

Supplementary Material: Measuring Social and Political Identities In Social Media Self-Descriptions

Clara Vandeweerd, Gregory Eady, Frederik Hjorth, and Peter Thisted
Dinesen

March 12, 2026

Contents

A	Data collection	2
B	Most similar words	2
C	Identity-relevant and -irrelevant clusters	4
D	Manual cleaning of partisan identity cluster allocations	17
E	Validation results	18
F	Time trends for all clusters	20
G	Clusters most associated with each US state	22
H	Clusters most associated with politics	25

A Data collection

The social media bios of roughly 3.4 million active Twitter users who were geo-located to the United States were collected each day beginning in June 2020 and ending in June 2023. To define this sample, in May of 2020, we collected all of the profiles of users who followed at least one among a set of ideologically diverse media outlets that included *MSNBC*, *Huffington Post*, *The New York Times*, *The Washington Post*, *CNN*, *The Wall Street Journal*, *FOX News*, and *Breitbart News*. After these bios were collected, we subsetted these users by those who we classified as “active” users. An active user was defined as a user who tweeted in the last year; who sent at least 25 tweets over the lifetime of their account; and who had at least 10 followers. We then used the user text in the location field and whether a user’s most recent tweet¹ contained geo-location data to identify any user who either self-described themselves as being located in the US (e.g. “New York, NY” in their location field), or had geo-coordinates within the US based on their most recent tweet (Dredze et al., 2013). For the open text of the location field, the Carmen library (Dredze et al., 2013) string-matches this text to a location database. Any user whose location could not be identified was thus excluded from the final sample. In total, 22 million users (out of a total of 87 million) could be geographically located, of which 9.2 million were classified as being in the United States. The final subset of users consisted of 3.4 million of these users who were classified as “active” users.

To collect day-level panel data for these 3.4 million users, we set up a data collection script on an Amazon Web Services server that began collection each day at midnight to collect the bios of all users. At the time of collection, Twitter rate limits for the ‘users/lookup’ endpoint was 90,000 bios per 15 minutes, using the standard Twitter REST API. This meant that the collection of user profiles each day took roughly 3,400,000 users / 90,000 profiles (per 15 minutes) / 4 = ~ 9.5 hours. This collection began in June 2020 and ran until the ‘users/lookup’ end point of the Twitter REST API was curtailed in June 2023 as a result of changes in Twitter’s ownership.

B Most similar words

Table B1 below shows the face validity of the FastText word embeddings trained on our Twitter bio dataset by listing the most similar words to a number of key identity terms. Table C2 lays out the identity-relevant clusters, with their hand-picked names and most frequent words. We focus on the most frequent words from each cluster rather than, say, the words whose embeddings are closest to the spatial core of each cluster. The reason is that core words may be very unusual (e.g. “serial entrepreneur”), while frequent words capture the most common reasons for assigning a bio to a particular cluster (e.g. “professional”).

¹A user’s most recent tweet was returned as part of the call to the ‘users/lookup’ end point.

Table B1: Most similar words to some identity keywords of interest, according to the trained FastText word embedding model. Note that some of these words are rare in the corpus; any word used 20 times or more is included in the embedding space.





















Father	Mother	Republican	Democrat
proud father	mother/gigi	conservative republican	ex-democrat
father of 3 husband-father dad	mother/nana mother/mimi mother of 3	reagan republican ex-republican republican 🐘	centrist democrat liberal democrat progressive democrat
stepfather father of five step-father father of 2 father of twins father of four	mother of twins mother of 2 mother of 4 mother daughter stepmother step-mother	vote republican reformed republican anti-republican moderate republican republican until senate republican	lifelong democrat proud democrat moderate democrat democrat resister staunch democrat prouddemocrat
Red Sox	Lgbt	Black	Christian
boston redsox patriots redsox boston bruins redsox bruins patriots bruins sox bruins patriots red patriots celtics celtics bruins pats bruins	lgbtq lgbtqa lgbtqi lgbt 🏳️ lgbtqia lgbtq 🏳️ lgbtqia 🏳️ lgbtqrighs lgbtq blm lgbtqally	blackity black blacks black brown black af blackfoot blackout blackfish blackman black men black male	christian 🏠 born-again christian devout christian pro-life christian reformed christian orthodox christian christian pro-life christiana evangelical christian christian prolife
Tech	Music	Artist	Bitcoin
technology technolgy ed-tech hi-tech dctech technologys technology gadgets adtech tech startup technerd	reggae music trance music music feeds trap music newmusic music 🎵 sony music electronic music music concerts indiemusic	henna artist trapeze artist vfx artist artist painter artiste artist 🎨 graffiti artist resin artist mixed-media artist mosaic artist	bitcoins crypto bitcoin bitcoin 🚀 bitcoin eth bitcoin dogecoin bitcoin btc bitcoiner bitcoin crypto altcoin bitcoin ethereum

C Identity-relevant and -irrelevant clusters

Table C1: The 75 identity-relevant clusters, sorted by their frequency of occurring in bios, with their 20 most frequent words. Note that some clusters have fewer than 20 words.

