

1 Tag distribution and dataset statistics

The Reddit activity captured in the Politosphere dataset shows a steady intensification of political engagement over time, with increasing numbers of users, comments, and politically oriented subreddits throughout 2013–2017. This growth is punctuated by clear surges during U.S. election years—most notably in 2016, when activity peaks in correspondence with the presidential campaign (Fig. 1a). The creation and removal of subreddits by tag further indicate that many Democratic- and Conservative-aligned communities emerge around these periods, but often disappear shortly afterward. At the same time, the number of “Banned” groups rises sharply from 2016 onward, many of which vanish in subsequent years (Fig. 1b).

Although Reddit remains predominantly U.S.-based (over 50% of total traffic according to SimilarWeb [1]), European participation displays a gradual upward trend, contributing to a progressively more diversified user base. Together, these temporal dynamics define the empirical foundation for the structural and semantic analyses developed in the following sections.

This dataset constitutes the basis for all subsequent analyses, comprising 120,429,663 comments from 1,889,317 users across 498 politically oriented subreddits (January 2013–December 2017), obtained from the Reddit Politosphere dataset [2], together with 16,778,792 original submissions from the Pushshift Reddit dataset [3], covering 208,316 unique domains.

In our analyses, we rely on a core set of metadata fields. For comments, these include the comment, submission, and parent identifiers, scores, and subreddit-level information such as ban status and inferred political alignment. For submissions, we retain the text, submission ID, user and subreddit identifiers, scores, and domains, which enable us to examine both the linguistic content and the interaction networks underlying political discourse on Reddit.

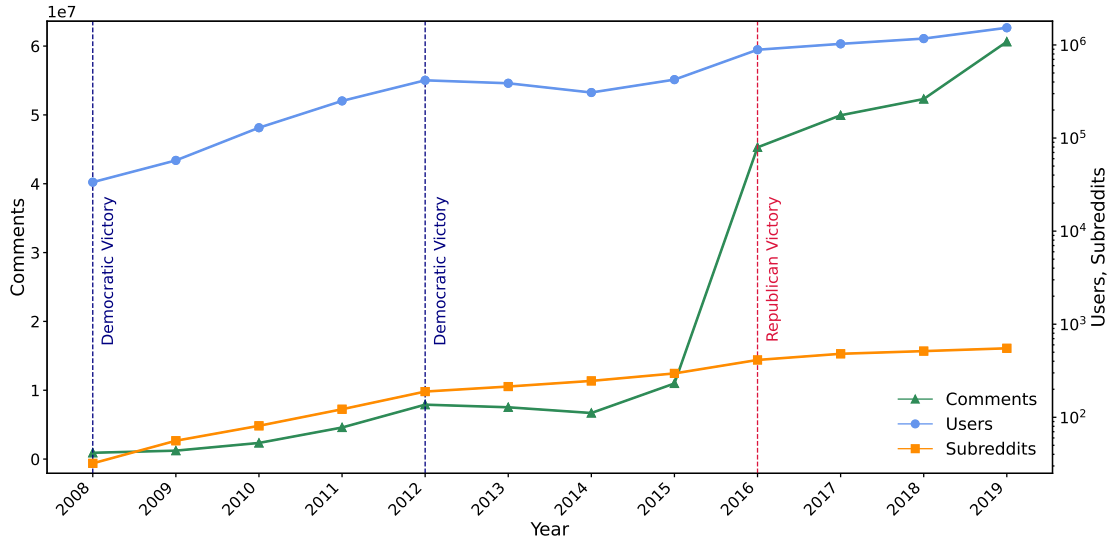
A subsequent key step involves defining a concise yet informative tagging scheme for subreddit communities, extending subreddit-level metadata — including inferred political alignment — with a set of manually curated topic tags covering 17 categories. Public subreddit descriptions were also used to assist in the manual labeling of communities into topic tags.

We also prepared the textual material for linguistic analysis, in order to make structural (interaction-based) and semantic (content-based) components directly comparable within a unified analytical framework. Figure 2 summarizes this methodological workflow—from the raw conversational structure of Reddit data (panel a), to the extraction of descriptive and textual content used for tag assignment and language processing (panel b), to qualitative examples of subreddits with multiple tags and pairwise textual similarity among them (panel c). This process provides the foundation for both network-level and semantic analyses discussed in later sections.

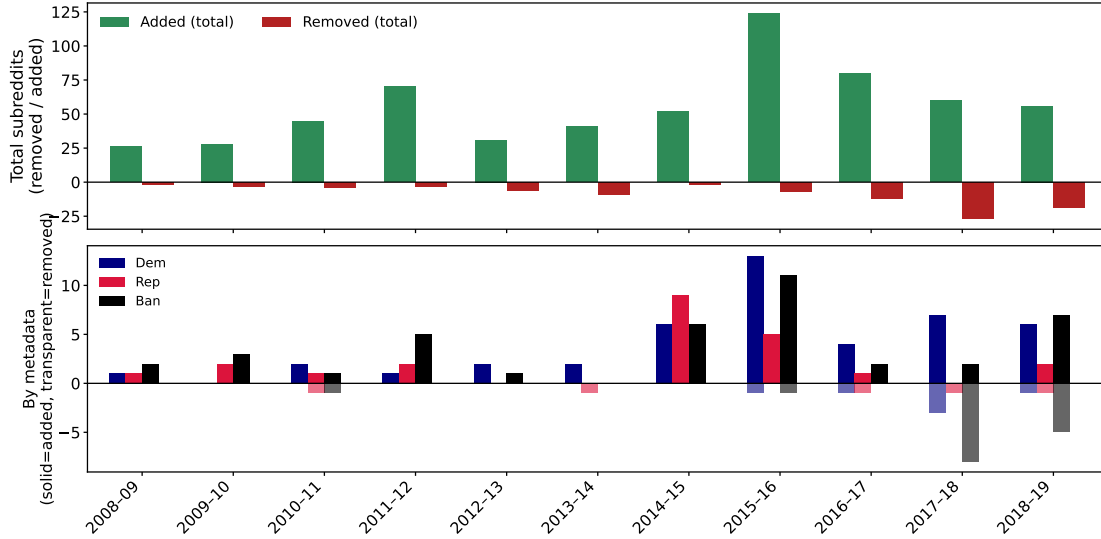
For clarity of presentation, detected network communities are assigned names and colors that reflect the distribution of tags among their constituent subreddits. Rather than adopting only the most frequent tag, we construct composite labels: the dominant tag is reported first, and additional tags are included whenever their frequency reaches at least half that of the preceding one. In this way, a community can receive a compound name such as “Far-Left/Dem” when both affiliations are strongly represented. The generic tag “Politics” is not allowed as a primary label, as it represents an umbrella category without specific partisan meaning. Each tag is also associated with a base color, and the color of a community is then obtained as the average of the colors of the tags appearing in its label. This procedure ensures that community visualizations capture both the dominant and secondary topical affiliations of the underlying subreddits. The resulting tagging scheme—an extension of the original Politosphere labels—is detailed in Table 1, and the complete list of tagged subreddits is reported in Table 2.

Table 1. Color scheme used throughout the figures to represent subreddit categories.

Color	Abbreviation	Full name
	Geop	Geopolitics
	News	News
	Far-Left	Far Left
	Econ	Economics
	PolTalk	Political Talk
	Politic	Politics (general)
	Dem	Democrats
	Lib	Liberals
	Cons	Conservatives
	Gun	Gun Rights
	Far-Right	Far Right
	Ban	Banned
	SocialJustice	Social Justice
	UK	United Kingdom
	Eur	Europe
	Canada	Canada
	Austr	Australia



(a) Annual activity in the Politosphere dataset, reporting the number of comments, active users, and political subreddits.



(b) Subreddit turnover by year, with additions and removals broken down by political tag.

Figure 1. Side-by-side comparison of annual activity (a) and subreddit turnover (b).

Table 2. Subreddits and associated tags.

Subreddit	Tags	Subreddit	Tags	Subreddit	Tags
2012Elections	Dem, Cons	2016Elections	Dem, Cons	2016.elections	Dem, Cons
2ALiberals	Lib, Gun	AOC	Politic, Dem	Abortiondebate	SocialJustice
ActiveMeasures	Geop, News	AgainstHateSubreddits	SocialJustice	AgainstTheChimpire	Ban
Agorism	Econ	Albertapolitics	Canada	AlexandriaOcasio	Dem
AltRightChristian	Ban	AmalaNetwork	Politic	AmericanPolitics	Politic
AnCap101	Far-Right	Anarchism	Far-Left	AnarchismOnline	Far-Left
AnarchistNews	News, Far-Left	Anarcho.Capitalism	Far-Left	Anarchy101	Far-Left
AntiSemitismInReddit	Far-Right	AntiTrumpAlliance	Dem	AnticommieCringe	Far-Right
AntifascistsofReddit	Far-Left	AnybodyButHillary	Dem	AnythingGoesNews	News
AprogressiveParty	Politic, SocialJustice	ArabIsraeliConflict	Geop	ArrestedCanadaBillC16	Canada
AsAGunOwner	Gun	AskALiberal	Lib	AskBernieSupporters	Dem
AskConservatives	Cons	AskDemocrats	Dem	AskEconomics	Econ
AskFeminists	SocialJustice	AskFemmeThoughts	SocialJustice	AskLibertarians	Lib
AskSocialScience	Politic	AskThe_Donald	Cons	AskTrumpSupporters	Cons
Ask_Politics	Politic	Ask_TheDonald	Cons	Askpolitics	Politic

