral College – or belonging to minority ethnicities who were key targets in the campaign [22]. In some cases, we should be aware that the same bots may be shared by several different political actors. A recent study of bot farms in Turkey has suggested that bots-for-hire would often share content on behalf of opposing parties or movements, and well as, in some cases, for non-political marketing campaigns [23].

Awareness of bot content on social media platforms is increasingly high, which has led some users to accuse others of being a “bot” even when they are not. In the 2019 UK general election, many users on Facebook used the catchphrase “Brilliant Boris” to reply to campaign posts by then Prime Minister Boris Johnson (see Fig. 5 below). The phrase was so ubiquitous that other users started to suspect the involvement of automation, however, an investigation by the British Broadcasting Corporation suggested that most or all of the “Brilliant Boris” commenters were genuine, and some were even satirically pretending to act like bots, either to ridicule Johnson’s voters, who were perceived to be machine-like in their support, or to add to the perception that bots were present [24]. Adding further complexity some campaigners who supported Johnson encouraged repeating particular “bot-like” catchphrases to “drive Remainers nuts”. These layers of catchphrases and counter-catchphrases, happening quickly and spontaneously on social media, can prove challenging to annotate accurately.



Figure 5: Supportive (left) and satirical (right) catchphrases used in the 2019 UK general election campaign on Facebook, captured by BBC News. Comments are copyright to their respective authors but have been anonymized by the BBC. [24]

7.3 Disinformation campaigns

Paid actors have also been known to manipulate the political conversation by posing as key voters or members of certain demographics. For example, various nation states have been known to engage in disinformation campaigns which engage real people to act as if they were representative of a political person or group, or encourage fringe elements of such groups to rise to greater prominence [25]. Like bots, disinformation has been observed to intensify in more politically active regions [26].

In the 2016 US presidential election, disinformation researchers and counter-disinformation actors identified a number of accounts which appeared to belong to real individuals or groups, but were engaged in spreading “fake news” to a potentially disruptive effect. In Fig. 6 below, we see some accounts flagged by the Clemson University Media Forensics Hub, one of which is authentic, but two of which are not. [27]

The spread and manipulation of catchphrases is often key to a successful disinformation campaign. Contemporary examples include the portmanteau catchphrases “scamdemic” and “plandemic” to refer to conspiracy theories regarding the Covid-19 pandemic. While not, arguably, inherently political, these can be leveraged by political actors to cast doubt on decision makers, the health industry, and in some cases, to spread political hatred such as racism and antisemitism [28].

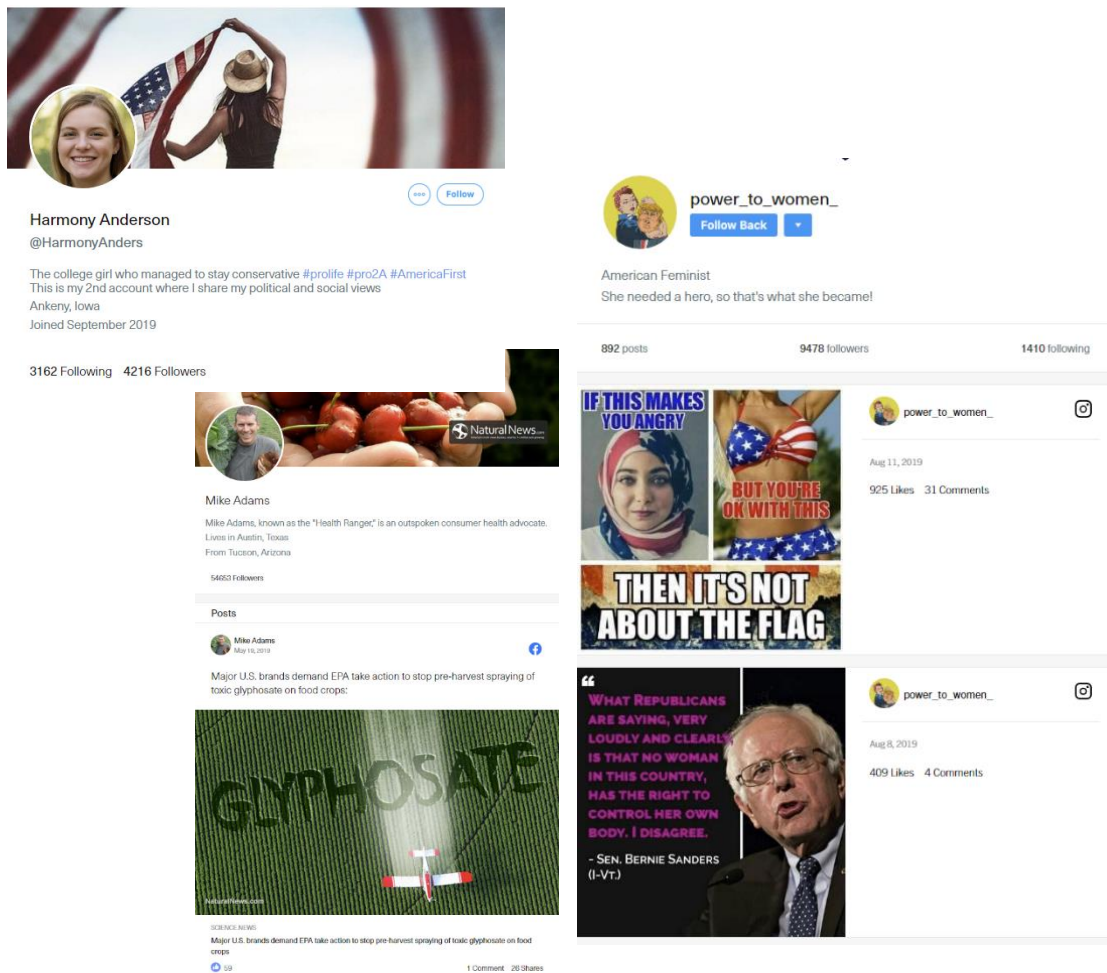


Figure 6: Three accounts identified by Clemson University's SpotTheTroll training initiative as potential disinformation actors. None of these are bots – they are or were authored and managed by real people. Two are disinformation accounts which the researchers connected to a Russian state-sponsored disinformation campaign. One is run by a US native and voter. Tweets and posts are copyright to their respective authors, as shown. For more details visit www.spotthetroll.org [27]

7.4 Mitigation techniques for disinformation, algorithmic bias, and bots

Many mitigation strategies for the above focus on **transparency**. This may refer to the transparency of an algorithm, a social media account, or a platform in general. Facebook’s “Ad Library” and other tools can be considered an effort at transparency.