Artist	Business Leadership	Positivity	Family Roles	Personal Growth
writer	business	fun	wife	passion
author	consultant	person	husband	experience
artist	leader	happy	2	personal
photographer	passionate about	funny	married	power
musician	expert	strong	kids	success
designer	development	crazy	4	focus
activist	sales	awesome	3	knowledge
blogger	management	cool	children	growth
creator	partner	kind	beautiful	purpose
actor	executive	positive	5	performance
poet	leadership	simple	daughter	experiences
singer	strategist	smart	amazing	action
storyteller	innovation	sweet	boys	process
graphic designer	advisor	active	girls	creativity
freelance	finance	super	happily married	value
singer songwriter	strategy	nice	wonderful	goals
actress	career	curious	best friend	opportunity
dancer	consulting	pretty	daughters	thinking
writer editor	analyst	weird	sons	vision
comedian	small business	loyal	lucky	driven


Community	Higher Education	Services	Business Finance	Mother
community	student	free	services	mom
education	university	come	service	mother
support	college	information	company	wife mom
center	graduate	use	solutions	wife mother
network	mba	welcome	industry	mom wife
public	ms	share	needs	mama
program	ma	daily	financial	dog mom
dedicated	major	ideas	agency	mommy
youth	bs	journey	insurance	grandmother
programs	masters	sharing	clients	grandma
association	ba	bring	market	sister
foundation	political science	inspired	companies	daughter sister
nonprofit	ohio state	meet	businesses	mom of 3
organization	penn state	seeking	firm	momma
society	degree	bringing	professionals	cat mom
communities	c o	guide	planning	aunt
families	b s	artists	experienced	mother of 3
institute	state university	search	investment	mother of two
resources	intern	join	investing	mother of 2
supporting	graduated	created	capital	mom of two





Hobby Emojis	Heart Mind	Technology	Selling	Helping
	heart	tech	online	help
	soul	technology	fashion	helping
	mind	engineer	products	building
	open	data	shop	lives
	times	product	style	create
	dreams	developer	unique	build
	lost	security	original	creating
	age	engineering	custom	mission
	sound	software	hair	grow
	sleep	internet	modern	leaders
	coming	software engineer	buy	connect
	face	architect	jewelry	impact
	turn	mobile	sell	help you
	brain	platform	color	serve
	away	analytics	line	growing
	walk	cloud	luxury	helping people
	fall	ai	store	inspire
	hand	technologist	makeup	opportunities
	stars	code	fine	drive
	inside	startups	plus	improve

Places	Trade	Marketing	Sports Teams	Senior Positions
city	master	marketing	nascar	director
official twitter	high	design	yankees	manager
area	model	digital	red sox	president
office	TM	creative	dallas cowboys	specialist
official	®	social media	blues	senior
county	▪	communications	boston sports	head
district	licensed	pr	philly sports	lead
south	class	content	roll tide	officer
department	bitcoin	brand	packers	vp
west	trade	marketer	cubs	chief
north	crypto	digital marketing	lakers	assistant
town	id	public relations	duke	coordinator
central	trader	communication	orange	executive director
union	cannabis	web	cowboys	associate
division	accountant	advertising	notre dame	vice president
valley	gold	brands	saints	sr
washington	affiliate	creative director	steelers	project manager
park	elite	branding	mets	administrator
regional	ad	digital media	angels	dir
official account	trading	seo	billsmafia	managing director

Volunteer	Business Services	Journalism	Enthusiast	Teacher
member	based	sports	enthusiast	teacher
currently	servicing	news	nerd	school
volunteer	specializing	editor	geek	learning
fellow	providing	journalist	junkie	high school
current	quality	reporter	aspiring	students
past	focused	journalism	fanatic	principal
previously	leading	magazine	addict	teaching
board	provides	weather	obsessed	learner
ex	located	contributor	amateur	teach
comms	provide	covering	aficionado	ed
board member	private	columnist	wannabe	librarian
candidate	innovative	anchor	guru	hs
chair	offering	journal	connoisseur	english
turned	full service	managing editor	extraordinaire	middle school
ambassador	premier	photojournalist	addicted	schools
staff	offers	fox	nut	academy
organizer	including	meteorologist	bookworm	lifelong learner
fmr	we provide	correspondent	freak	math
chapter	founded	reporting	coffee addict	language
rep	provider	cover	ninja	library






Hobbies	Music	Industry	Positive Emojis	Fan
friends	music	commercial	💕	fan
enjoy	world	field	🎯	proud
play	art	safety	🔥	sports fan
reading	photography	construction	🕶️	supporter
running	arts	car	👉	huge
outdoors	jazz	air	👑	lifelong
cooking	r b	driver	🙏	die hard
watching	hip-hop	control	💋	sports fanatic
playing	hip-hop rap	metal	💍	life long
especially	edm	large	😍	big fan
my favorite	hiphop	complete	😊	diehard
cars	classical music	aviation	rip	huge fan
exploring	dance electronic	industrial	😍	cubs fan
fishing	latino music	advanced	😘	hockey fan
course	soundtrack	manufacturing	😊	sports nut
gardening	electronic music	driving	🕊️	die-hard
enjoys	gospel music	technician	cashapp	mets fan
laughing	music concerts	equipment	💕	yankees fan
hiking	digital creators	operator	😈	rabid
eating	indie experimental	delivery	😁	wrestling fan