Subreddit	Tags	Subreddit	Tags	Subreddit	Tags
AusPol	Austr	AustraliaLeftPolitics	Austr	AustralianPolitics	Austr
AutoNewspaper	News	BCPolitics	Politic	BDS	Geop
BadSocialScience	Politic	BannedFromThe_Donald	Cons	BasicIncome	Politic
BeardTube	PolTalk, Politic	BerlinTruckAttack	News	BernieSanders	Dem
BernieSandersSucks	Cons	Beto2020	Dem	BitcoinDiscussion	Econ, Politic
BlueMidterm2018	Dem	BreakingNews24hr	News	BritishPolitics	UK
CAVDEF	Politic	CNNmemes	News, Politic	COMPLETEANARCHY	Far-Left
California_Politics	Politic	CanadaPolitics	Canada	CanadianPolitics	Canada
Capitalism	Far-Left	CapitalismVSocialism	Far-Left	Cascadia	Politic
CatholicPolitics	Politic	China_Debate	Geop	Classical_Liberals	Lib
ColoradoPolitics	Politic	Communalists	Far-Left	CommunismWorldwide	Far-Left
Conservative	Cons	ConservativeLounge	Cons	ConservativeMeta	Cons
ConservativeNewsWeb	News, Cons	ConservativesOnly	Cons	Conservatives_R_Us	Cons
DNCleaks	News	DankLeft	Far-Left	DarkEnlightenment	Far-Right
Dave_Rubin	PolTalk	DeathtoAmeriKKKa	Far-Right	DebateAltRight	Ban
DebateAnarchism	Far-Left	DebateCommunism	Far-Left	DebateFascism	Ban
DebateaCommunist	Far-Left	DemocraticSocialism	Far-Left	Democrats2020	Dem
DescentIntoTyranny	Politic	Digital_Manipulation	Politic	DitchMitch	Dem, SocialJustice
DonaldTrumpWhiteHouse	Cons	Donald_Trump	Cons	DrainTheSwamp	Cons
Drumpf	Politic	ENLIGHTENEDCENTRISM	Politic	EarthStrike	SocialJustice
EcoInternet	SocialJustice	Economics	Econ	EducatingLiberals	Lib
Egalitarianism	SocialJustice	ElizabethWarren	Dem	EmergingRisks	News, Econ
EndFPTP	Politic	EndlessWar	Politic	EnoughCapitalistSpam	Far-Left
EnoughCommieSpam	Far-Right	EnoughIDWspam	News, Politic	EnoughLibertarianSpam	Lib
EnoughObamaSpam	Cons	EnoughPaulSpam	Dem	EnoughTrumpSpam	Dem
Enough_AOC_Spam	Dem	Enough_Sanders_Spam	Cons	EuropeMeta	Eur
EuropeanFederalists	Eur	EuropeanSocialists	Eur	ExplainBothSides	Politic
FLgovernment	Politic	FULLCOMMUNISM	Far-Left	FULLDISCOURSE	Far-Left, Politic
FakeProgressives	Far-Left	FeMRADebates	SocialJustice	FemraMeta	SocialJustice
FoxFiction	News, Politic	Foxhidesinfo	News, Politic	FreeEuropeNews	Eur
FreePolDiscussion	PolTalk, Politic	FreeSpeech	News	FreeSpeechWorld	News, Ban
Freethought	News	FriendsofthePod	Dem	Fuckthealtright	Far-Left, Dem
Full_news	News	GAPol	Politic	GBPolitics	UK
GUARDIANauto	News	GaryJohnson	Lib	GenZedong	Far-Left
GeneralStrikeUSA	SocialJustice	GoldandBlack	Far-Left	Government_is_lame	Far-Left
GrassrootsSelect	Dem	GreenAndPleasant	Far-Left, UK	GreenNewDeal	SocialJustice
GreenParty	SocialJustice	GunsAreCool	Gun	HBDstats	Ban
HanAssholeSolo	Far-Right	HateSubsInAction	SocialJustice	HeadlineCorrections	Politic
Hillary	Dem	HillaryForAmerica	Dem	HillaryForPrison	Cons
HongKongProtest	Geop, News	IAMALiberalFeminist	SocialJustice	IDontLikeRPolitics	Ban
IWG	Politic	IWW	SocialJustice	Identitarians	Ban, SocialJustice
ImABlue	Ban	Impeach_Trump	Dem	ImpeachmentWatch	Politic
IndoPakDialogue	Geop, Politic	IntellectualDarkWeb	Politic	Intelligence	Politic
IntelligenceNews	News	InternationalNews	News	IronFrontUSA	Far-Left, SocialJustice
IslamUnveiled	Ban	Israel	Geop	IsraelPalestine	Geop
IsraelSubredditWatch	Geop	Israel_Palestine	Geop	JamesDamore	SocialJustice
JoeBiden	Dem	JordanPeterson	Politic	Jreg	Politic
Kamala	Dem	KasichForPresident	Cons	KeepOurNetFree	Politic
Keep_Track	Politic	KochWatch	Econ	Kossacks_for_Sanders	Dem
Labour	News, UK	LabourUK	News, UK	LateStageCommunism	Far-Left, Politic
LateStageImperialism	Geop, Far-Left	LateStageSocialism	Far-Left	Le_Pen	Far-Right, Eur
LeftWingMaleAdvocates	Politic	LeftWithoutEdge	Far-Left	LeftieZ	Far-Left
LeftistHotTakes	Far-Left	Lessig2016	Dem	LevantineWar	Geop
LibDem	Dem, Lib	Liberal	Lib	Liberalist	Lib
LiberalsvsNazis	Lib	Libertarian	Lib	LibertarianDebates	Lib
LibertarianFreeState	Lib	LibertarianLeft	Far-Left, Lib	LibertarianPartyUSA	Politic, Lib
LibertarianUncensored	Lib	LouderWithCrowder	PolTalk	MENACConflicts	Geop
MarchAgainstNazis	Far-Left, SocialJustice	MarchAgainstTrump	Dem	MarchForNetNeutrality	Politic
MarchForScience	Politic	Marco_Rubio	Cons	MarketAnarchism	Far-Left
Marxism	Far-Left	MassachusettsPolitics	Politic	MedicareForAll	SocialJustice
MensRightsMeta	SocialJustice	MetaRepublican	Cons	MideastPeace	Geop, Politic
MissouriPolitics	Politic	MobilizedMinds	Politic	ModernPropaganda	News, Politic
MoreTankieChapo	Far-Left, Ban	Mueller	Geop	MurderedByAOC	Dem
NATOrussianconflict	Geop	NOWTTYG	Gun	NSALeaks	News
NationalSocialism	Far-Right, Ban	NegaRedditRedux	Politic	NeutralPolitics	Politic
NeutralTalk	Politic	NeverTrump	Dem	NewJerseyuncensored	Politic
NewPatriotism	Politic, Far-Right	New_Jersey_Politics	Politic	NewsWhatever	News
Newsy	News	NoFilterNews	News	NoNetNeutrality	Politic
NonAustrianEconomics	Econ	NorthKoreaNews	Geop, News	Objectivism	Politic
Occupy	News, Far-Left, SocialJustice	OntarioPolitics	Canada	OperationPullRyan	Cons
Oregon_Politics	Politic	OurPresident	Dem	Our_Politics	Politic
POLITIC	Politic	POTUSWatch	Politic	Palestine	Geop
PalestineCircleJerk	Geop	PalestineIntifada	Geop	PanicHistory	Politic
PeoplesPartyofCanada	Canada	Pete_Buttigieg	Dem	Physical_Removal	Far-Right, Ban
Polcompball	Politic	Policy2011	UK	PoliticalCompass	Politic
PoliticalCompassMemes	Politic	PoliticalDiscussion	Politic, SocialJustice	PoliticalHorrorStory	Politic
PoliticalHumor	Politic	PoliticalHumour	Politic	PoliticalMemes	Politic
PoliticalOpinions	Econ, Politic	PoliticalPhilosophy	Politic	PoliticalVideo	Politic
Political_Revolution	Far-Left, Dem, SocialJustice	Political_Tumor	Politic	Political_Tweets	Politic

Subreddit	Tags	Subreddit	Tags	Subreddit	Tags
Postleftanarchism	Far-Left	PragerUrine	Lib	Pragmatism	Politic
PraxAcceptance	Politic, Lib, Ban	PresidentWarren	Dem	PresidentialRaceMemes	Politic
PublicLands	News, Politic	Pyongyang	Geop	QualitySocialism	Far-Left
RedEnsign	UK, Canada	ReddLineNews	News	RedditCensors	Politic
RedsKilledTrillions	Far-Right	Reform_The_DNC	Dem	Republican	Cons
RepublicanValues	Cons	RepublicansForSanders	Cons	Right_Wing_Politics	Cons, Ban
RightwingLGBT	Far-Right, Ban, SocialJustice	RiseUPP	SocialJustice	Romney	Cons
RsocialismMeta	Far-Left	RussiaLago	Geop	SRSDiscussion	SocialJustice
SandersForPresident	Dem	SargonofAkkad	Far-Right, UK	SelfAwarewolves	Politic
SethRich	Ban	ShitLiberalsSay	Far-Left	ShitNeoconsSay	Ban
ShitPoliticsSays	Politic	ShitPoppinKreamSays	Politic	ShitRConservativeSays	Dem
ShitThe_DonaldSays	Dem	Shitstatistssay	Politic	ShittyDebateCommunism	Politic
Sino	News, Far-Left	SmugIdeologyMan	Politic	SocialDemocracy	Politic
SocialismVCapitalism	Far-Left	Socialism_101	Far-Left	Sorosforprison	Ban
SpeechFree	News	StillSandersForPres	Dem	StormfrontorSJW	Politic
SyndiesUnited	SocialJustice	SyrianRebels	Geop	TYT	PolTalk
TedCruzForPresident	Cons	TennesseePolitics	Politic	TexasPolitics	Politic
ThanksObama	Dem	The3rdPosition	Far-Right, Ban	TheColorIsBlue	News, Politic
TheColorIsRed	News, Politic	TheLeftCantMeme	Cons	TheMajorityReport	News
TheMotte	Politic	TheNewRight	Far-Right, Ban	TheNewsFeed	News
TheRecordCorrected	News	The_Cabal	Ban	The_Congress	Politic
The_Donald	Cons, Ban	The_Donald_CA	Cons	The_Europe	Ban, Eur
The_Farage	Far-Right, UK	The_Leftorium	Far-Left	The_Mueller	Politic
The_MuellerMeltdown	Politic	ThisButUnironically	Politic	ThreeArrows	Far-Left
TiADiscussion	News, Politic	TimCanova	Dem	TommyRobinson	Far-Right, UK
TraaButNoCommies	Far-Right, SocialJustice	TrueCatholicPolitics	Politic	TruePoliticalHumor	Politic
TrueReddit	News, Politic	TrueTrueReddit	News, Politic	True_AskAConservative	Cons
TrumpCriticizesTrump	Cons	TrumpForPrison	Dem	Trump_Watch	Cons
Trumpgret	Cons	Trumpgrets	Cons	UK_Politics	UK
UMukhasimAutoNews	News	USNEWS	News	UkrainianConflict	Geop
Ultraleft	Far-Left	UnbiasedCanada	Canada	VaushVidya	Far-Left, SocialJustice
VirginiaPolitics	Politic	VoteBlue	Dem	Vote_Trump	Cons
WatchRedditDie	Politic	WayOfTheAloha	Dem	WayOfTheBern	Dem
WhatsMyIdeology	Politic	WhereAreTheChildren	SocialJustice	WhereIsAssange	SocialJustice, Austr
WhiteRights	Far-Right, Ban	WikiInAction	Politic	WikiLeaks	News
YangForPresident	Dem	YangForPresidentHQ	Dem	YangGang	Dem
YemeniCrisis	Geop	Zionism	Geop	abetterworldnews	Geop, News
accidentallycommunist	Far-Left	acteuropa	Eur	actualconspiracies	Politic
actualliberalgunowner	Lib, Gun	agitation	SocialJustice	alltheleft	Far-Left
altnewz	News	altright	Far-Right, Ban	anarcho_primitivism	Far-Left
anarchocommunism	Far-Left	anarchomemes	Far-Left	anarchy	Far-Left
antifa	Far-Left, Ban	antifapassdenied	Far-Left	antiwar	SocialJustice
arizonapolitics	Politic	askaconservative	Cons	askhillarysupporters	Dem
atheismplus	Politic	austrian_economics	Econ	badeconomics	Econ
badgovnofreedom	Politic	badpolitics	Politic	benshapiro	Cons
bernie	Dem	bernieblindness	Dem	besteuropa	Eur
betternews	News	brealism	Far-Left, SocialJustice	brexit	UK
calexit	Politic	canadaleft	Far-Left, Canada	capitalism_in_decay	Far-Left
censorship	Politic, SocialJustice	centerleftpolitics	Far-Left	centrist	Politic
chinareddits	Geop	chomsky	Far-Left, SocialJustice	climateskeptics	Politic
communism	Far-Left	communism101	Far-Left	communists	Far-Left
conservativecartoons	Cons	conservatives	Cons	conspiracyfact	Politic
conspiratoracy	Politic	craftofintelligence	News	cyberlaws	Politic
daverubin	PolTalk, Cons	debateAMR	SocialJustice	debatepoliticalphil	Politic
democraticparty	Dem	democrats	Dem	demsocialist	Far-Left
demsocialists	Far-Left	dirtbagcenter	Politic	distributism	Econ
donaldrump	Cons, Ban	dsa	Far-Left	ecomonitor	Econ
economy	Econ	enoughpetersonspam	UK	enoughsandersspam	Dem
esist	News	europeannationalism	Far-Right, Ban, Eur	europeans	Eur
europeanunion	Eur	europees	Eur	evolutionReddit	Politic
exlibertarian	Lib	fascist	Far-Right, Ban	feminismformen	SocialJustice
fivethirtyeight	Politic	foreignpolicy	Geop	fullstalinism	Far-Left
futuristparty	Politic	geopolitics	Geop	georgism	Politic
globalistshills	Politic	googogahgah	Politic	government	Politic
gravelforpresident	Dem	greed	Econ, Politic	gue	Politic
gulag	Far-Left	guncontrol	Gun	gunpolitics	Gun
hillaryclinton	Dem	historicalrage	Politic	holocaust	Ban
illinoispolitics	Politic	indianmuslims	Geop	inslee2020	Dem
inthemorning	PolTalk	inthenews	News	iranpolitics	Geop
irishpolitics	Geop, Eur	irredeemables	Cons, Ban	israelexposed	Geop
jillstein	Dem, SocialJustice	jimmydore	PolTalk, Dem	justicedemocrats	Dem
killthosewhodisagree	Politic	labor	Econ, SocialJustice	law	Politic
leftcommunism	Far-Left	liberalgunowners	Lib, Gun	libertarianaustralia	Lib, Austr
libertarianmeme	Lib	libtard	Ban	marxism_101	Far-Left
media_criticism	Politic	metaNL	Lib	metanarchism	Far-Left
militant	Far-Left	mnt.economics	Econ	moderatepolitics	Politic
monarchism	Politic	mormonpolitics	Politic	mutualism	Far-Left
ndp	Dem, Canada	neoconNWO	Cons	neoliberal	Lib
neoprogs	SocialJustice	neutralnews	News	nevadapolitics	Politic
neveragainmovement	Gun	new_right	Far-Right, Ban	nonmorons	Politic