A similar concept to transparency is **explainability**. Since the precise mechanisms of algorithms are often not readily understood, even by computer scientists who employ them, it is sometimes unclear how transparency offers any insight into how algorithmic bias may come about. The concept of explainability extends this to being able to help a user understand how algorithmic decisions are made, however simply.

Ethicists have suggested that **accountability** is also important; that digital events involving algorithmic influence should be understood in terms of their social context. In this respect, it may be argued that an algorithm may itself have a sociopolitical nature; it is not necessarily “just a tool” [29].

Social media enthusiasts, in particular Twitter owner Elon Musk, have attempted to curtail the influence of bots on their systems, but with mixed success. In some cases, a certain level of bot use can be a positive thing, in making automation accessible to everyday users. Bots on platforms such as Discord and Twitter often have a thoroughly positive utility, offering desired feature extensions and content. This may make a comprehensive ban on automated accounts undesirable.

Awareness of the rapidly changing regulation space is important, as regulation, self-regulation and law, particularly political advertising, plays a key role in the dissemination – or lack thereof – of political catchphrases. France is in the process of restricting the use of Facebook, Twitter, and TikTok from government-issued phones [30]. The EU is considering laws that would mandate further transparency for platforms to catalogue their political adverts []. Meta, in response to existing or threatened regulation, has issued temporary restrictions on political advertising in periods of high sensitivity, and is reportedly considering a blanket ban on political advertising in the EU []. Other proposed regulation includes offering users additional controls over their content, such as mandating that a purely chronological feed – free from recommendation algorithm input – should always be available on all platforms [].

8. Political catchphrases on social media and analysed via computational methods, in academic research

As noted above, there is a substantial volume of research into political catchphrases and slogans on social media. While I have not attempted a comprehensive literature review, some key literature of interest is mentioned below. They are listed in reverse chronological order.

8.1 Books

Brasted, Howard et al. (Ed.)
Religion, Extremism and Violence in South Asia
Springer Nature, 2022

While generally interesting as contextual information for recent political trends in South Asia, Chapter 4 of this book, by **Stuti Bhatnagar**, is entitled *Social Media and Hindu Extremism in India*. It focuses on the role of the ruling BJP in India as a political actor on social media, and intersections between nationalism and violent extremism. As the BJP itself and prominent figures therein are some of the most significant social media actors globally, this has some extremely useful explanations, although it specifically focuses on the darker elements of nationalist political activity.

McLeod, Kembrew and Zimdars, Melissa

Fake News: Understanding Media and Misinformation in the Digital Age
MIT Press, 2020

This book focusses specifically on the catchphrase “fake news” and the history of its development from 2016 to 2020. Although it is not technically deep, there is some interesting exploration as to how catchphrases in general have developed, and how the digital age affects this. It is written in a very engaging style and is suitable for both academic and non-academic readers.

Curini, Luigi and Franzese, Robert

The SAGE Handbook of Research Methods in Political Science and International Relations
SAGE Publications, 2020

As well as being generally a good resource for political science research, chapter 29 of this book, by Curini and **Robert A. Fahey**, focuses on *Sentiment Analysis and Social Media*. Sentiment analysis is a key tool in the data scientist toolkit for social media. This chapter is written primarily for a social science audience seeking to understand this, and uses some pertinent political content on social media as an example.

Lilleker, Darren G., et al.

Social Media Campaigning in Europe
Routledge, 2019

This is a brief but excellent summary of social media campaigning; while it does not say much about slogans, it is extremely useful for understanding the context of a social media campaign and the whys and wherefores of political actors in the European social media space. There is also some important attention to the rapidly changing space of social media regulation in Europe.

Penny, Joel

The Citizen Marketer: Promoting Political Opinion in the Social Media Age
Oxford University Press, 2017

Joel Penny, an Assistant Professor at the School of Communication and Media at Montclair State University, has written this book on “hashtag activism” and other forms of political communication on social media. There are mentions of catchphrase use throughout, although it is not often labelled as such.

Partington, Alan and Taylor, Charlotte

The Language of Persuasion in Politics: An Introduction
Taylor & Francis, 2017

This textbook, by two linguistics professors, examines the use of political language (offline and online) from a linguistic standpoint. There is some useful treatment of social media politicking via language, including slogans, although somewhat dated.

8.2 Some Relevant Papers: General

Note: DOI is given as a hyperlink where available. Where DOI is not available, I have attempted to provide a link to the document from the title.

Note: Given the changing nature of politics on social media, special attention is given to papers written since 2016. This is not to suggest earlier papers are not valuable, and researchers may wish to include other earlier, seminal works not listed here.

Irene Russo et al. have proposed and tested a methodology for analysing “proto-slogans”, that is, lexical items uttered in political communities on social media that, while not necessarily being catchphrases themselves, are similarly predictive of ideological leanings. This is extremely valuable and interesting work, given the difficulties with NLP and predictive linguistics. See [Share and Shout: Proto-Slogans in Online Political Communities](#) (JLCL, 2022).

Lee Artz has written on the potency of social media to create social movements, focusing on the creation of the catchphrase “Black Lives Matter” from an online hashtag. See *Social Media and Social Movements* (Protest, 2022). <https://doi.org/10.1163/2667372X-01030003>

Ioana Literat and **Neta Nliger-Vlienchik** have written an oft-cited analysis of 1651 youth-created videos created for the 2016 US Presidential Election. Many of these are described or linked via catchphrases. As video analysis is computationally expensive, scholarship of this type is rare and valuable. *Youth collective political expression on social media: The role of affordances and memetic dimensions for voicing political views* (New Media & Society, 2021).

<https://doi.org/10.1177/1461444819837571> They have also written other papers of interest in this space, many of which also have general applicability.