Father	Tv Film	Foods Drinks	Spiritual	Online
dad	producer	coffee	human	com
father	media	wine	nature	follow
husband father	entertainment	beer	beauty	check out
father husband	tv	eat	light	email
husband dad	film	pizza	happiness	page
dad husband	video	sarcasm	kindness	website
proud father	filmmaker	craft beer	body	follow us
parent	award-winning	chocolate	self	visit
family man	theatre	drink	spiritual	available
grandfather	production	good food	joy	tips
brother	award winning	cheese	spirit	info
christian husband	studio	tea	natural	contact
father of 3		fresh	laughter	check
daddy	theater	bbq	universe	dm
son brother	television	bourbon	magic	site
father of two	videos	ice cream	healing	latest
father of 2	films	whiskey	human being	app
dog dad	productions	tacos	compassion	link
uncle	winning	oxford comma	wisdom	subscribe
grandpa	executive producer	taste	balance	booking




Academia	Justice	Entrepreneur	Identity Politics	Pop Culture
social	peace	owner	 	all things
research	justice	entrepreneur	blm	books
professor	freedom	founder		movies
researcher	social justice	ceo	black	pop culture
policy	government	co-founder	blacklivesmatter	comedy
phd	human rights	investor	matter	movie
phd student	fighting	llc	feminist	disney
assistant professor	fight	business owner	black lives	star wars
ph d	democracy	founder ceo	vegan	shows
scholar	equity	philanthropist	gay	country music
studies	equality	co-owner	queer	horror
phd candidate	rights	visionary	lgbtq	comics
studying	movement	president ceo	ally	reality tv
prof	humanity	inventor	millennial	fantasy
psychology	liberty	coo	pronouns	classic rock
academic	protect	cofounder	humanist	sci-fi
grad student	diversity	managing partner		comic
associate professor	race	ceo founder	pride	broadway
public policy	war	serial entrepreneur	bi	drama
study	civil rights	cfo	unapologetic	music movies





Writing	Outdoor	Healthcare	Disease	Military Police
book	beach	healthcare	survivor	retired
writing	earth	care	mental health	veteran
words	water	nurse	child	vet
read	land	medical	warrior	army
stories	sun	rn	social worker	military
story	ocean	medicine	cancer	us army
blog	farmer	practice	adult	marine
i write	wild	public health	cancer survivor	pilot
write	road	md	disabled	veterans
poetry	farm	physician	pain	navy
series	ride	health care	autism	army veteran
publisher	outdoor	doctor	awareness	air force
amazon	outside	therapy	recovery	leo
press	mountains	registered nurse	childrens	police
published	plant	clinical	adults	law enforcement
publishing	lake	hospital	breast cancer	force
fiction	garden	patient	sex	us navy
reviews	mountain	nursing	addiction	captain
songs	sunshine	patients	counseling	usmc
indie	summer	trained	treatment	soldier

Entertainment	God	Northeast	Nature Emojis	Sports Disciplines
local	god	nyc	🌟	football
events	jesus	new york	💜	baseball
house	blessed	ny	🖤	golf
chef	faith	dc	⚡	soccer
club	christ	boston	💚	basketball
party	lord	new yorker	🌟	hockey
event	grateful	miami	💖	tennis
street	grace	nj	🦋	nfl
restaurant	child of god	pa	🌻	wrestling
bar	jesus christ	brooklyn	🌹	sport
favorite	gods	philly	💛	nba
virtual	god first	atl	🌱	college football
hospitality	god fearing	new england	⭐	athletics
museum	thankful	ct	🐉	basketball coach
celebrity	pray	philadelphia	🌟	head coach
hotel	saved	pittsburgh	🌸	track
destination	saved by grace	new jersey	🌿	fantasy football
season	philippians 4	baltimore	💎	mlb
kitchen	blessings	washington dc	💜	ball
bartender	faithful	buffalo	👉	softball

Social Studies	Sustainability	Coaching	Liberal Democrat	Conservative Republican
politics	space	coach	resist	country
history	energy	educator	liberal	conservative
political	project	certified	democrat	trump
interested	environment	speaker		maga
culture	systems	mentor	progressive	patriot
humor	sustainability	trainer	blue	 
interests	green	therapist	 	republican
issues	environmental	coaching	resistance	proud american
historian	system	counselor	resister	trump supporter
economics	urban	practitioner	theresistance	god bless
philosophy	planet	facilitator	gop	libertarian
literature	sustainable	motivational speaker	fbr	constitution
religion	climate	author speaker	proud democrat	maga kag
government politics	projects	public speaker	dem	christian conservative
arts culture	safe	motivator	bidenharris2020	president trump
general news	housing	speaker author	biden harris	our country
current events	architecture	keynote speaker	democratic	trump2020
business finance	transportation	servant leader	vote blue	kag
topics	climate change	coaches	democrats	conservative christian
feminism	economic	presenter	voteblue	parler

Pronouns	Countries Flags	Us And World	American Emojis	Animals
she her		global		dogs
he him		usa		dog
they them	latina	international		animals
she they	spanish	u s		cats
he they	bilingual	national		cat
she hers		united states		animal
he his		nation	 	humans
she ella	languages	worldwide		pets
he haw	puerto rico	humanitarian		pet
she elle		womens		rescue
		united	 	babies
		americas		horses
	mexico	uk	 	puppies
	 	states	 	puppy
	solo	africa	 	crazy cat
	paris	china	 	pup
		intl		fur babies
	et	caribbean	 	pups
		canada	 	kitty
	translator	india	 	furry