Subreddit	Tags	Subreddit	Tags	Subreddit	Tags
nra	Gun	nrnx	Politic	nyspolitics	Politic
obama	Dem	obamacare	Dem	occupywallstreet	Far-Left
onguardforthee	Canada	overpopulation	Politic	pol	Politic, Ban
politicalcartoons	Politic	politicalfactchecking	News, Politic	politicalhinduism	Geop
politics	Politic	politicsdebate	Politic	postnationalist	SocialJustice
prochoice	SocialJustice	progressive	SocialJustice	progun	Gun
prolife	Cons, SocialJustice	propaganda	News	qualitynews	News
race	Ban	randpaul	Cons	realworldpolitics	News
redacted	News	republicans	Cons	restorethefourth	Far-Left, SocialJustice
revolution	Far-Left	rojava	Geop	ronpaul	Lib, Cons
samharris	Dem	scotus	Politic	secondamendment	Gun
seculartalk	PolTalk	shitfascistssay	Far-Left	shitguncontrollerssay	Gun
shitleftistsay	Cons	shitneoliberalismsays	Far-Left	shittankiessay	Far-Left
slatestarcodex	Politic	slatestarcodex_cw	Politic	smuggies	Ban
socialanarchism	Far-Left	socialism	Far-Left	socialjustice101	Far-Left, SocialJustice
stevencrowder	PolTalk, Cons	stopadvertising	News	stupidpol	Far-Left
syriancivilwar	Geop	syrianconflict	Geop	taxmarch	SocialJustice
terrorism	Politic	the_meltdown	Politic	thedavidpakmanshow	PolTalk
thenewcoldwar	Politic	thenewsrighnow	News	theredpillright	Far-Right
thomasjefferson	Politic	tories	Cons, UK	trollfare	Geop
trump	Cons	trump16	Cons	trumptweets	Cons
tuesday	Cons	tulsi	Dem	tytonreddit	PolTalk
ukipparty	UK	ukpolitics	UK	ukright	Far-Right, UK
uncensorednews	News, Ban	unfilter	News, Politic	union	Far-Left, SocialJustice
usnews	News	uspolitics	Politic	venezuelancivilwar	Geop
wakinguppodcast	PolTalk	walkaway	Cons	wexit	Canada
willis7737_news	News	worldevents	Geop, News	worldpolitics	Geop, News
worldtoday	Geop, News	yimby	Politic		

2 Effectiveness and advantages of statistical validation

Our approach employs the bipartite configuration model (BiCM; see Methods) to filter out random co-occurrences and correct for degree bias before projection onto the subreddit layer. As detailed in the main text, the validated projection is obtained after applying the Revealed Comparative Advantage (RCA) filter, retaining only links with $RCA \geq 1$ [4]. For each projected link, p -values are computed analytically using a Poisson approximation of the bipartite connection probability [5, 6]. Statistical significance is then assessed using a threshold $\alpha = 10^{-4}$, with multiple-hypothesis correction through the False Discovery Rate (FDR) procedure [7].

Figure 3 illustrates the effect of validation by comparing bipartite degree with projected degree in 2016. Without validation, degree correlations remain near-perfect and the adjacency matrix (ordered by degree) exhibits a nested, uninformative structure. After validation, the trivial correlation is broken, yielding a structure with clear community separation. A visual comparison of the two adjacency matrices and the corresponding partitions is shown in Fig. 4.

For 2016, we examined the variation of modularity Q and density δ under different filtering strategies. As shown in Fig. 5, validated projections obtained with the BiCM display a smooth, monotonic increase in Q and a corresponding decrease in δ as the significance threshold α becomes more stringent. Importantly, the resulting community structure remains stable and informative even at relatively low values of α .

By contrast, in the unvalidated case—illustrated in Fig. 6 for the user–subreddit projection—the signal is much weaker. Modularity never reaches values comparable to the validated networks, and reducing density requires discarding large portions of the network. Even at the modularity peak ($C = 6.50$), the resulting structure remains unstable and fails to provide coherent communities.

The community partitions resulting from these procedures are compared in Fig. 7. Validated networks yield stable partitions across thresholds, whereas in the unvalidated case, even at the threshold corresponding to the modularity peak ($C = 6.50$), the detected clustering differs substantially from that observed at lower cutoffs. For consistency, we matched validated and unvalidated cases at the same network size, using $\alpha = 10^{-2}, 10^{-4}, 10^{-6}, 10^{-8}$, corresponding to $C = 1.65, 2.57, 2.80, 3.15$.

Finally, to assess partition stability, we generated pairs of projections (validated vs. non-validated) with the same number of subreddits and compared their Louvain partitions using information–theoretic measures [8–11]. Given a partition C with K clusters, the Shannon entropy is defined as:

$$S(C) = - \sum_{\kappa=1}^K P(\kappa) \log P(\kappa), \quad \text{where} \quad P(\kappa) = \frac{n_{\kappa}}{n}. \quad (1)$$

Here, n_{κ} is the size of cluster κ , and n is the total number of elements.

The mutual information (MI) quantifies the similarity between two partitions C and C' of the same data by measuring the

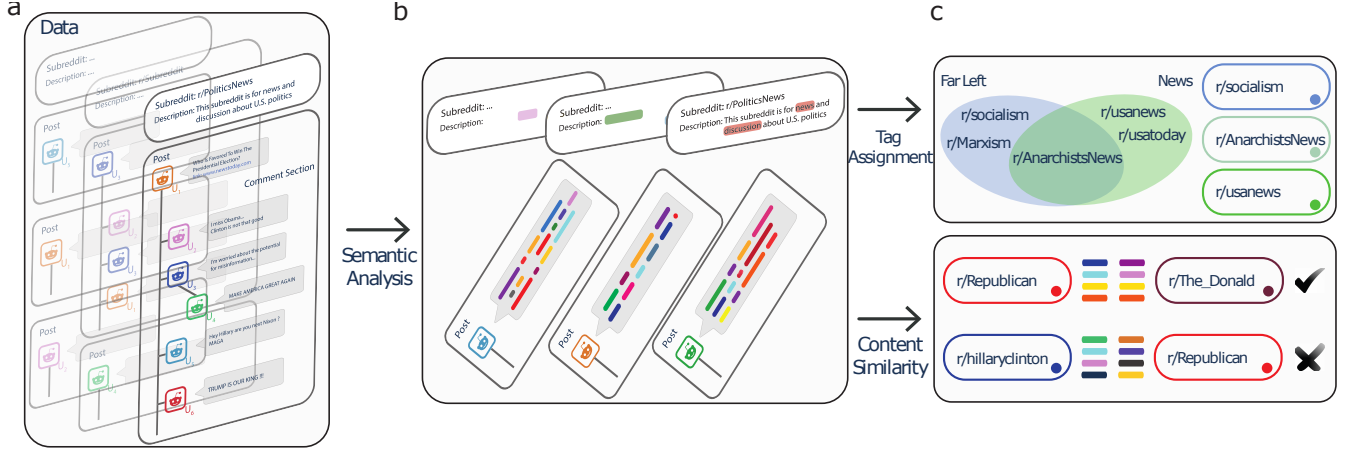


Figure 2. Semantic analysis and tag assignment pipeline. Panel a: Example of Reddit conversation data, including posts and comment threads. Panel b: Separation between subreddit descriptions (used for manual tag assignment) and post texts (used for linguistic analysis). Panel c: Qualitative examples of tag assignment (e.g., “Far Left”, “News”) and content similarity across subreddits based on cosine distance. This figure summarizes how subreddit metadata and text content were jointly processed to extract thematic tags and compute linguistic similarities. The detailed description of the linguistic analysis is provided in Supplementary Section S6.

shared information between them:

$$MI(C, C') = S(C) + S(C') - S(C, C').$$

To normalize MI, we use the Normalized Mutual Information (NMI):

$$NMI(C, C') = \frac{2MI(C, C')}{S(C) + S(C')}. \quad (2)$$

A more refined measure, Adjusted Mutual Information (AMI), accounts for expected MI between random partitions:

$$AMI(C, C') = \frac{MI(C, C') - \mathbb{E}[MI(C, C')]}{\max[S(C), S(C')] - \mathbb{E}[MI(C, C')]} \quad (3)$$

Lastly, we computed the Variation of Information (VI) [11], which quantifies the distance between two partitions:

$$VI(C, C') = S(C) + S(C') - 2MI(C, C'). \quad (4)$$

Since $VI \leq \log(n)$, we normalized the score by $\log(n)$ for comparability across datasets of different sizes.

To ensure comparability across validated and non-validated cases, we grouped in the validated projections all subreddits that did not survive statistical filtering into a single residual cluster. Summarizing, we computed Shannon entropy $S(C)$, Normalized Mutual Information (NMI), Adjusted Mutual Information (AMI), and the Variation of Information (VI, normalized by $\log n$).

Figure 8 shows that validated networks consistently yield high NMI, AMI, and 1-NVI (one minus the normalized Variation of Information) across thresholds, indicating robust and repeatable partitions. In contrast, unvalidated networks produce unstable results, with large fluctuations across thresholds and generally lower similarity between partitions.

3 Community detection algorithm comparison

Communities in our user–subreddit and domain–subreddit networks were first detected by maximizing modularity using the Louvain algorithm [12].

Modularity Q quantifies the quality of a network division into modules, with higher values indicating dense intra-module connections and sparse inter-module links. It is defined as

$$Q = \frac{1}{2m} \sum_{ij} (A_{ij} - \frac{k_i k_j}{2m}) \delta(C_i, C_j),$$

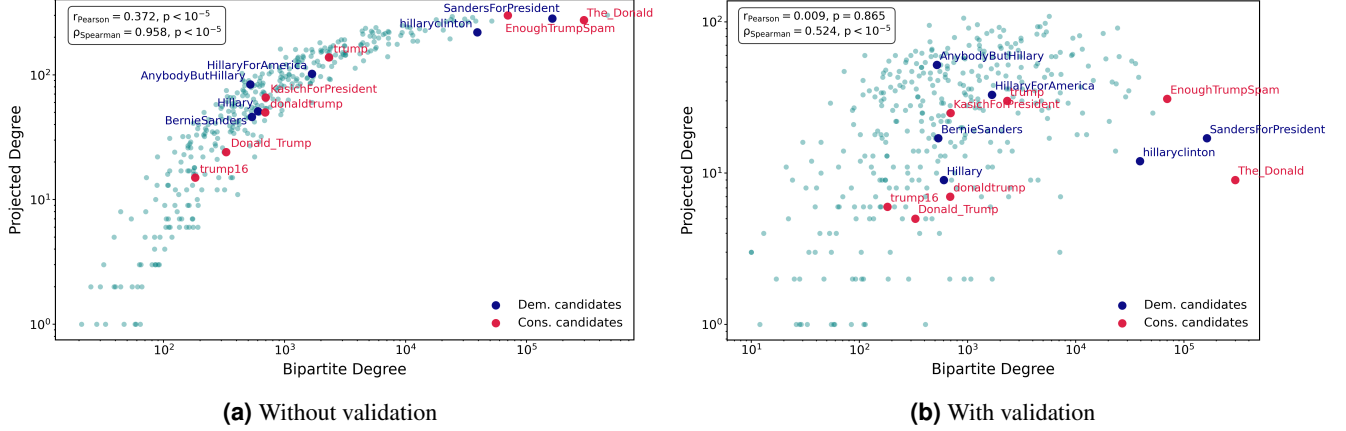


Figure 3. Correlation between subreddit degree in the bipartite network and degree in the projected network (log–log), shown (a) without and (b) with statistical validation. Each point represents a subreddit; highlighted labels mark communities associated with Democratic (blue) and Conservative (red) candidates. Pearson’s r and Spearman’s ρ (with p -values) are reported within panels. Validation reduces spurious links and alters the dependence between bipartite and projected degree, clarifying the role of politically salient subreddits.

where A_{ij} is the weight of the edge between nodes i and j , k_i and k_j are their degrees, m is the total edge weight, and $\delta(C_i, C_j) = 1$ if nodes i and j share the same community, 0 otherwise. The Louvain procedure begins by assigning each node to its own community, then iteratively moves nodes to neighboring communities whenever such moves increase Q . Once no single-node moves can improve modularity, communities are aggregated into “super-nodes” and the process repeats until a global maximum is reached.