Constance de Saint Laurent et al. have written on anti-immigrant community building via hashtag activism. This is an excellent demonstration of how catchphrase analysis can be used to map communities online in the social sciences. See *Malevolent Creativity and Social Media: Creating Anti-immigration Communities on Twitter* (Creativity Research Journal, 2019).

<https://doi.org/10.1080/10400419.2020.1712164>.

8.3 Some Relevant Papers: Regional use; Specific elections

While some research exists in almost all recent elections and regions with notable political activity, I present a few papers here regarding specific ones which may be of interest, or as a starting point for further research.

Manuel Garrido-Lora (University of Seville) has co-written several relevant papers on the subject regarding the use of political slogans in Spanish elections. Some are only available in Spanish, but many are also available in English translation. Of particular interest is *Strategy and creativity in the use of political slogans: A study of the elections held in Spain in 2019* (Communication in Society, 2022). <https://doi.org/10.15581/003.35.3.155-171>

Wenhao Bi has written on the reception of political catchphrases by young Chinese people; he particularly focuses on popular reception of the 2016 presidential election in Taiwan. See *Playing politics digitally: young Chinese people's political feelings on social media platforms* (Cultural Studies, 2022). <https://doi.org/10.1080/09502386.2021.1912808>

Muhammad Umair Chaudhary et al. performed an analysis of social media impact in the 2018 Pakistani general election, and in doing so analysed the dissemination of political slogans during that period. The paper is quite short, but their methodology may be of interest. See *Use Of Social Media In Electoral Process During General Elections 2018 In Punjab, Pakistan* (Multicultural Education, 2021). <https://doi.org/10.5281/zenodo.5142596>

Leticia Cesarino writes on the use of catchphrases in Brazilian populism (original in Portuguese). See *How social media affords populist politics: remarks on liminality based on the Brazilian case* (Tabalhos em Linguística Aplicada, 2020) <https://doi.org/10.1590/01031813686191620200410>.

Ruth Breeze has written an informative paper on Twitter use by populists in the UK, with some interesting observations about key catchphrases. See *Exploring populist styles of political discourse in Twitter* (World Englishes, 2020). <https://doi.org/10.1111/weng.12496>

Alida Maria Silletti has collected a database of political slogans used in the French election 2017, and written an interesting paper (in French) on how they spread online. A fantastic example of how slogans may “translate” to catchphrases. See *Les slogans de l'élection présidentielle française: des "petites phrases" potentielles?* (Lingue e Linguaggi, 2019). <https://doi.org/10.1285/i22390359v29p315>

Xin Xin and **Falik Hartig** have written separate but similarly themed papers on the use of social media catchphrases by the Chinese government, leveraging state agencies to aid in their dissemination. See Hartig, *Political slogans as instruments of international government communication* (Journal of International Communication, 2018) <https://doi.org/10.1080/13216597.2018.1444664> and Xin, *Popularizing party journalism in China in the age of social media: The case of Xinhua News Agency* (IAMCR, 2018) <https://doi.org/10.1177/2059436418768331>. There has also been research into the *reception* of such slogans; see **Bi** above.

9. Technical resources and approaches

9.1 Data science basics – tools and principles

Key takeaways:

- Most computational analysis will require familiarity with a data viewing tool (such as a spreadsheet application) and a programming language (such as Python or R).
- Computational analysis can handle qualitative or quantitative data, but it may be useful to annotate, or at least observe, your data typology early in the process.
- For some types of analysis, it may be useful to build familiarity with non-relational data structures, particularly graph databases.

Key tools

Every data scientist’s toolkit will be a little bit different, depending on their personal preferences and what they need for a given project.

For political data, I tend to rely on **Python**, a programming language, and **Microsoft Excel**. Generally, we will want at least a working familiarity with one technical analysis tool such as Python, R, or Scala, and with a basic data explorer such as Excel, Google Sheets, etc. It is possible to run a number of data analysis techniques in Python without any prior Python experience, as much of the code for use in such situations has already been written. There’s a brief technical demonstration using Excel and Python below.

Key principles

Simplifying greatly, data normally comes in two forms, qualitative and quantitative, and in two formats, relational and non-relational. You may also hear the term “no-sequel” or “NoSQL” used among programmers; this is effectively saying the same thing as “non-relational” (SQL is a common relational database format).

Most datasets you encounter involving political catchphrases will include qualitative and quantitative data. Qualitative data includes user-generated text, hashtags, ratings, and social science

annotations. Quantitative data includes advertising analytics and views and reactions on social media. It is useful to mentally annotate what is in the database before proceeding, as this may save time later on in computational data science and in predicting potential errors before they arise.

Political catchword data may be relational or non-relational. Relational databases are organized in rows (also called “tuples”) and columns. Where columns represent individually identifiable data types, they can also be called “dimensions”.

A non-relational database is where there are connections among data, but the connections are not defined by tables of columns and rows. A good pertinent example of a non-relational dataset is a collection of Twitter users who all follow, like, and retweet one another. Some analysts may visualize this as a cloud with various users connected to one another. The below example in *Fig. 7* is an analysis by market analysts Brandwatch of transphobic catchwords on Twitter being used by different communities, including by trans people themselves as critique or in an effort to “reclaim” certain pejorative terms.

Sometimes, the same data can be stored and reproduced in a relational or non-relational way. For example, virtually all websites rely on relational data to function and display information to a user; social media is no exception. If, in our analysis, we’re mostly concerned with the volume and identification of political catchphrases, we probably want a relational structure. However, if we are mapping the *spread* of a catchphrase, we may wish to consider reformatting the data to a non-relational structure. The most common subtype you are likely to encounter is a **graph database**.

Non-relational databases are not covered in the technical demonstration, but there is a useful guide on PythonSimplified.com: <https://pythonsimplified.com/introduction-to-graph-databases/>.