Radio Podcast	Alum	Social Media	Gamer	Origins
host	alum	ig	gamer	american
podcast	grad	instagram	youtube	independent
voice	alumni	follow me	game	citizen
radio	alumna	sc	games	immigrant
dj	alumnus	insta	player	irish
talk	alumn	snapchat	gaming	italian
listen	fsu alum	facebook	video games	jewish
co-host	mba grad		content creator	french
podcaster	wvu alum	fb	anime	proudly
	uga alum	snap	channel	latino
podcasts	mizzou alum	tiktok	youtuber	indian
personality	alumnae	add me	twitch	canadian
podcast host	state alum	soundcloud		mexican
weekly	ucf alum	ig fb	youtube channel	americans
station	usc alum	instagram snapchat	streamer	heritage
morning	msu alum	teamfollowback	stream	african
fm	auburn alum	instagram facebook	video game	hispanic
voices	uga grad	facebook instagram	pc	native american
heard	hbcu grad	tumblr	plays	german
radio host	mizzou grad	fb ig	main	british

Fitness	Christian	Real Estate	Law	Science
fitness	christian	real estate	attorney	science
training	pastor	realtor	law	scientist
dance	church	agent	lawyer	bio
instructor	catholic	broker	legal	lab
yoga	christ follower	estate	matters	stem
athlete	follower of jesus	realty	cpa	data scientist
fit	follower of christ	homes	tax	biology
mma	follower	commercial real		computer science
gym	minister		law firm	data science
personal trainer	jesus follower	selling	judge	physics
triathlete	servant	realtor [®]	criminal justice	biotech
train	ministry	property	personal injury	evolution
exercise	bible	nmls	defense	chemistry
boxing	gospel	properties	federal	
classes	theology	banker	crime	neuroscience
crossfit	worship	mortgage	immigration	sciences
yoga teacher	ministries	realestate	paralegal	biologist
competitive	preacher	residential	employment	scientific
athletes	born again	keller williams	practicing	
pt	kingdom	sellers	criminal defense	mit

Family	Professional	Health Wellness	Travel	Food
family	professional	health	travel	food
god family	pro	lifestyle	traveling	food wine
family friends	profession	wellness	adventure	food drink
family and friends	pr professional	healthy	adventures	foods
faith family	hr professional	nutrition	travels	food and drink
fam	professionally	parenting	travel agent	foodies
familyfirst	professionalism	fitness wellness	travelling	food drinks
familyman	semi-professional	health wellness	trips	foody
familys	profesional	health fitness	travel ✈️	raw food
my familys	paraprofessional	health and wellness	traveled	food bev
family 👨👩	unprofessional	healthy living	next adventure	foodporn
twdfamily	pga professional	registered dietitian	new adventures	food 🍷
extended family	professions	well-being	travelers	food 🍴
family1st	hospitality professional	nutritionist	luxury travel	
trancefamily	professional wrestler	diet	travel bug	
family-man	professional napper	weight loss	outdoor adventures	
spn family	semi professional	lifestyle blogger	outdoor adventure	
	professional firefighters	healthier	travel tourism	
	professionally trained	healthy lifestyle	well traveled	
	profesh	plant based	traveling ✈️	

Table C2: The 25 clusters we judged to be irrelevant to identity, sorted by their frequency of occurring in bios, with their 20 most frequent words. Note that some clusters have fewer than 20 words.

Passionate	Common Fillers	Future	Being Cool	Change
lover	know	way	rock	change
advocate	think	better	baby	truth
friend	going	future	king	hope
traveler	oh	right	aka	hate
foodie	yes	today	queen	true
runner	long	place	band	believe
believer	probably	need	miss	i believe
animal lover	shit	laugh	white	end
avid	die	dream	red	trust
reader	left	learn	hot	stop
dog lover	you know	everyday	pop	choice
passionate	fuck	money	half	men
cook	getting	run	star	respect
dreamer	dont know	watch	hip hop	word
maker	hello	hard	bitch	speak
world traveler	ask	look	guitar	stand
golfer	said	try	sing	reality
political junkie	mean	every day	boss	facts
that order	cause	one day	da	fear
occasional	whats	smile	roll	common sense

Names	Titles	Disclaimers	Life	Cities	States	Tweets
dr		opinions are	life	chicago		twitter
john		tweets are	live	california		account
mr		tweets	life to the fullest	ca		tweet
david		views are	lifes	los angeles		general
brown		opinions	called life	san diego		post
michael		views		colorado		following
james		thoughts		seattle		tweeting
mark		not endorsements		san francisco		tweet about
joe		opinion		houston		rt
jr		views expressed		bay area		related
jack		endorsements		detroit		updates
mike		endorsement		dallas		random
mrs		my employer		las vegas		follow back
chris		thoughts are		arizona		interesting
paul		opinions expressed		austin		memes
smith		not necessarily		sf		feed
rose		retweets are		st louis		personal account
williams		opinions mine		az		retweets
max		rts are		midwest		retweet
taylor		rts		denver		occasionally