To assess the robustness of these partitions, we compared Louvain results with those obtained from the Infomap algorithm [13, 14], which frames community detection as an information-theoretic optimization. Infomap simulates random walks on the network under the assumption that walkers spend more time within communities than between them. It employs the “map equation” to measure the theoretical code length $L(M)$ required to describe a random-walk trajectory given a community partition:

$$L(M) = q_{\sim} S(\mathcal{Q}) + \sum_i p_{\circ}^i S(P^i),$$

where $q_{\sim}$ is the probability of exiting a community, $S(\mathcal{Q})$ is the entropy of inter-community transitions, $p_{\circ}^i$ is the probability of taking steps within community i , and $S(P^i)$ is the entropy of intra-community movements. Infomap iteratively refines the partition to minimize $L(M)$, with the best partition yielding the shortest code length.

For the 2016 data, we generated networks at multiple significance thresholds and a non-validated baseline using a fixed interaction cutoff (see Methods and SI Section S3). We then applied both Louvain and Infomap to each network and quantified agreement via normalized mutual information (NMI), adjusted mutual information (AMI), and variation of information (VI). As shown in Fig. 9, although both algorithms produce highly similar partitions, statistically validated networks achieve consistently higher NMI and AMI (and lower VI) than the non-validated baseline, demonstrating stronger consistency between methods. These improvements are accompanied by a systematic increase in modularity and a corresponding decrease in code length as the validation threshold becomes more stringent, further indicating that statistical filtering sharpens community structure.

Finally, to provide a complementary comparison, we also employed a stochastic block model (SBM) approach [15, 16], which infers hierarchical block structure by optimizing a minimum-description-length objective over possible partitions. This model accommodates nested communities and automatically selects the number of blocks. Figure 9 also reports the comparison between Louvain and SBM partitions: here too, validated networks yield higher concordance between algorithms, demonstrating the stability of our community assignments across different detection paradigms.

4 Polarization and user labels in interaction-based partitions

To analyze polarization in more detail, we constructed polarization matrices that capture how users labeled with a given affiliation populate different subreddit communities. Formally, for each community c' we consider the set of its active users $W_{c'}$,

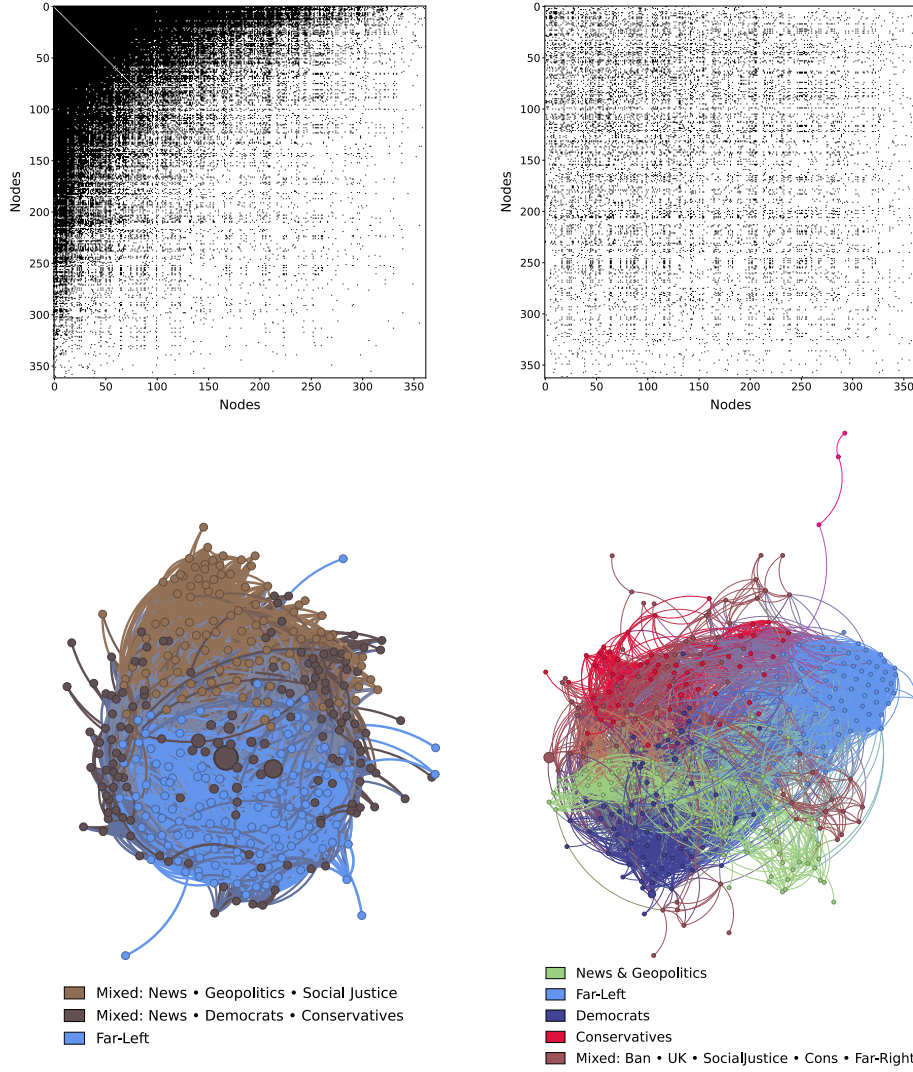


Figure 4. Comparison of subreddit networks in 2016, with the same number of nodes. In the unvalidated case, the adjacency matrix (nodes ordered by bipartite degree) exhibits a nested and uninformative structure, and the resulting partition lacks clear modularity. By contrast, the validated network reveals a block structure in the adjacency matrix and yields clearer, more coherent communities when modularity is maximized.

and for each possible label c we compute the fraction of $W_{c'}$ whose main label is c , dividing by the total number of users in $W_{c'}$ (see Methods). The resulting entry $P_{c'c}$ therefore represents the proportion of users active in community c' that are labeled as c , so that each row of the polarization matrix is normalized to 1. Diagonal entries P_{cc} define the polarization index of community c , quantifying the extent to which its activity is self-focused, with higher values indicating stronger polarization. These matrices provide a general framework to study polarization, which we apply at different levels of aggregation—from tag-based groups to network-detected communities.

(i) Tag communities.

First, we constructed polarization matrices where rows correspond to topic tags assigned to subreddits and columns to tags propagated onto users ($N \times N$; see Methods). Users are labeled *a priori* according to the main tag of the subreddits in which they comment. The resulting matrices are strongly diagonal, indicating clear within-tag alignment. Annual polarization matrices (Fig. 10) and donut charts (Fig. 11) provide direct visualizations of these distributions. Mann–Whitney U tests confirm the statistical significance of this diagonal structure, showing robust polarization within tag-defined communities (Table 3).

(ii) Network communities.

Second, we applied the Louvain algorithm to the user–subreddit interaction networks, yielding communities of subreddits

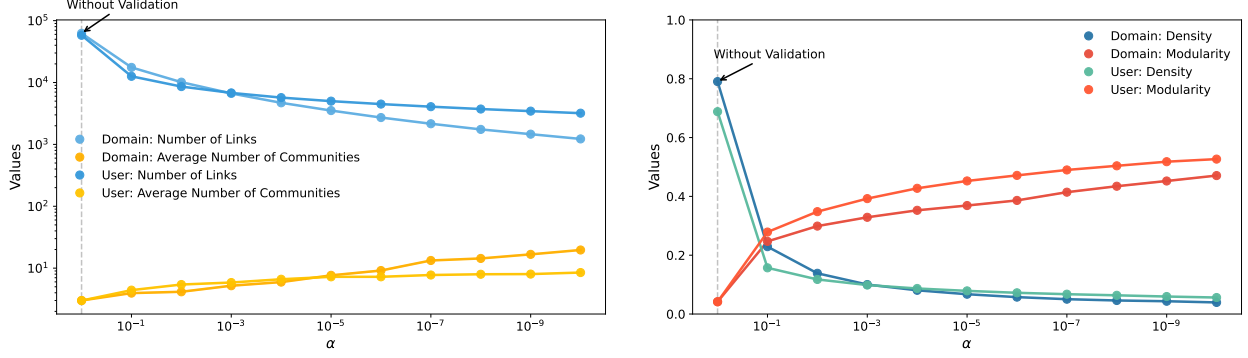


Figure 5. Effect of statistical validation on projected subreddit networks (2016), derived from both user–subreddit and domain–subreddit analyses. As the significance threshold α becomes more stringent, validated projections display a monotonic increase in modularity Q and a corresponding decrease in density δ .

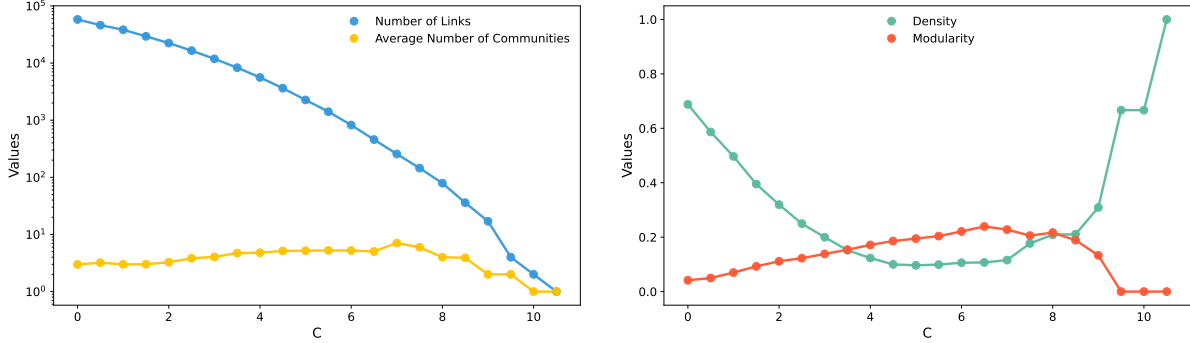


Figure 6. Effect of weight thresholding without validation for the 2016 user–subreddit projection. Modularity Q and density δ fluctuate erratically as the cutoff $C = \log(\text{co-occurrence weight})$ increases, and even at the peak of modularity ($C = 6.50$) the resulting partitions remain unstable.

($C \times C$ matrices). In this case, users are labeled by their assigned network community. The corresponding polarization matrices (Fig. 12) and donut charts (Fig. 13) again reveal highly diagonal structures, validating that network partitions capture strong polarization patterns.

(iii) Community–tag cross analysis.

Finally, we examined the internal makeup of network communities by comparing them with user tags, producing mixed $C \times N$ matrices. These matrices measure how each detected community is populated by users with different propagated tags. Results are summarized in Fig. 14, while normalized Shannon entropies (Table 4, see also Eq. 5) quantify the degree of topical diversity within each community in terms of user-tag composition.

For each community detected in the subreddit network, we computed the entropy of its user-tag distribution. Each user is assigned a tag, and for a given community we consider the empirical distribution $\{p_i\}$ over tags i . The normalized Shannon entropy is defined as

$$H_{\text{norm}} = -\frac{1}{\log_2 N_{\text{tags}}} \sum_{i=1}^{N_{\text{tags}}} p_i \log_2 p_i, \quad (5)$$

where N_{tags} is the total number of distinct tags. By construction, $H_{\text{norm}} \in [0, 1]$: values close to 0 indicate communities dominated by a single tag (low diversity), while values close to 1 correspond to communities where tags are more evenly distributed (high diversity).