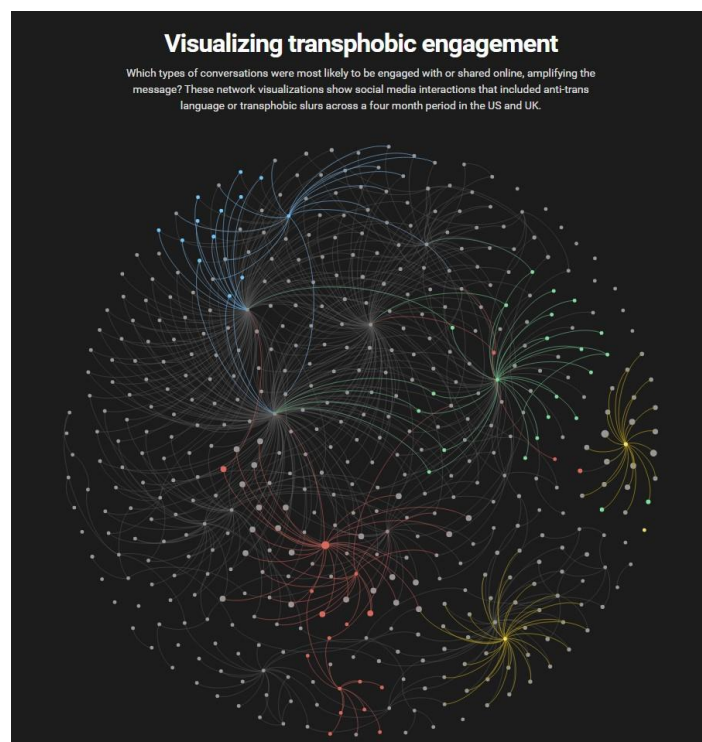


Figure 7: A non-relational visualization expressing interactions between different Twitter users. Each colour represents a different Twitter community, linked by mutual engagement, which has repeated transphobic catchphrases or slurs.
© Brandwatch, 2019 []

9.2 Key datasets for Reddit, Facebook, and Instagram

Key takeaways:

- **Reddit** has a very accessible API.
- **Facebook, Reddit, and Instagram** can be explored using Meta's research tools, such as the **Meta Ad Library** and **CrowdTangle**.
- You can also find and download relevant datasets on **Google's datasetsearch** and on **Kaggle** (for all platforms).

Google datasetsearch

You can search for relevant datasets using Google's "datasetsearch" at:

<https://datasetsearch.research.google.com/>

Not every database available will be here, but many will. If you're conducting research in this area, it would be good to check back here every few weeks or months to see if there are new additions. I recommend using the terms "political catchphrases", "political slogans" and "political".

Meta tools (Facebook and Instagram, 2018-present)

As part of a self-regulating approach to increase transparency following the 2016 US Presidential Election, Facebook has segregated their political advertising library and made it open access.

Not all political ads will appear here, but most will. To search these ads, we can access the Meta Ad Library at <https://www.facebook.com/ads/library/>. From there, we select our desired location of inquiry and in "Ad Category" select "Issues, Elections or Politics". We can also add keywords to this search; this may be a good place to enter candidate catchphrases which we believe to be in use from prior research.

In this library, we can see how many impressions (views) each ad has had, when it ran, how much was spent on it, and who paid for it. We can download a csv (comma separated values file) of this data, up to three times a day. A csv is a relational database separated by commas; each line is a row and each column in that row is separated by a comma (or sometimes a semicolon).

CrowdTangle (Facebook, Instagram, and Reddit)

CrowdTangle is a research tool made by Meta which acts as a limited API for their platforms. It's a little bit more powerful than the ad browser in that it lets you retrieve data on and monitor posts from all public content producers (not just advertisers, and not just political advertisers). You can apply for an account if you are a principal investigator in a research project. See more at

<https://www.crowdtangle.com>

Kaggle (Most platforms, selected periods 2019-present)

Kaggle has created a number of useful datasets, including a dataset of Indian political tweets, a dataset of Democrat vs Republican tweets, Indonesian political hoaxes, and more. You can download these directly from Google datasetsearch, or you can visit Kaggle.com.

Twitter and Facebook hashtags

Twitter has severely restricted access to their API recently, and moreover, has a rather unreliable search feature. However, it is extremely easy to search hashtags. Simply enter the hashtag into Twitter's search feature, or click on the hashtag if you see it, and all public tweets including that hashtag will be visible. This also works on Facebook.

TAPS Dataset (Twitter, 2016-2020)

Harvard has collected an extremely useful catalogue of political tweets in English from 2016 to 2020. This is a big database, so handle with care! You can explore selections of the dataset here:

<https://dataverse.harvard.edu/dataset.xhtml?persistentId=doi:10.7910/DVN/QG1HQF>

Reddit API (2008-present)

Reddit is one of the most openly accessible social media sites for researchers, and is also home to a large degree of political content. I have not used the Reddit API myself, but there's a very useful guide on doing so with Python here: <https://towardsdatascience.com/how-to-use-the-reddit-api-in-python-5e05ddfd1e5c>

Database of Stopwords

Stopwords are words which are likely not relevant for most textual analyses, like articles and prepositions. We can use a database of stopwords to help programmatically exclude these from a text corpus when using NLP. There are several of these available, but a commonly used one is the University of Glasgow's here: http://ir.dcs.gla.ac.uk/resources/linguistic_utils/stop_words. There is also one built into Python's NLTK toolkit, which you can import along with your modules. Be cautious using this for catchphrase analysis, however, as some of our catchphrases may include stopwords.

Other tools

A number of tools designed for marketers have been used and mentioned in this survey, such as **Brandwatch** and **Trackalytics**. There are also repositories, such as **GitHub** and **Statista**, which may be used by researchers. There are also tools designed for research collaboration, such as Zhouhan Chen's **Safe Link Network**. These tools and repositories may be very useful in viewing data on social media which, while public, may be challenging to collect or collate.

While marketing tools are often very useful, it is important to be aware of limitations; they are not often very closely reviewed, and their attention to political data can be minimal. For example, on social media aggregator ViralPitch, the account of the Indian President is listed as representing a male (she is not). Given hers are some of the largest accounts on social media globally, this type of mislabelling could significantly skew data analyses (for example, popularity by gender).

9.3 Gathering data in private networks such as WhatsApp and Discord

Relevant useable datasets of WhatsApp and Discord data, for example, tend to be absent due to the private nature of these networks. However, if one is studying extremism, marginalized groups, gamers, or other communities, or if one simply has an interest in these platforms, this will invariably necessitate data gathering in these contexts.