Cities States2	Gender Age	Striving	Working	Origin Adjectives
state	girl	living	time	born
texas	man	loving	work	native
florida	guy	looking	working	forever
michigan	old	trying	making	raised
southern	woman	loving life	works	born and raised
tx	young	just trying	job	resident
ohio	single	taking	full time	texan
fl	lady	enjoying	worked	born raised
atlanta	boy	trying to make	working hard	transplant
nc	kid	living the dream	currently working	lived
va	dude	finding	day job	grew up
georgia	gal	enjoying life	working towards	roots
virginia	female	raising	who works	was born
new orleans	small town	livin	workin	birth
ga	male	trying to be	spare time	originally
cleveland	just a guy	hoping	ive worked	moved
minnesota	chick	striving	worked hard	grown
alabama	guy who	chasing	working stiff	hoosier
iowa	your average	girl living	full-time job	jersey girl
maryland	guys	searching	currently unemployed	bred

Numbers	Good	Common Numbers	Stuff	Years Awards
21	good	100	things	2020
20	best	24 7	big	2021
22	great	st	stuff	year
19	real	50	little	champion
18	bad	1 2	small	2019
23	greatest	pm	thing	2018
6	most important	phone	lots	winner
24	worst	est	places	2016
7	worlds best	40	little bit	2017
17	pretty good	99	other things	2022
10	really good	3rd	other stuff	2015
16	even better	hours	many things	2014
25	damn good	90	things happen	2x
13	good guys	number	take pictures	2nd
8	absolute best	miles	simple things	2013
11	good deed	000	cool stuff	2012
12	good egg	0	big ideas	champ
15	greatest joy	million	these things	award
9	mans best	counting	bits	awards
14	good measure	60	things done	champions

Timespans	People	New	Home	Schedule
years	people	new	home	day
20 years	women	1	team	night
30 years	folks	1st	group	days
25 years	ppl	second	teams	week
long time	peoples	newest		weekend
over 20	people places	newly		weekends
over 30	interesting people	newer		nights
40 years	meeting people	anew		during the day
yrs	ordinary people			superhero by night
15 years	people suck			nighttime
10 years	genuine people			nights and weekends
years of experience	people seem			nights sleep
over 25	like-minded people			dayz
35 years	ppls			2day
over 40	rude people			nightside
22 years				night and weekend
23 years				nightowl
18 years				
24 years				
21 years				

D Manual cleaning of partisan identity cluster allocations

First, among the 1790 words from the partisan/ideological clusters we manually filtered out 142 that did not have sufficiently clear partisan content (e.g. “votes”, “commonsense”), and 255 that referred to specific issues. In the Democrat/liberal cluster, many common issue terms referred to mask-wearing and vaccination; perhaps not unexpected, given the salience of the issue in the observed period, and the deep political divide surrounding it in the US (Pew Research Center, 2021). In the Republican/conservative cluster, issue-related terms were overwhelmingly related to gun rights. Not-clearly-partisan terms accounted for 8% of the bio tokens that were originally classified as either Democrat/liberal or Republican/conservative, and issue terms accounted for 10%.

Second, some references to parties, politicians or ideologies are hostile (e.g. “I hate Hillary Clinton”), causing users to end up in the wrong partisan/ideology cluster. To filter out hostile uses, we created a dictionary of around 200 phrases that, in combination with a political term, were likely to signal negativity (e.g. “can’t stand*”, “disgusted by*”). We also identified 42 terms from the partisan/ideological clusters that seemed likely to attract hostility (e.g. “Republicans”, “Dems”); together, these terms were mentioned around 55,000 times in the bios. We then flagged all combinations of a hostile phrase with a political term. We found around 1,800 hostile uses (with negative mentions of Donald Trump accounting for over half of them), and reversed the partisan cluster association of those terms in those bios. The relatively small number of negative mentions of political terms is in line with Lee et al. (2022), who find that negative partisan identities (based mainly on out-party dislike) are relatively uncommon.

To validate our approach to political identification, we hand-coded 200 random occurrences of a political term in a bio. These are terms that fall into either the Conservative/Republican or the Liberal/Democrat cluster, excluding terms we hand-labeled as irrelevant or issue-related (see main text). We found seven hostile uses, of which one was caught by our dictionary of hostile phrases. We found two further cases where a term usually associated with one partisan side was appropriated by another (e.g. “Proud American”, a conservative cluster term, used by a liberal user).

Thus, in a total of eight cases (4%), a bio was ascribed to a political identity cluster opposite to the one intended. Note that in the “predictability of politics” application (main article section 5.2), we required bios to have *more* identifying terms from one partisan/ideological cluster than the other in order to classify it as belonging to that partisan side. Five out of the eight missed hostile/appropriated uses occurred in bios that also had at least as many correctly ascribed political terms (e.g. “Proud American, #TheResistance”). Thus, in this application, only three bios (1.5%) would have ended up in the opposite-than-intended partisan cluster.

Finally, we also found nine cases (5%) where a normally politicized term was used in a non-political way (e.g. “country”, another conservative term, used in the phrase “country music”). These bios were ascribed to a political cluster, although the term they mentioned was in fact apolitical.