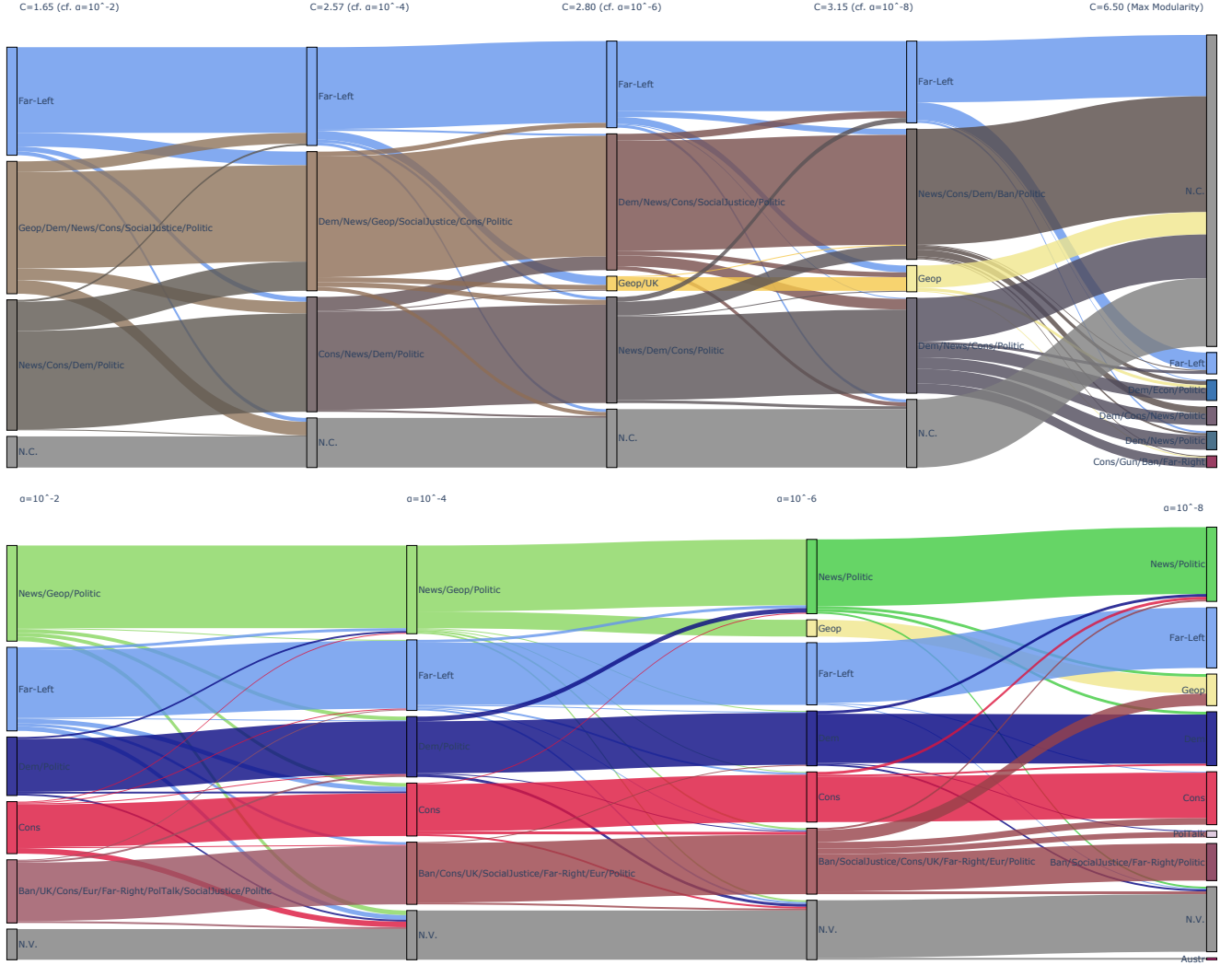


Figure 7. Flowcharts of subreddit communities in 2016. (a) *Unvalidated*: partitions obtained at increasing thresholds C , defined as the logarithm of the co-occurrence weight in the projected network. For the first four cutoffs, the numbers in parentheses indicate the reference significance levels α used in the validated case to enable like-for-like comparison. The last partition corresponds to the modularity peak at $C = 6.50$. (b) *Validated*: partitions obtained at different significance thresholds α . Validated communities are more stable across thresholds and achieve higher modularity than their unvalidated counterparts.

To track how user labels evolve over time, we generated annual Sankey diagrams (Fig. 15), illustrating flows of users between label categories. These flows reveal a marked rise in Democratic-labeled users from 2015 onward, with both Conservative and Banned groups expanding significantly around the 2016 elections. In particular, we observe strong transitions into Banned: in 2015–2016 about 11.9% of Conservatives and 15.6% of Democrats moved into Banned, while in 2016–2017 these percentages rose to 21.9% and 3.9%, respectively. Cross-flows between Democrats and Conservatives are also visible (4.5% and 4.9% in 2015–2016; 8.4% and 1.8% in 2016–2017). The baseline group sizes for Democrats and Conservatives, which serve as reference values for these flows, are reported in Table 5. Multi-label assignment can introduce fluctuations: for instance, a large cohort of Democrats appears in 2015 (58.5k equivalent users), and about 24.4% of those moving into Ban later return to Democrats in 2016–2017.

Together with the entropy scores and flow dynamics, these results highlight the heterogeneity and volatility of political communities on Reddit, with polarization intensifying around election cycles and moderation contributing to the emergence of Banned groups.

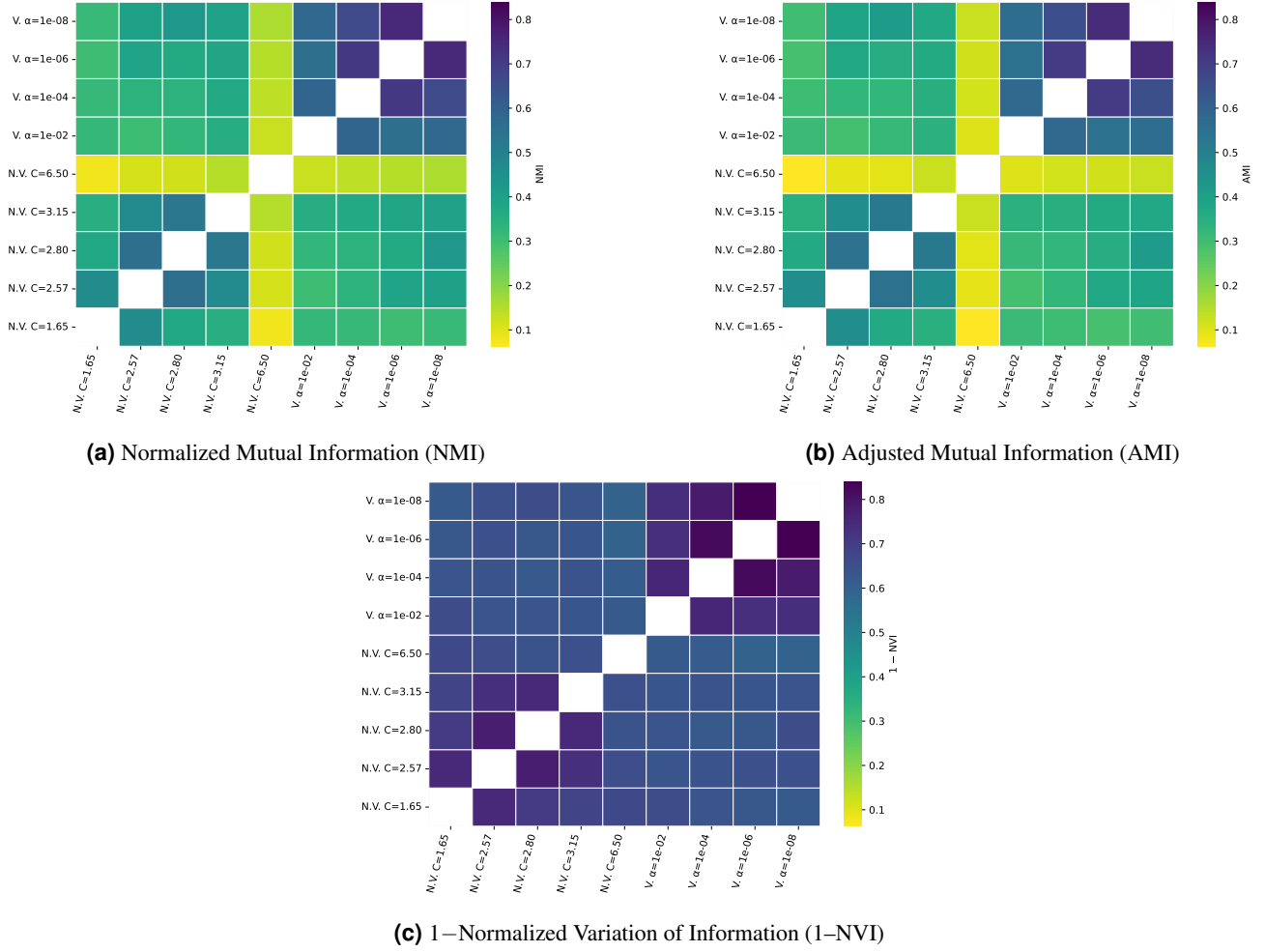


Figure 8. Similarity between community partitions in validated and unvalidated subreddit networks (2016). Each panel reports a heatmap of pairwise similarity across network variants (diagonal masked). Metrics shown: (a) Normalized Mutual Information (NMI), (b) Adjusted Mutual Information (AMI), and (c) 1 - NVI, i.e., one minus the normalized Variation of Information. Validation yields consistently higher similarity, confirming the robustness of community partitions.

Table 4. Entropy of subreddit communities identified in the subreddit networks, computed according to the distribution of user tags populating each community.

Year	Community	Entropy
2013	Geop	0.432
2013	News/SocialJustice/Politic	0.740
2013	Far-Left	0.711
2013	UK	0.043
2013	Cons/Lib/Politic	0.673
2013	SocialJustice/Politic	0.693
2013	Canada	0.071
2014	Geop/News/Politic	0.656
2014	Far-Left	0.700
2014	Lib	0.623
2014	UK/Eur/Politic	0.608
2014	Cons/Gun/Dem/Econ	0.753
2014	SocialJustice	0.569
2014	Ban/Far-Right	0.375
2014	PolTalk	0.031
2014	Canada	0.045
2015	Geop	0.504

Year	Community	Entropy
2015	News/Politic	0.653
2015	Far-Left	0.778
2015	Econ/Lib/Far-Left/Politic	0.599
2015	Cons/Dem	0.669
2015	SocialJustice/Politic	0.738
2015	Ban/UK/Far-Right	0.350
2015	Canada	0.017
2015	Austr	0.028
2016	News/Geop/Politic	0.846
2016	Far-Left	0.796
2016	Dem/Politic	0.692
2016	Cons	0.706
2016	Ban/UK/SocialJustice/Cons/Far-Right/Eur	0.511
2016	Austr	0.043
2017	Geop	0.548
2017	News/Politic	0.856
2017	Far-Left	0.766
2017	Dem/Politic	0.671
2017	Cons/Lib/Politic	0.808
2017	Ban/SocialJustice/Far-Right/Cons	0.507

5 Impact of tag removal on the polarization index

To uncover interdependencies among tag-based communities, we performed a leave-one-tag-out analysis: each tag was removed in turn from the set of possible subreddit labels and we recomputed the polarization index ρ for the remaining tags following the procedure described in **Results**. This approach highlights indirect relationships driven by shared user bases and the mutual reinforcement of community coherence.