Discord has an API which allows the use of Discord bots which can be invited to a server. Conceivably, with server users' permission, such bots could be used to gather and aggregate data from a specific server. While I do not know of such a technique being employed, it is technically feasible.

WhatsApp and similar chat networks have an added layer of inaccessibility, in that they are often end-to-end encrypted; that is, the users sending and receiving the data have effectively full control over the spread of that data, unless it is copied in some way by someone who has been invited to that chatroom. If you are invited to a relevant group, WhatsApp does include an export feature where you can easily create a file of the chat on your system or connected server. However, given the inherently private nature of WhatsApp, this should be agreed with the chat participants if used in research, and should probably not be published, except, arguably, in an anonymized or pseudonymized form.

I have used Discord chat data in conjunction with a formal research interview, using a documented and ethically approved technique. This is, arguably, the most obvious way of using private networks – i.e., to treat them as private conversations (however largely populated they may be) and use validated social science instruments designed for that purpose.

Invariably, research of private groups will require a high degree of ethical awareness and oversight, as well as a working familiarity of privacy laws in the EU and UK, most notably, GDPR and EPrivacy. One may not always be able to rely exclusively on a university's policies or the policies of research tools, as there is a wide degree of non-compliance across these networks.

9.4 Some relevant computational data science methods

Data exploration

Data “exploration” refers to examining key features of your data, such as the size, shape, and type of your data. If your dataset is small enough, popular software programs like Excel are often very useful for this. Larger datasets, however, will often require proprietary software or other tools.

Python has a function called “describe” which will tell you how large a relational database is in columns and rows. You may also want to use visualizations (see below) to assist your exploration.

In a dataset containing political catchphrases, this may be able to tell you how many items you have, roughly what they are, and how the data is organized.

Data visualization

“Visualization” refers to seeing your data. For example, a chart or a graph is a visualization. Python has a lot of modules one can download and add to Python to do visualizations. The most popular are **matplotlib** and **seaborn**.

Data cleanup

“Cleanup” refers to organizing your data in a way where you can process it more effectively. This is particularly necessary if you want to do machine learning to it. This might involve removing some dimensions in the data which you don't need or which have a lot of blank values. We may also need to make sure our data is being properly interpreted – for example, are numbers being interpreted as numbers? Do some numbers representing qualitative data, such as user ratings?

Dimensionality reduction and feature reduction

If you have a dataset with very high dimensionality, you may want or need to trim this down. This is particularly important if you need to do machine learning. This is true for programmatic reasons – ML approaches are often less effective with high-dimensionality datasets – but it's also sometimes true for logical reasons. For example, some of the data in your dataset may simply be not relevant to your present inquiry, or you may only wish to examine part of it. In a political catchphrase dataset from social media, you will often have both time series data (e.g., when a social media post or advertisement was made), information about its spread and reception (like shares, “likes”, etc.), and some descriptive information, such as who posted it and where. Often, with political data, we find ourselves cutting out one of these and analysing it in isolation, or sometimes each in turn. This can be done with a variety of data science methods; the most common is **principal component analysis**, or **PCA**, an algorithmic function which attempts to computationally find the most distinguishing dimensions in a dataset. Various PCA methodologies are easily achievable in Python.

Sentiment analysis

Sentiment analysis is a means of categorizing natural language in terms of its sentiment, that is, positively or negatively. Several sentiment analysis training datasets for machine learning exist

online; by comparing a corpus of political catchphrase lexical items to these, it may be possible for researchers to analyze the general negativity or positivity of those catchphrases, or their context.

Cluster analysis

Cluster analysis is a popular methodology for finding groupings in the data, and is somewhat common in political analyses. It can be done visually by an expert, but it can also be done with the assistance of machine learning. For example, if you are mapping the ideologies of individuals onto a left-right axis or “political square”, you might find that those individuals clump together in different areas of the chart, depending on their demographics and the political landscape of their country or region. In data analysis, we have machine learning tools we can use to find clusters in an n-dimensional space (although this becomes less viable at increasingly large dimensionalities). Cluster analysis and dimensionality reduction often go hand-in-hand in terms of narrowing down a dataset to a smaller number of features. This can add understanding and clarity to data, particularly data about human societies.

Knowledge graphs

Knowledge graphs are a form of data architecture which attempts to find, create, and/or document links between data or data dimensions. A knowledge graph is also known as a **semantic network**. Search engines and social media employ knowledge graphs in the backend to inform recommendation algorithms. For example, if we find that users who like cats are more likely to vote Labour in a UK general election, we can add an “edge” to a graph to express this connection. Political catchphrases can inform knowledge graph edges, linking a community with a political actor or campaign. A theoretical use of knowledge graphs in this context would be to engage in ideological profiling of users on the basis of catchphrase use. This is a large field of research, and may be considered an academic specialism in its own right.

9.6 Technical demonstration of basic catchphrase data exploration and analysis

The following paragraphs suggest a methodology for processing political catchword data from a download of Facebook posts. It uses Python and Microsoft Excel.

This demonstration is also available as a **Jupyter Notebooks** (.ipynb) file, and as a pdf. You can download these from www.emyn.ulusive.net/catchphrases

9.6.1 Getting the data

For this purpose, we’ll use CrowdTangle (discussed above) to retrieve a set of recent Facebook posts from US Governors. This function has already been set up by my research department, the Knowledge Media Institute at the Open University, so I just need to select our US Governors feed from a series of options.

Meta offers an algorithm on CrowdTangle to identify posts which are “overperforming” vs other posts by a set of authors. For my dataset, I’ve chosen overperforming posts from the past 30 days, and downloaded a csv of this data (see *Fig. 8*).

Looking at the CrowdTangle display allows us to do some initial data exploration. For example, we can see that the posts have view counts, reactions in the form of “likes” and other icons expressing feelings, comments, etc. We can also see that some posts have images or video – we won’t be able to download these to the csv except as a link, so it might be important to annotate some aspects of them now, if we think they’re meaningful. For example, I might note that the top result is a clip from Fox News, which is known for conservative-leaning content.

We also note where we might find catchphrases – most notably, in the body text of each post.

9.6.2 Understanding the data's context

At present, a small majority of the governors of US states are members of the Republican party. This is also a politically active season, with a number of candidates for President announcing their candidacy. This includes the Governor of Florida, Ron DeSantis. Depending on the project, I will probably do some further contextual research at this stage, but this may suffice for present purposes.