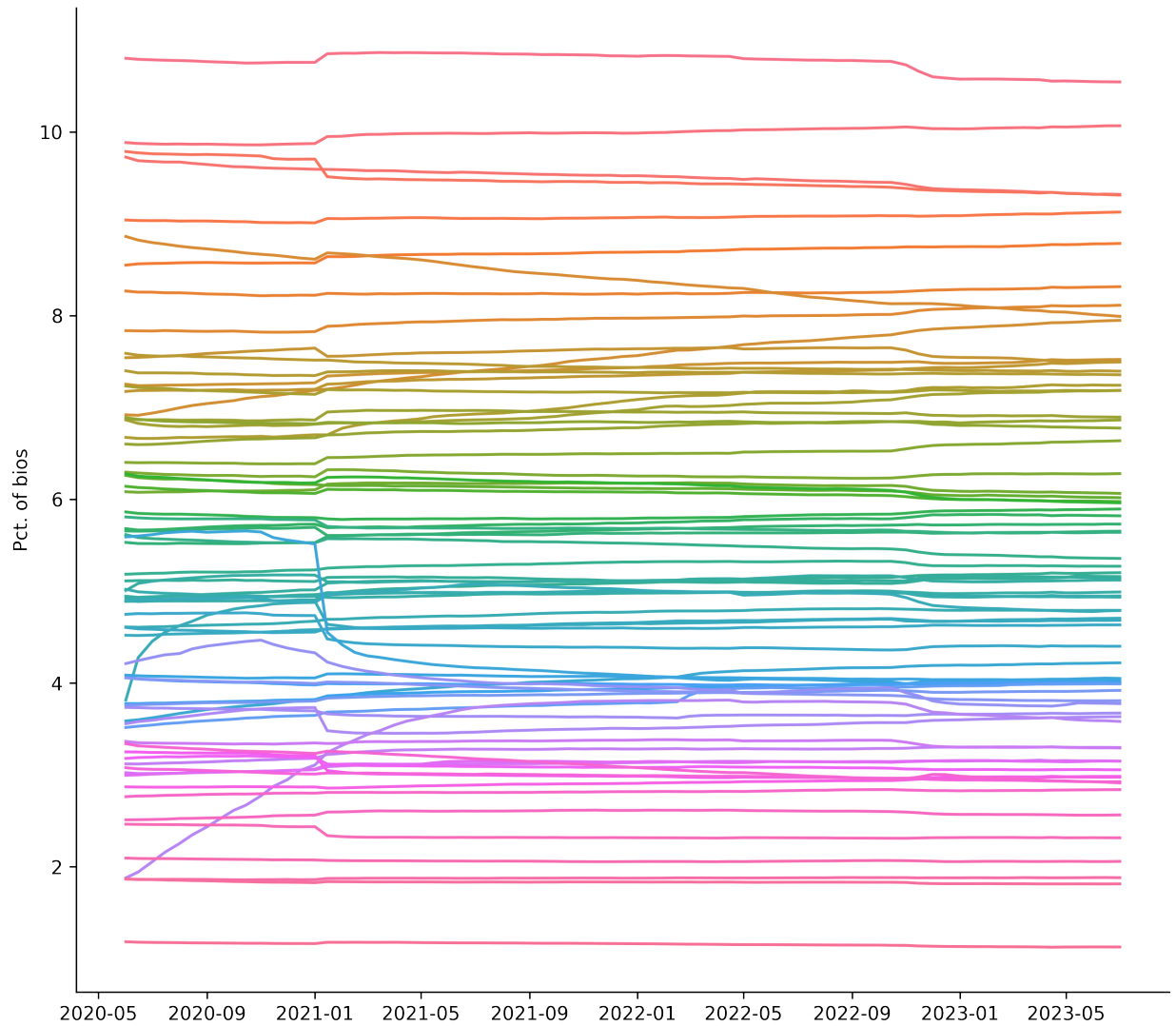
E Validation results

Table E1: Results of comparing our algorithm’s assignment of bios to clusters against three human coders. “Validation” is Krippendorff’s alpha between the algorithms ratings of whether a cluster was present in a bio against the modal rating of three coders. “Intercoder” is the alpha between coders. Based on ratings of 3000 bio-cluster combinations.

Cluster	Validation	Intercoder	Cluster	Validation	Intercoder
Food	1.00	0.97	Fitness	0.60	0.75
Pronouns	1.00	0.93	Social Studies	0.60	0.55
Animals	1.00	0.97	Pop Culture	0.60	0.58
Sports Disciplines	0.95	0.83	Us And World	0.59	0.46
Northeast	0.95	0.77	Music	0.55	0.86
Entrepreneur	0.95	0.83	Tv Film	0.55	0.61
Christian	0.95	0.97	Radio Podcast	0.54	0.77
Health Wellness	0.85	0.80	Marketing	0.54	0.54
Family	0.85	0.77	Entertainment	0.54	0.40
Alum	0.85	0.80	Journalism	0.53	0.65
Father	0.85	0.93	American Emojis	0.53	0.88
Travel	0.85	0.80	Business Finance	0.51	0.36
Mother	0.85	0.93	Law	0.49	0.61
Social Media	0.80	0.93	Outdoor	0.49	0.67
Foods Drinks	0.80	0.93	Academia	0.49	0.69
Conservative Republican	0.80	0.93	Countries Flags	0.47	0.79
Enthusiast	0.75	0.70	Disease	0.47	0.77
Coaching	0.75	0.55	Personal Growth	0.44	0.55
Healthcare	0.75	0.80	Spiritual	0.43	0.44
Identity Politics	0.75	0.76	Selling	0.43	0.83
Online	0.75	0.51	Hobby Emojis	0.41	0.64
Technology	0.75	0.68	Community	0.41	0.68
God	0.75	0.86	Family Roles	0.40	0.93
Higher Education	0.70	0.66	Business Leadership	0.38	0.48
Sports Teams	0.70	0.80	Hobbies	0.35	0.70
Artist	0.70	0.72	Sustainability	0.35	0.71
Real Estate	0.70	0.89	Teacher	0.35	0.65
Writing	0.70	0.89	Origins	0.34	0.63
Science	0.65	0.66	Places	0.31	0.11
Business Services	0.65	0.76	Nature Emojis	0.28	0.53
Positive Emojis	0.65	0.69	Positivity	0.20	0.52
Military Police	0.64	0.92	Trade	0.17	0.89
Gamer	0.64	0.69	Services	0.16	0.53
Liberal Democrat	0.64	0.89	Justice	0.13	0.63
Senior Positions	0.60	0.57	Industry	0.10	0.41
Professional	0.60	0.57	Heart Mind	0.09	0.31
Fan	0.60	0.76	Volunteer	-0.32	-0.09
Helping	0.60	0.60			