For each focal tag t and year y , we quantified sensitivity by sequentially omitting every other tag $s \neq t$ and measuring the resulting deviations in ρ_t . These deviations were standardized as

$$z_{t,y}(s) = \frac{\rho_t^{(-s)}(y) - \mu_t(y)}{\sigma_t(y)},$$

where $\rho_t^{(-s)}(y)$ is the polarization of t after removing s , while $\mu_t(y)$ and $\sigma_t(y)$ are the mean and standard deviation across all such exclusions in year y . Equivalently, the color scale reports how many standard deviations each value deviates from the annual across-tag mean for the focal tag.

In the bar plot of **Figure 3** (main text), we show polarization values for each tag together with shaded bars indicating the effect of excluding “Banned,” benchmarked against the across-exclusion baseline.

Figure 16 expands this view, reporting heatmaps for five focal tags: rows correspond to the excluded tag s , columns to years y , and colors encode $z_{t,y}(s)$ in units of σ . Positive values indicate that removing s increases the focal tag’s polarization relative to its annual mean; negative values indicate a decrease. These patterns reveal how other tags impact the focal tag through overlap and coupling of user populations. Consistent with the main text, omitting “Banned” produces a pronounced drop in the polarization of Conservative and Democratic groups, while excluding “Far Right” substantially reduces the polarization of Banned subreddits, especially in the pre-2016 election period.

6 Textual patterns and shifts in similarity

In this section, we examine how cosine-based textual similarity evolves within communities defined by both topic tags and network-derived interactions. Each subreddit’s textual corpus was aggregated into a single document, combining all posts and comment texts over the observation period. The resulting documents were represented as vector embeddings of terms, built from the vocabulary described in Methods. Pairwise cosine similarity between these vectors provides a measure of linguistic proximity among subreddits.

Similarities were then averaged within tag-based groups and within network-based communities, yielding two complementary perspectives on linguistic cohesion. Figures 17 and 18 report the annual cosine-similarity matrices for 2013–2017, showing pairwise similarities among tag-based and interaction-based communities, respectively. In both cases, diagonal entries (within-community similarity) are systematically higher than off-diagonal values, and their growth over time indicates an

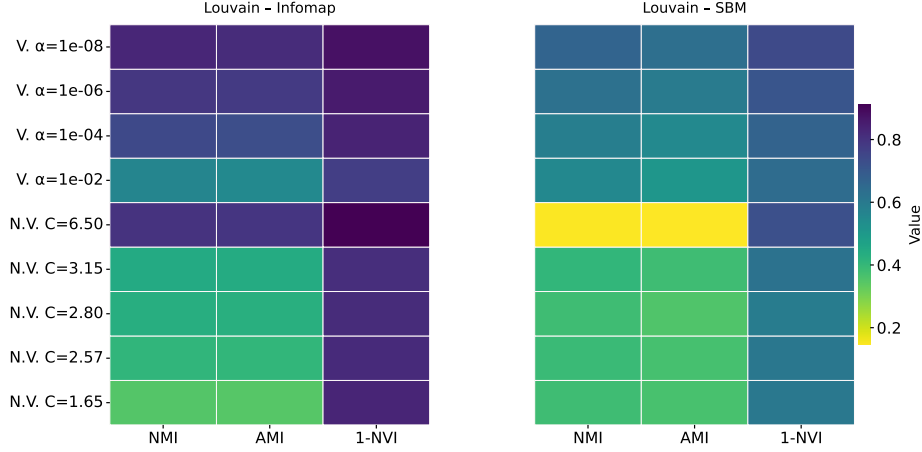


Figure 9. Agreement between community detection algorithms on validated and non-validated networks. The left panel compares Louvain and Infomap partitions, while the right panel compares Louvain and stochastic block model (SBM) partitions. Agreement is quantified using normalized mutual information (NMI), adjusted mutual information (AMI), and variation of information (VI). In both cases, statistically validated networks yield higher NMI, AMI and 1-NVI (and $1 - VI$ (equivalently, lower VI)) than the non-validated baseline, demonstrating stronger consistency across detection paradigms.

Year	Communities p -value	Tags p -value
2013	1.42×10^{-5}	3.12×10^{-10}
2014	5.73×10^{-7}	2.82×10^{-12}
2015	5.73×10^{-7}	2.51×10^{-12}
2016	7.23×10^{-5}	3.96×10^{-12}
2017	6.42×10^{-5}	3.35×10^{-12}

Table 3. Mann–Whitney U test results confirming that within-community similarity (diagonal values) is significantly higher than cross-community similarity. Comparisons are shown both across network and tag communities.

overall increase in textual cohesion within groups. At the same time, cross-community distances also tend to decrease, pointing to a progressive convergence in language use across different subreddit clusters.

To assess the robustness of this signal, we conducted Mann–Whitney U tests comparing each diagonal entry with the distribution of off-diagonal values. The results, summarized in Table 6, confirm that within-community similarity is significantly stronger than cross-community similarity.

We then examined temporal shifts in textual similarity while controlling for the increasing number of subreddits over time. Restricting the analysis to the subset of subreddits active from 2014 through 2016, we observed a significant average increase in cosine similarity for the Far Left, Democratic, News, and Conservative communities. Finally, we validated these distributional changes using two-sample Kolmogorov–Smirnov tests [17], directly comparing each year with its predecessor and contrasting 2014 with 2016. The resulting distributions, illustrated in Fig. 19, confirm a systematic upward shift in textual cohesion over time.

Beyond this general rise in linguistic cohesion, qualitative inspection of the cosine-similarity matrices reveals meaningful

Year	2013	2014	2015	2016	2017
Democrats	3 241.0	3 473.0	58 558.5	130 357.0	126 993.5
Conservatives	12 257.8	12 397.5	12 274.0	52 936.0	46 958.5

Table 5. Weighted number of users in Democratic and Conservative groups across years. These quantities represent the baseline sizes of the two groups that generate the flows in Fig. 15. Values are fractional (non-integers) because multi-label assignment is possible, i.e., the same user can contribute to multiple groups.

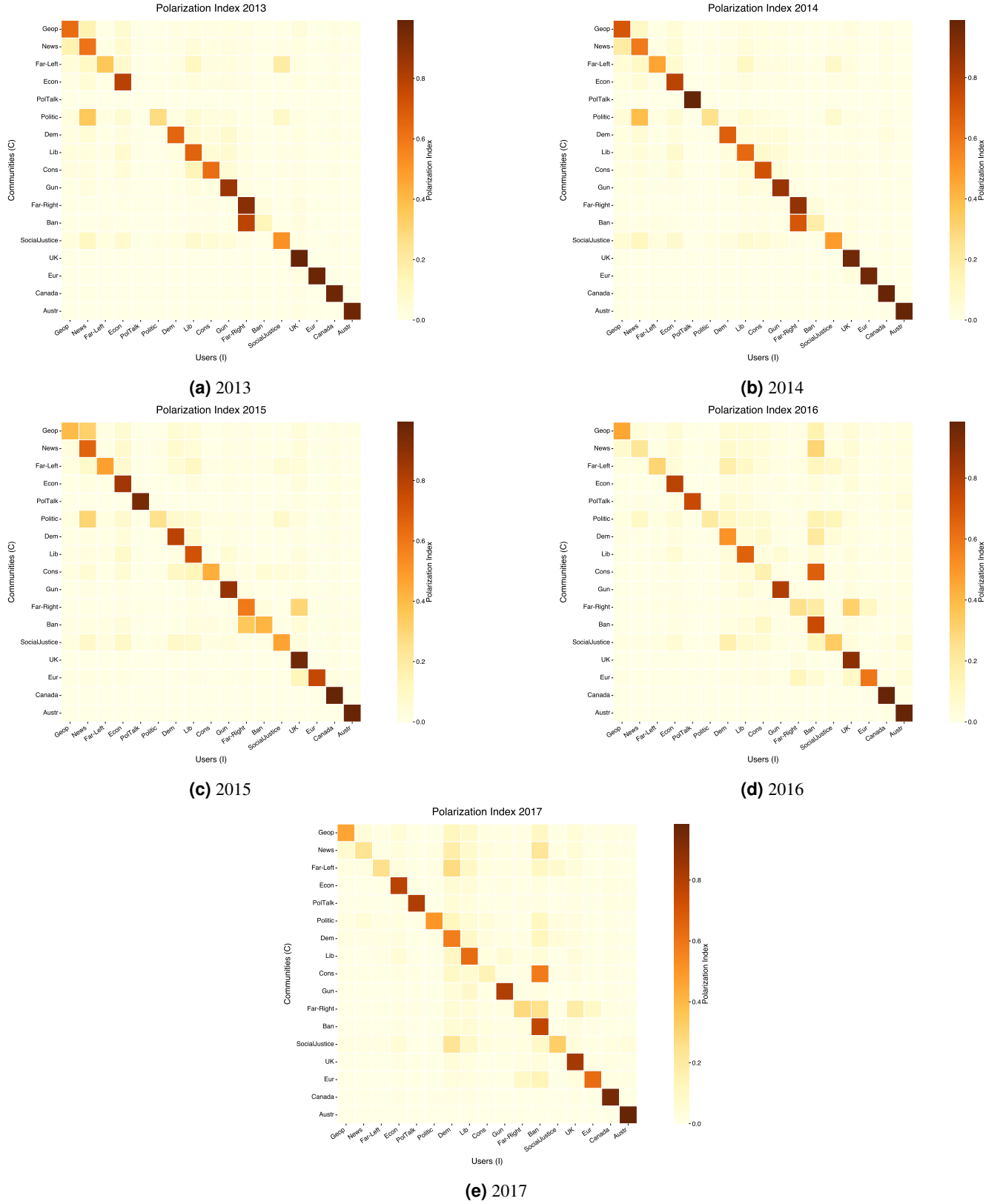


Figure 10. Polarization index of tag-based communities, 2013–2017. Each panel shows a heatmap of polarization values, with row labels corresponding to subreddit tags and column labels to the tags propagated to users. Diagonal entries capture within-tag polarization, while off-diagonal entries represent cross-tag interactions.