9.6.2 Initial data exploration in Excel

Since this is a pretty small dataset, it's not going to be computationally difficult to open it up and make sense of it in an everyday office program like Excel. We could use Python for this, but for smaller datasets, Excel can be more clear and more useable, especially for people unfamiliar with code.

Viewing our downloaded csv file in Excel, we can see that various features have been organized in columns, and each post is a row. That's what we would normally expect. We also note where the body text is – it's in a column called "Message". While I'm here, I'm going to copy and paste this data to a separate Excel document, and save that as a csv, "messages.csv". I now have two datasets – one which has only the post body text authored by the Governors' Facebook pages, and another which is the original, unchanged csv. I'm also going to save the message text as a .txt file, "messages.txt". This is so we can do naive NLP without considering who wrote what, and it will be quicker to process in Python when doing so.

If you're working in certain languages, you may notice that Excel doesn't display some characters accurately. Don't worry about this – they should be fine in the raw csv. If you're unsure, open the csv in a text editor, like Notepad, and ensure that the original characters have been saved.

We can also note that we have some quantitative data (such as the number of likes) and some qualitative data (the message text, links, etc). We also have timeseries data (the date and time of each post's creation).

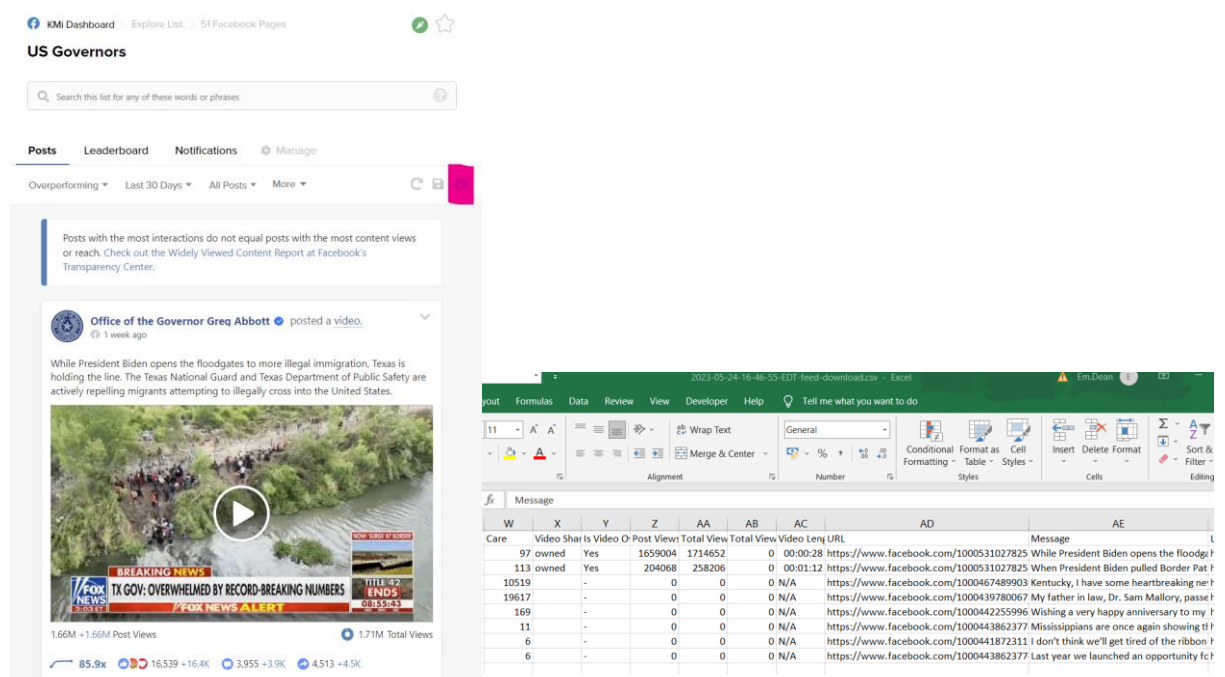


Figure 8: Left: Sample screenshot from the CrowdTangle platform. The download to csv icon is highlighted in pink. Right: the resulting csv, opened in Microsoft Excel.

9.6.3 Finding catchphrases in the “message” text

Python offers us a number of tools for NLP and textual analysis. We’re going to start with a basic tool in Python called “counter”. We’re also going to open the messages.txt file we just made, and make our csv into a Python list. Once we’ve done so, we can simply count the number of words that appear using “counter”.

First, let’s import our Counter module, which is in “collections”. We’re also going to need “defaultdict”, which adds a few common Python commands. Lastly, we’ll import “re”, which are regular expression operations or “RegEx”, which lets us process some files more easily.

```
from collections import Counter
from collections import defaultdict
import re
```

Next, we’ll open our file “messages.txt”. I’ve specified the encoding as UTF-8, because that’s how I encoded the text file. You may need to change this depending on your system.

```
file_path = r'messages.txt'

with open(file_path, encoding="utf8") as file:
    text = file.read()

counts = Counter(re.findall('\w+', text))
print(counts)
```

When I call “print(counts)”, Python tells us what the most common words are. Unsurprisingly, these are all generally common words in English, like “the” and “to”. However, we can scroll down through this list to find something more interesting.

We can use a stopwords dictionary like the University of Glasgow one noted in “datasets” above to trim this down. But we want to be careful to not exclude catchphrases that include these words. Conversely, we might also want to add words to our stopwords dictionary, like “com” and “ly”, internet domain suffixes which are being picked up by the counter. However, this dataset is small enough to parse as is for now.

Other than our generally common words, some frequent words used are “Mother” (29), “family” (36), “nation” (43), “border” (29), and “Texas” (75).

This is where some contextual knowledge is important - Mother’s Day, in the US, is in May, and several governors posted about this. We might also think that the Governor of Texas is overrepresented in this dataset. But if we compare our word counts to the source data, we will find that actually, this isn’t the case – several governors have been recently posting about Texas in conjunction with anti-migration messaging.

We can also use some simple code to look for hashtags in the text.

```
textList = text.split()
for i in textList:
    if(i.startswith("#")):
        x = i.replace("#", '')
        print(x)
```

Here, I've created a new variable called "textList" which splits our text into a list of words. We then have a "for loop" which looks for items in the list starting with "#", and then lists those without the leading "#".