F Time trends for all clusters

Figure F1: Percentage of bios containing all identity-relevant clusters, at biweekly intervals from June 2020 to June 2023.



- | | | | |
|---------------------|-------------------|-------------------------|-----------------|
| Artist | Volunteer | Outdoor | American Emojis |
| Business Leadership | Senior Positions | Writing | Pronouns |
| Positivity | Business Services | Pop Culture | Radio Podcast |
| Family Roles | Journalism | Healthcare | Animals |
| Personal Growth | Teacher | Identity Politics | Gamer |
| Community | Hobbies | Military Police | Alum |
| Services | Music | Disease | Origins |
| Business Finance | Enthusiast | Entertainment | Christian |
| Higher Education | Industry | God | Real Estate |
| Hobby Emojis | Positive Emojis | Nature Emojis | Fitness |
| Technology | Father | Conservative Republican | Social Media |
| Mother | Tv Film | Northeast | Law |
| Helping | Fan | Sustainability | Science |
| Heart Mind | Foods Drinks | Sports Disciplines | Family |
| Selling | Spiritual | Countries Flags | Professional |
| Trade | Entrepreneur | Coaching | Health Wellness |
| Places | Online | Social Studies | Travel |
| Sports Teams | Academia | Liberal Democrat | Food |
| Marketing | Justice | Us And World | |

G Clusters most associated with each US state

Table G1: The three identity-relevant clusters that occur most disproportionately in each state, as measured by their χ^2 statistic. Sample sizes are the number of non-empty bios (on the reference date) that could be ascribed to each state.

State	Top 1 Identity	χ^2	Top 2 Identity	χ^2	Top 3 Identity	χ^2	n
Alabama	God	1849.3	Christian	1490.4	Sports Teams	795.1	22541
Alaska	Outdoor	987.6	Military Police	84.3	Hobbies	33.8	3065
Arizona	Conservative Rep.	690.1	Liberal Democrat	401.3	Military Police	395.9	40606
Arkansas	Christian	935.1	God	784.3	Family Roles	485.9	11344
California	Tv Film	6357.6	Technology	4821.5	Artist	2998.5	306308
Colorado	Outdoor	727.4	Sustainability	256.9	Technology	191.2	40776
Connecticut	Places	260.0	Teacher	160.5	Business Services	134.8	22016
Delaware	Academia	43.5	Community	40.9	Business Services	27.2	1786
District of Columbia	Volunteer	9637.4	Us And World	8847.9	Academia	4141.5	57895
Florida	Real Estate	3103.0	Countries Flags	2865.7	Conservative Rep.	1566.9	141476
Georgia	God	1517.8	Positive Emojis	796.9	Christian	536.1	77368
Hawaii	Outdoor	881.0	Spiritual	265.6	Military Police	106.8	6678
Idaho	Outdoor	100.5	Hobbies	68.1	Real Estate	62.9	5715
Illinois	Marketing	157.9	Business Leadership	147.1	Pronouns	122.2	88936
Indiana	Father	536.9	Family Roles	515.5	Higher Education	366.2	30403
Iowa	Teacher	441.7	Father	355.0	Family Roles	281.5	13593
Kansas	Sports Teams	314.2	Higher Education	281.0	Family Roles	227.0	13028
Kentucky	Christian	958.3	Family Roles	677.5	Sports Teams	619.5	19894
Louisiana	Positive Emojis	590.1	God	550.5	Conservative Rep.	244.2	23268
Maine	Journalism	60.9	Pronouns	31.7	Foods Drinks	29.9	2333
Maryland	Healthcare	993.5	Places	363.6	Community	358.0	40989
Massachusetts	Academia	1843.2	Science	1498.8	Technology	1162.0	59146
Michigan	Liberal Democrat	178.0	Mother	172.4	Higher Education	169.5	52706
Minnesota	Pronouns	391.0	Enthusiast	298.0	Father	234.9	36892
Mississippi	Higher Education	1349.6	God	1191.3	Christian	1007.1	9791
Missouri	Father	311.7	Family Roles	264.8	Teacher	219.4	33338