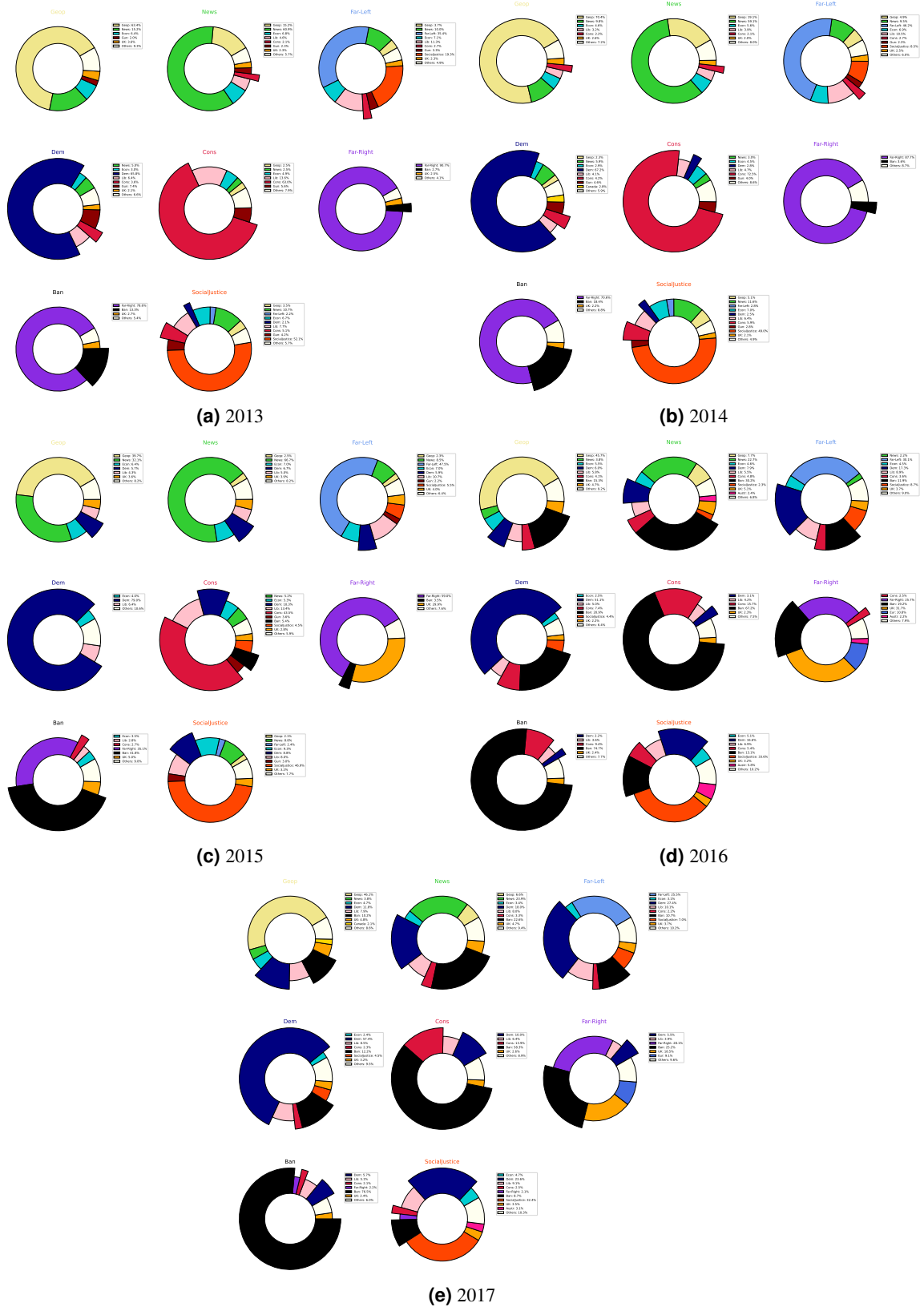
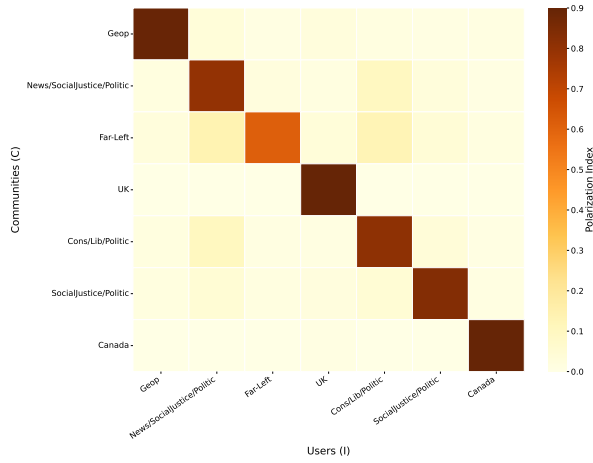
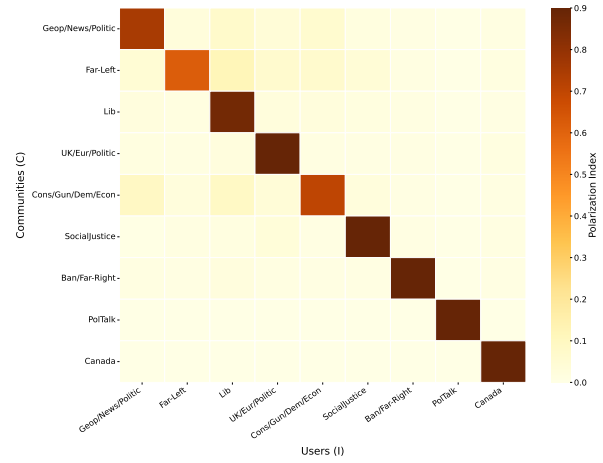


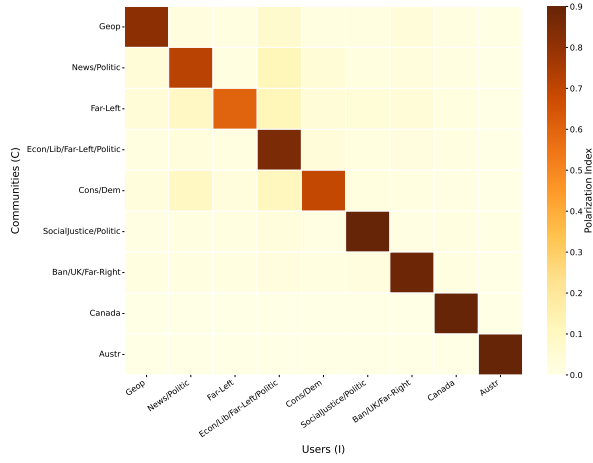
Figure 11. Donut charts showing the polarization of subreddit communities from 2013 to 2017. Each donut represents the distribution of users with assigned tags within a given community, highlighting the relative weight of Democrats, Conservatives, and Banned users in its composition. Categories below 2% are aggregated into “Others”.



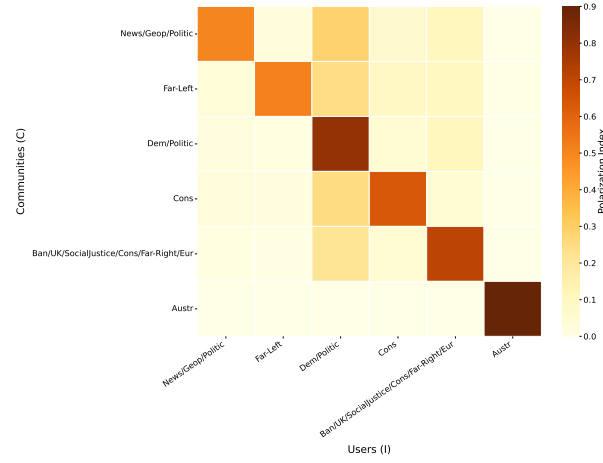
(a) 2013



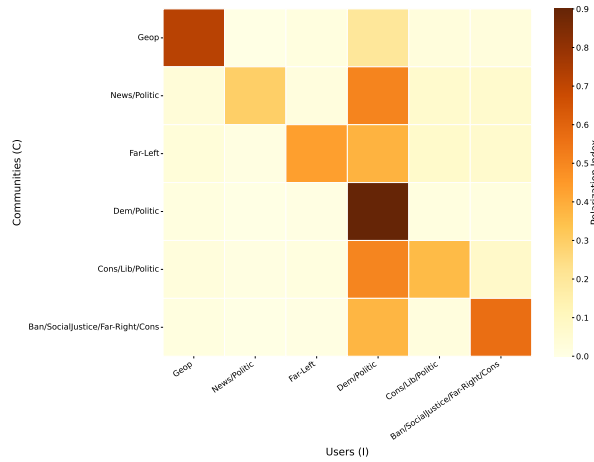
(b) 2014



(c) 2015



(d) 2016



(e) 2017

Figure 12. Polarization index of tag-communities in terms of users, 2013–2017. Each panel shows a heatmap of the polarization matrix between communities identified in the subreddit networks. Row labels indicate the detected communities, while column labels represent users with tags propagated from subreddit communities.

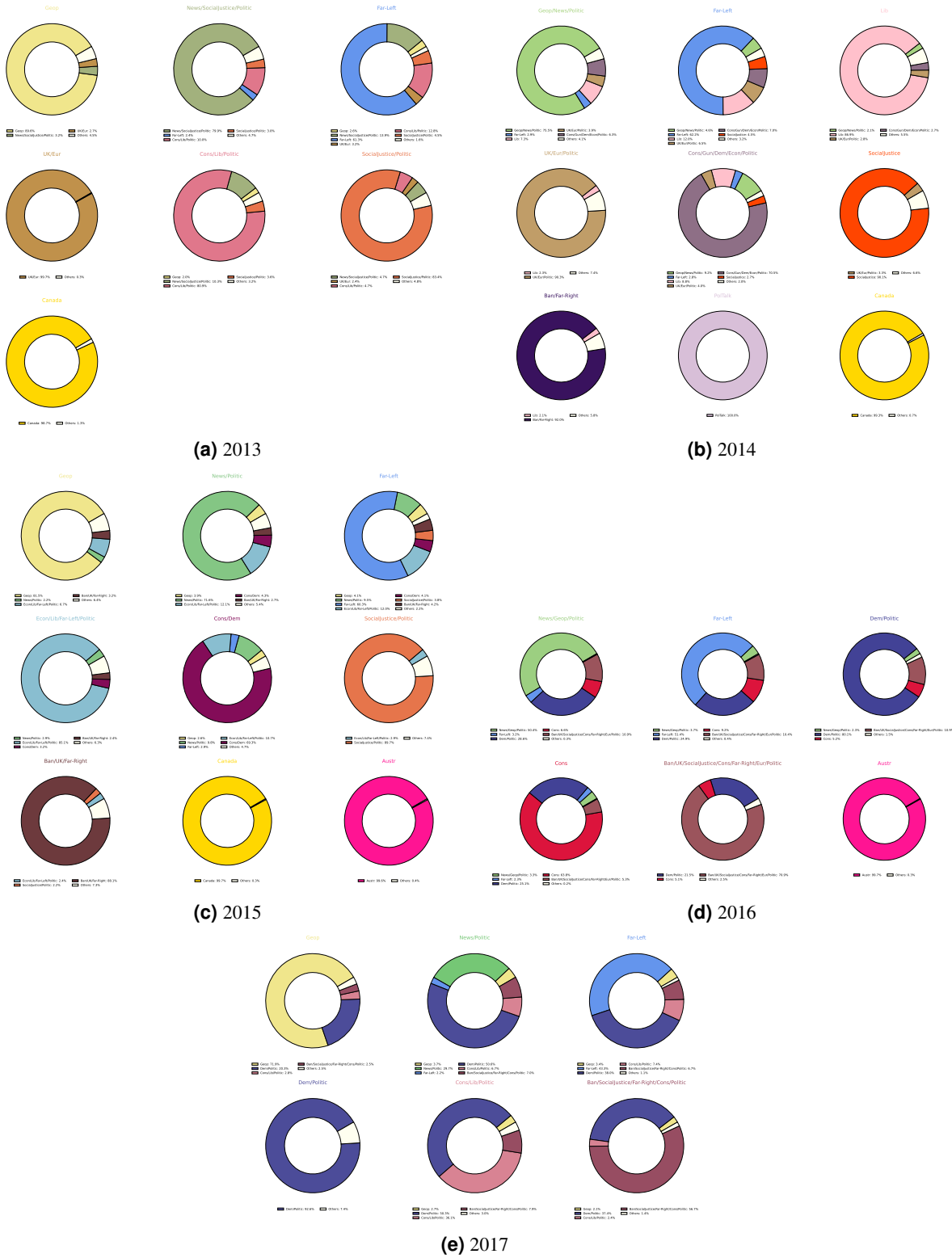


Figure 13. Donut charts showing the polarization of subreddit communities identified in the validated networks from 2013 to 2017. Each donut illustrates the distribution of labeled users populating the corresponding community. Categories representing less than 2% are aggregated into “Others”.