Some of these won't be catchphrases, but some might. With a little political knowledge, we might sort some wheat from the chaff. For example, "#NationalLibraryWeek" might be connected to recent political discussion concerning the banning of books. Even tastier, we can see there's a hashtag used many times, "#OperationLoneStar", which seems to refer to efforts on the part of the Texas government to restrict movements of immigrants, refugees and asylum seekers on the border with Mexico.

9.6.4 Finding catchphrases with an initial visual analysis

One of the most basic things we can do with word frequencies is to create a "word cloud", which makes words larger based on the frequency of their use in the dataset. This is a type of visual analysis, and lets us view the data in a more unstructured way.

We can do this using many different tools on the internet, but while we're in Python, we can use the **wordcloud** module. This includes a stopwords module. We'll also need a few other modules – **numpy** and **pandas**, which are popular maths modules, and **matplotlib**, which we'll use to make the visualization. When we import them "as", we can make shortcuts so it's quicker to write our code.

```
import numpy as np
import pandas as pd
from os import path
from PIL import Image
from wordcloud import WordCloud, STOPWORDS, ImageColorGenerator
import matplotlib.pyplot as plt

wordcloud = WordCloud().generate(text)

wordcloud = WordCloud(max_font_size=50, max_words=100,
background_color="white").generate(text)
plt.imshow(wordcloud, interpolation='bilinear')
plt.axis("off")
plt.show()
```

Looking at the data in this way might let us spot some things we didn't before. For example, "proud" is a commonly used word, as is "school". There are also multiple words we could consider as a single catchphrase, for example, "families" and "family" are both prominent, and we might want to look at our source data to see if these are used in similar phrases.

We could go further with this data! If you want to copy this code and try a few more data science methods, please download the ipynb file here and experiment: www.emyn.ulusive.net/catchphrases

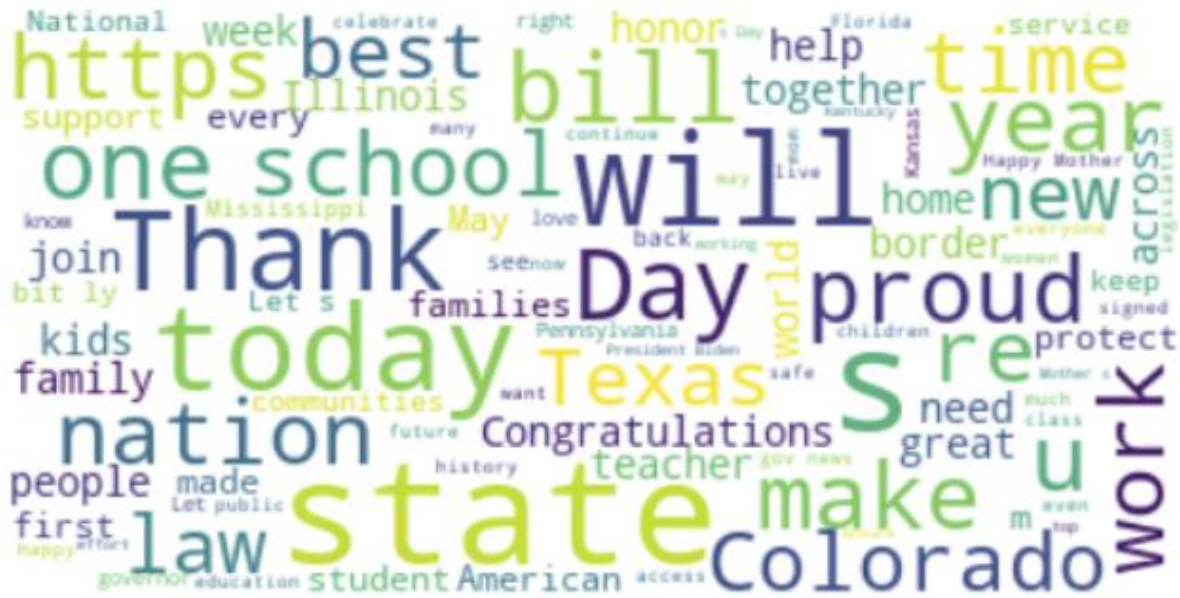


Figure 9: Wordcloud of US Governors' Facebook posts, April 25 – May 25, 2023

10. Concluding remarks

The spread of political catchphrases in the digital sphere has been rapidly accelerated by social media use and by social media architecture itself. This has given impetus to increasing collections of datasets concerning political catchphrase use, as well as computational research.

This survey has suggested contextual information, key tools, and methods to approach political catchphrases using computational methods, with a focus on social media as an emerging corpus of political activity.

11. Works Cited

Note: Additional suggested reading is listed in Section(8) above.

- [1] "catchphrase," *Merriam-Webster's New American Dictionary*. Merriam-Webster, 2023. Accessed: May 30, 2023. [Online]. Available: <https://www.merriam-webster.com/dictionary/catchphrase>
- [2] "catchword," *Oxford English Dictionary*. Oxford University Press, 2023. Accessed: May 30, 2023. [Online]. Available: <https://www.oed.com/view/Entry/28833>
- [3] W. Safire, *The New Language of Politics: A Dictionary of Catchwords, Slogans, and Political Usage*. New York: Collier Books, 1972.
- [4] "slogan," *Merriam-Webster's New American Dictionary*. 2023. Accessed: May 31, 2023. [Online]. Available: <https://www.merriam-webster.com/dictionary/slogan>
- [5] B. Woodward, "'Fake news': A guide to Trump's favourite phrase – and the dangers it obscures," *the Independent*, Oct. 02, 2020. Accessed: May 31, 2023. [Online]. Available: <https://www.independent.co.uk/news/world/americas/us-election/trump-fake-news-counter-history-b732873.html>
- [6] T. Helne and T. Hirvilammi, "Wellbeing and Sustainability: A Relational Approach," *Sustainable Development*, vol. 23, no. 3, pp. 167–175, May 2015, doi: 10.1002/SD.1581.