State	Top 1 Identity	χ^2	Top 2 Identity	χ^2	Top 3 Identity	χ^2	n
Montana	Outdoor	453.8	Places	91.2	Hobbies	35.5	3829
Nebraska	Father	327.1	Sports Teams	239.5	Teacher	215.6	10675
Nevada	Entertainment	269.3	Positive Emojis	209.5	Conservative Rep.	200.0	20997
New Hampshire	Conservative Rep.	136.8	Liberal Democrat	108.7	Outdoor	77.1	7280
New Jersey	Teacher	153.6	Places	100.5	Business Services	63.0	50883
New Mexico	Origins	130.6	Liberal Democrat	77.7	Countries Flags	73.5	7426
New York	Tv Film	4806.3	Artist	2997.3	Writing	1882.8	220998
North Carolina	Christian	547.5	Higher Education	476.4	God	455.0	54529
North Dakota	Sports Teams	133.2	Outdoor	63.7	Sports Disciplines	60.0	2449
Ohio	Fan	556.5	Family Roles	529.5	Mother	467.9	79140
Oklahoma	Christian	829.2	Family Roles	665.2	God	655.1	16996
Oregon	Pronouns	966.8	Identity Politics	685.4	Liberal Democrat	600.7	29488
Pennsylvania	Healthcare	238.5	Higher Education	224.4	Places	160.9	72315
Rhode Island	Places	174.9	Academia	137.2	Teacher	76.5	6110
South Carolina	Social Media	1557.0	God	1198.5	Christian	914.9	27652
South Dakota	Father	119.5	Places	87.1	Family Roles	52.4	2808
Tennessee	Christian	2219.1	God	2030.9	Family Roles	845.4	46845
Texas	God	2271.1	Christian	848.4	Higher Education	747.5	168586
Utah	Father	200.8	Higher Education	125.6	Outdoor	90.0	12878
Vermont	Outdoor	156.2	Community	123.4	Sustainability	108.5	3436
Virginia	Military Police	919.3	Places	816.5	Teacher	563.9	53400
Washington	Technology	2373.4	Pronouns	1141.2	Identity Politics	282.8	39645
West Virginia	Higher Education	309.8	Family Roles	231.3	Mother	210.1	7517
Wisconsin	Sports Teams	380.1	Fan	190.0	Father	138.6	26550
Wyoming	Outdoor	156.8	Places	23.3	Conservative Rep.	19.9	1608

H Clusters most associated with politics

This Appendix shows the numerical results behind our analysis of which clusters are most associated with either Liberal/Democrat or Conservative/Republican bios (see Figure 5 in the main paper). Table H2 shows the clusters that are used most disproportionately by Liberal/Democrat users, and Table H3 does the same for Conservative/Republican users.

Table H2: Identity-relevant clusters that are mentioned proportionally the most by Liberal/Democrat users, compared to Conservative/Republican ones. First columns show the percentage of users in each group who mention the cluster. Last column shows the ratio between the two.

	Lib./Dem.	Cons./Rep.	Ratio L to C
Pronouns	6.5%	0.6%	10.60
Identity Politics	31.9%	5.7%	5.62
Science	4.3%	1.6%	2.72
Nature Emojis	7.9%	3.0%	2.61
Sustainability	5.4%	2.6%	2.07
Animals	11.1%	5.5%	2.00
Enthusiast	9.3%	4.8%	1.93
Academia	5.4%	2.8%	1.92
Spiritual	8.3%	4.3%	1.91
Artist	15.4%	8.2%	1.89
Teacher	7.1%	3.7%	1.89
Disease	11.2%	6.1%	1.83
Hobby Emojis	25.2%	14.8%	1.70
Healthcare	7.3%	4.4%	1.66
Alum	2.4%	1.5%	1.66

Table H3: Identity-relevant clusters that are mentioned proportionally the most by Conservative/Republican users, compared to Liberal/Democrat ones. First columns show the percentage of users in each group who mention the cluster. Last column shows the ratio between the two.

	Lib./Dem.	Cons./Rep.	Ratio C to L
God	3.1%	19.2%	6.27
Family	2.1%	8.4%	3.99
Christian	3.9%	11.9%	3.10
Us And World	3.8%	8.6%	2.25
Military Police	14.5%	29.3%	2.02
Father	6.9%	13.9%	2.01
Real Estate	1.2%	2.3%	1.97
American Emojis	10.4%	19.8%	1.91
Origins	6.8%	11.6%	1.70
Sports Disciplines	3.3%	5.2%	1.57
Industry	4.6%	7.0%	1.54
Professional	1.8%	2.8%	1.53
Entrepreneur	3.4%	4.7%	1.39
Business Services	2.5%	3.4%	1.34
Family Roles	19.6%	25.9%	1.32

References

- M. Dredze, M. J. Paul, S. Bergsma, and H. Tran. Carmen: A Twitter geolocation system with applications to public health. In *Workshops at the twenty-seventh AAAI conference on artificial intelligence*, 2013.
- A. H.-Y. Lee, Y. Lelkes, C. B. Hawkins, and A. G. Theodoridis. Negative partisanship is not more prevalent than positive partisanship. *Nature human behaviour*, 6(7):951–963, 2022.
- Pew Research Center. America is exceptional in its political divide, 2021. URL <https://www.pewtrusts.org/en/trust/archive/winter-2021/america-is-exceptional-in-its-political-divide>. Accessed June 2023.