- [7] D. Garland, "The emergence of the idea of 'the welfare state' in British political discourse", doi: 10.1177/09526951211034412.
- [8] S. Kemp, "The Changing World of Digital in 2023," *Meltwater*, Jan. 26, 2023. <https://www.meltwater.com/en/blog/changing-world-of-digital> (accessed May 31, 2023).
- [9] K. Alkiek, "Unpacking Political Activity on Reddit," *Medium*, Aug. 31, 2022. <https://medium.com/@kenan.r.alkiek/political-users-on-reddit-83926d7354c6> (accessed May 31, 2023).
- [10] J. Hollingsworth, "Why K-pop fans being credited with disrupting Trump's rally shouldn't be a surprise | CNN," *CNN*, Jun. 22, 2020. Accessed: May 31, 2023. [Online]. Available: <https://edition.cnn.com/2020/06/22/asia/k-pop-fandom-activism-intl-hnk/index.html>
- [11] "News Consumption in the UK: 2022," Jul. 2022. Accessed: May 31, 2023. [Online]. Available: https://www.ofcom.org.uk/__data/assets/pdf_file/0027/241947/News-Consumption-in-the-UK-2022-report.pdf
- [12] "Demographics of Social Media Users and Adoption in the United States," *Pew Research Center*, Apr. 07, 2021. <https://www.pewresearch.org/internet/fact-sheet/social-media/> (accessed May 31, 2023).
- [13] K. Buchholz, "How Politically Divided Are Social Media Networks?," *Satista*, Feb. 28, 2023. <https://www.statista.com/chart/29395/political-preference-social-media/> (accessed May 31, 2023).
- [14] F. Huszár, S. I. Ktena, C. OBrien, L. Belli, A. Schlaikjer, and M. Hardt, "Algorithmic amplification of politics on Twitter," *Proc Natl Acad Sci U S A*, vol. 119, no. 1, p. e2025334119, Jan. 2022, doi: 10.1073/PNAS.2025334119/SUPPL_FILE/PNAS.2025334119.SAPP.PDF.
- [15] S. Tan, "Global: To what extent do people use social media to catch up on the news?," *YouGov*, Jul. 29, 2022. <https://business.yougov.com/content/43271-global-social-media-as-source-of-news-2022> (accessed May 31, 2023).
- [16] P. Taylor, "Fixed broadband penetration in Indonesia 2001-2021," *Statista*. <https://www.statista.com/statistics/517739/fixed-broadband-subscriptions-per-100-inhabitants-in-indonesia/> (accessed May 31, 2023).
- [17] "The State of Mobile Internet Connectivity 2021 Connected Society," Sep. 2021. Accessed: May 31, 2023. [Online]. Available: <https://www.gsma.com/r/wp-content/uploads/2021/09/The-State-of-Mobile-Internet-Connectivity-Report-2021.pdf>
- [18] "Most Liked Politician on Facebook," *Trackalytics*, May 2023. <https://www.trackalytics.com/facebook/fanpage/likes/category/Politician/page/1/> (accessed May 31, 2023).
- [19] "Top 200 Most Followed Twitter Accounts," *ViralPitch*, 2023. <https://viralpitch.co/topinfluencers/twitter/top-200-twitter-influencers/> (accessed May 31, 2023).
- [20] "Top Instagrammers: Top 200 Most Followed Instagram Accounts," *ViralPitch*, 2023. <https://viralpitch.co/topinfluencers/instagram/top-200-instagram-influencers/> (accessed May 31, 2023).

- [21] F. P. Santos, Y. Lelkes, and S. A. Levin, "Link recommendation algorithms and dynamics of polarization in online social networks," *Proc Natl Acad Sci U S A*, vol. 118, no. 50, p. e2102141118, Dec. 2021, doi: 10.1073/PNAS.2102141118/SUPPL_FILE/PNAS.2102141118.SAPP.PDF.
- [22] I. Kottasova, "A third of pro-Trump tweets are generated by bots," *CNN Business*, Oct. 18, 2016. Accessed: May 31, 2023. [Online]. Available: <https://money.cnn.com/2016/10/18/technology/twitter-bots-donald-trump-hillary-clinton/>
- [23] T. Elmas, "Analyzing Activity and Suspension Patterns of Twitter Bots Attacking Turkish Twitter Trends by a Longitudinal Dataset," Apr. 2023, doi: 10.1145/3543873.3587650.
- [24] J. D'Urso, "The real people pretending to be 'Boris bots' on Facebook - BBC News," *BBC*, Nov. 01, 2019. <https://www.bbc.co.uk/news/blogs-trending-50218615> (accessed May 31, 2023).
- [25] S. Bradshaw and P. N. Howard, "The Global Disinformation Order 2019 Global Inventory of Organised Social Media Manipulation 2019 GLOBAL INVENTORY OF ORGANISED SOCIAL MEDIA MANIPULATION [i]," 2019. Accessed: May 31, 2023. [Online]. Available: <https://demtech.oii.ox.ac.uk/wp-content/uploads/sites/12/2020/08/CyberTroop-Report19.pdf>
- [26] M. Pratelli, M. Petrocchi, F. Saracco, and R. De Nicola, "Swinging in the States: Does disinformation on Twitter mirror the US presidential election system?," vol. 17, Mar. 2023, doi: 10.1145/nnnnnnnn.nnnnnnnn.
- [27] "Spot the Troll: Internet Troll and Disinformation Quiz," *Clemson University*, Oct. 2020. <https://spotthetroll.org/> (accessed May 31, 2023).
- [28] H. D. Lanier, M. I. Diaz, S. N. Saleh, C. U. Lehmann, and R. J. Medford, "Analyzing COVID-19 disinformation on Twitter using the hashtags #scamdemic and #plandemic: Retrospective study," *PLoS One*, vol. 17, no. 6, p. e0268409, Jun. 2022, doi: 10.1371/JOURNAL.PONE.0268409.
- [29] L. Amore, *Cloud Ethics*. Durham, North Carolina: Duke University Press, 2020.
- [30] A. Mascellino, "France Bans TikTok, Other 'Fun' Apps From Government Devices," *Infosecurity Magazine*, Mar. 2023. <https://www.infosecurity-magazine.com/news/france-bans-tiktok-fun-apps-govt/> (accessed May 31, 2023).