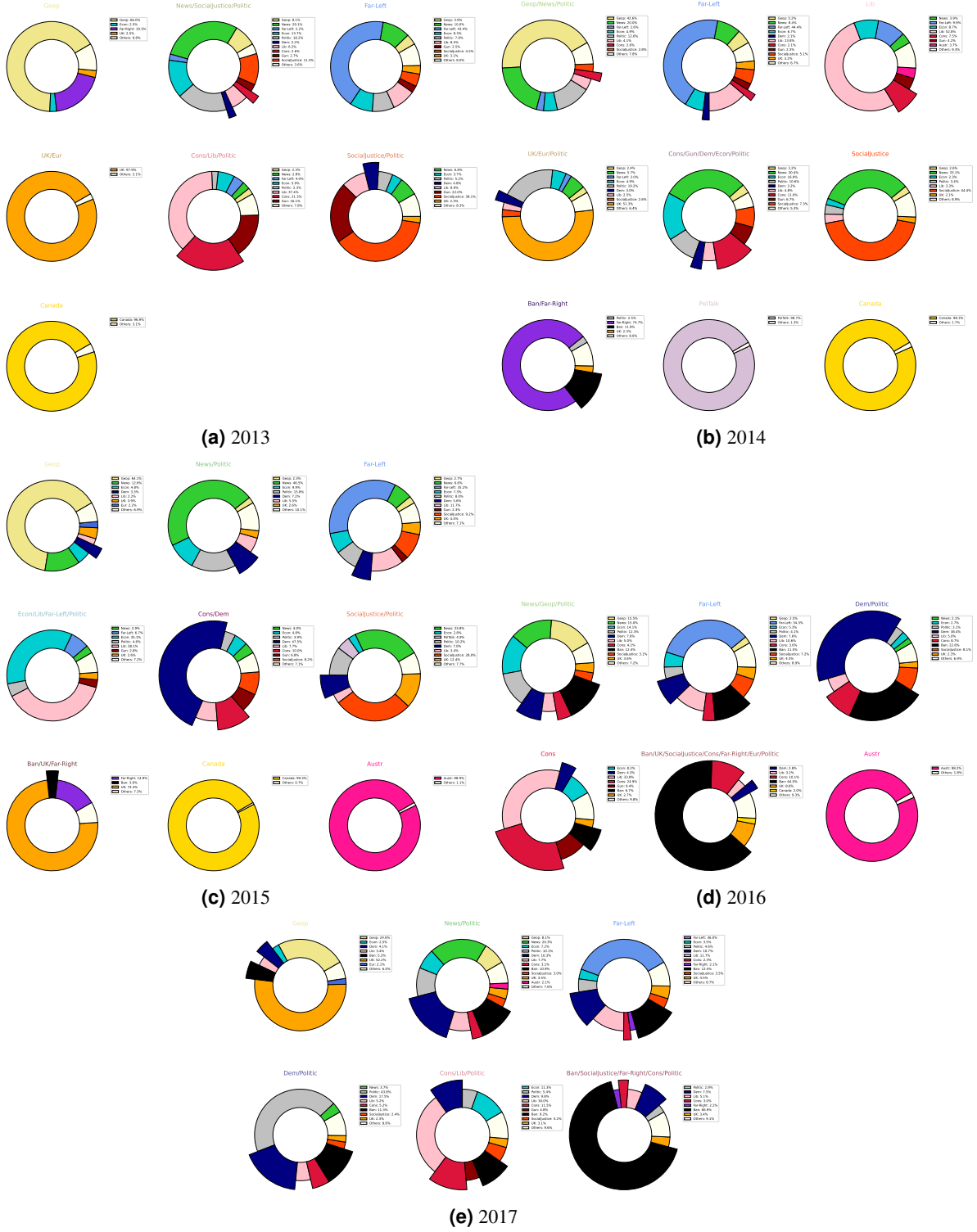


Figure 14. Polarization of subreddit communities identified in validated networks, analyzed at the user level. Each panel shows the population of network communities in terms of user tags for the years 2013–2017.

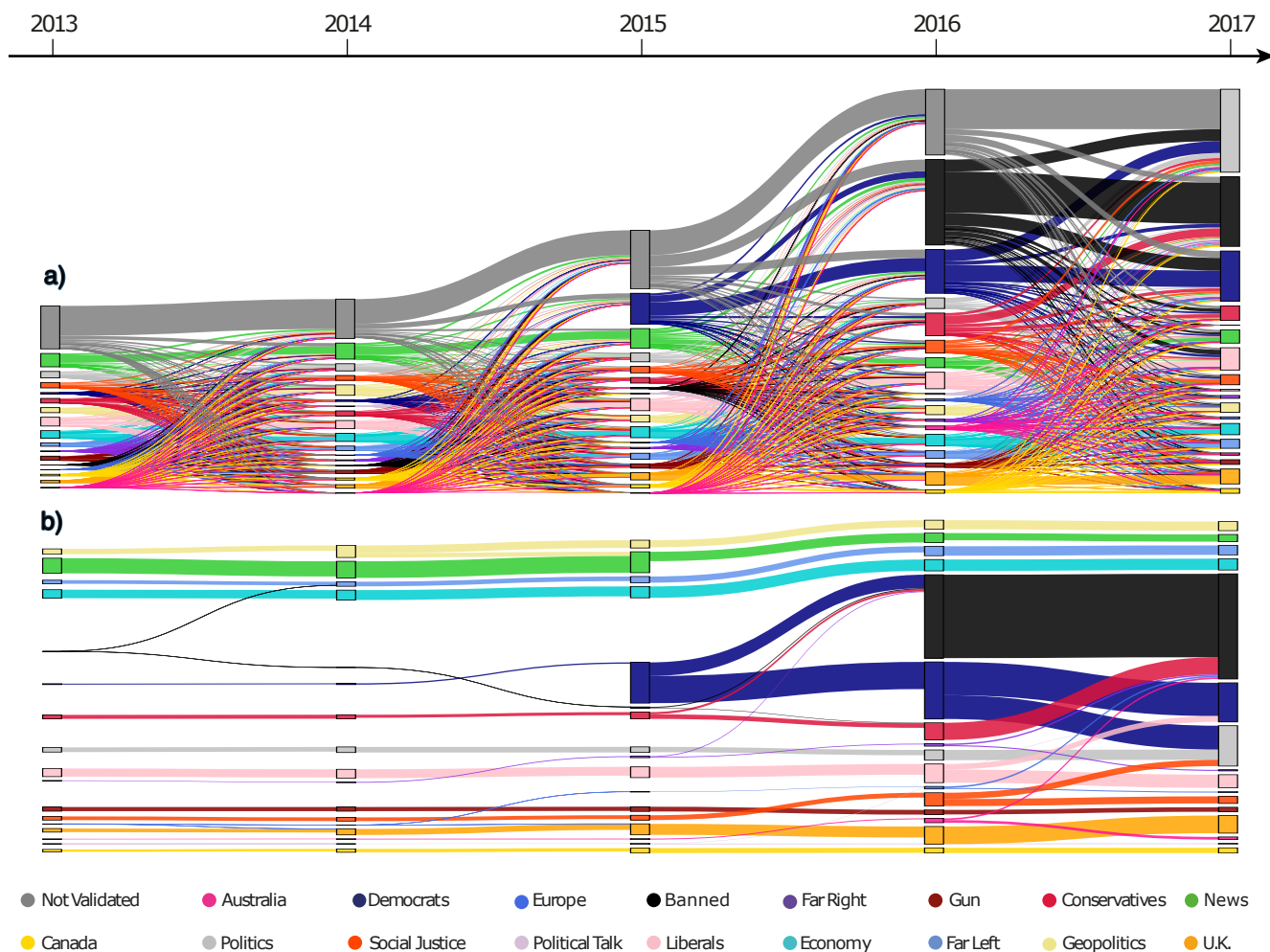


Figure 15. Flows of users across tag-defined groups, 2013–2017. (a) Full flowchart including all transitions; (b) restricted view including only flows representing at least 5% of the source community.

cross-group patterns. Before the 2016 elections, Banned subreddits display linguistic proximity to both Far-Right and Social Justice communities, reflecting mixed ideological influences. During the election year, however, the strongest convergence occurs between Democratic, Conservative, and Banned communities, indicating a temporary alignment of discourse around shared political narratives. This convergence is more evident in textual similarity than in interaction-based networks, where these groups remain structurally separated. Together, these results provide additional evidence for a gradual softening of linguistic boundaries, despite the ideological divisions observed in the main text.

Year	Communities p -value	Tags p -value
2013	2.53×10^{-5}	6.99×10^{-7}
2014	2.33×10^{-5}	4.30×10^{-10}
2015	9.75×10^{-6}	8.68×10^{-11}
2016	3.50×10^{-4}	1.53×10^{-10}
2017	3.11×10^{-3}	9.52×10^{-10}

Table 6. Mann–Whitney U test p -values for diagonal similarity significance, reported by year. Since only one set of tests was performed, the same values apply to both community- and tag-based comparisons.

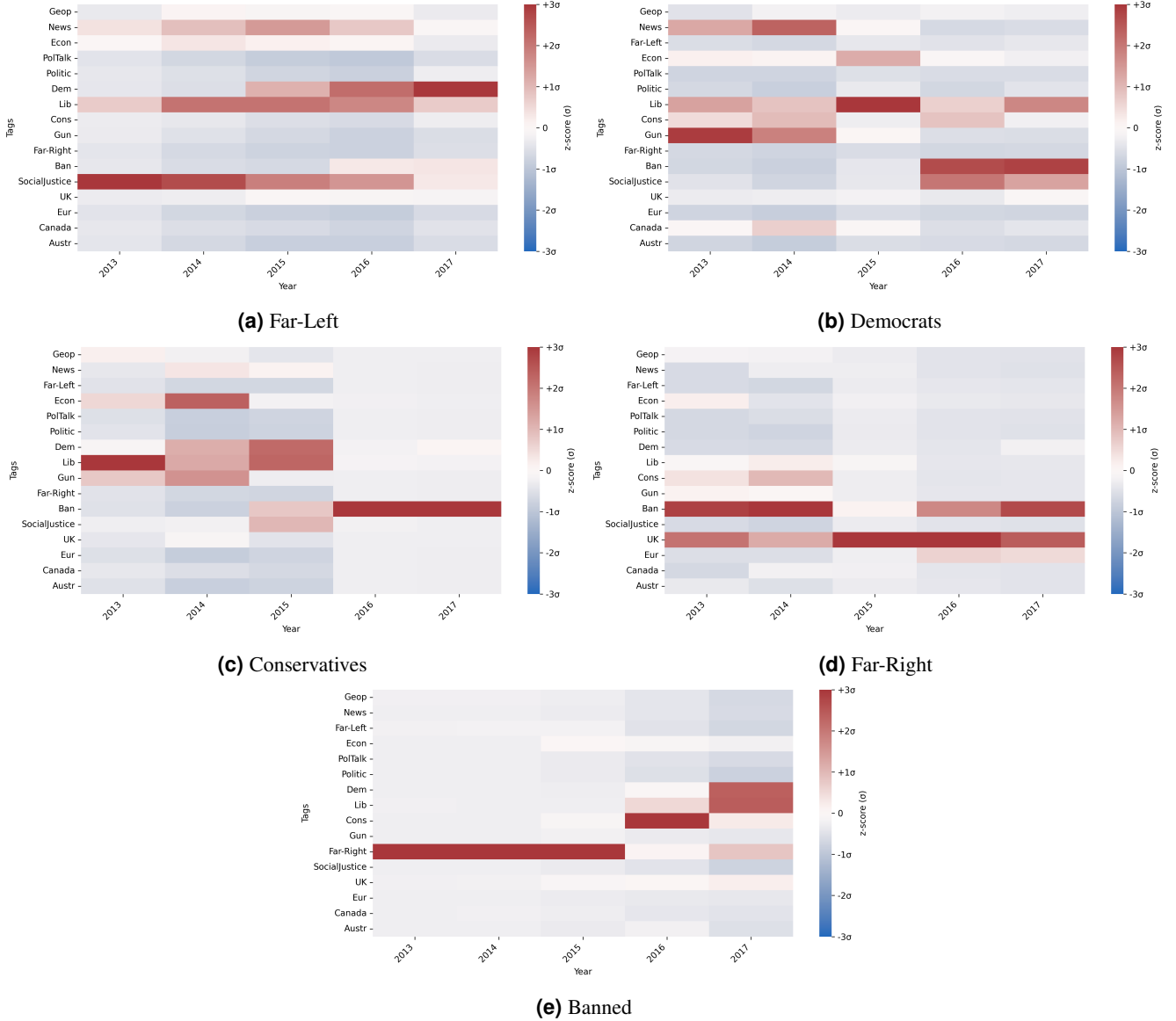


Figure 16. Leave-one-tag-out polarization maps for five reference tags. In each panel, rows indicate the *excluded* tag and columns the year. For every year, we recompute the polarization index of the reference tag after removing all label assignments of one other tag at a time, and report the standardized deviation as a z-score (in σ units) relative to the across-tag distribution in the same year. Positive values indicate that removing the excluded tag increases the reference tag’s polarization above its annual mean; negative values indicate a decrease. Warm/cool colors highlight tags whose removal consistently amplifies/attenuates the reference tag’s polarization, revealing overlap and coupling of user populations.

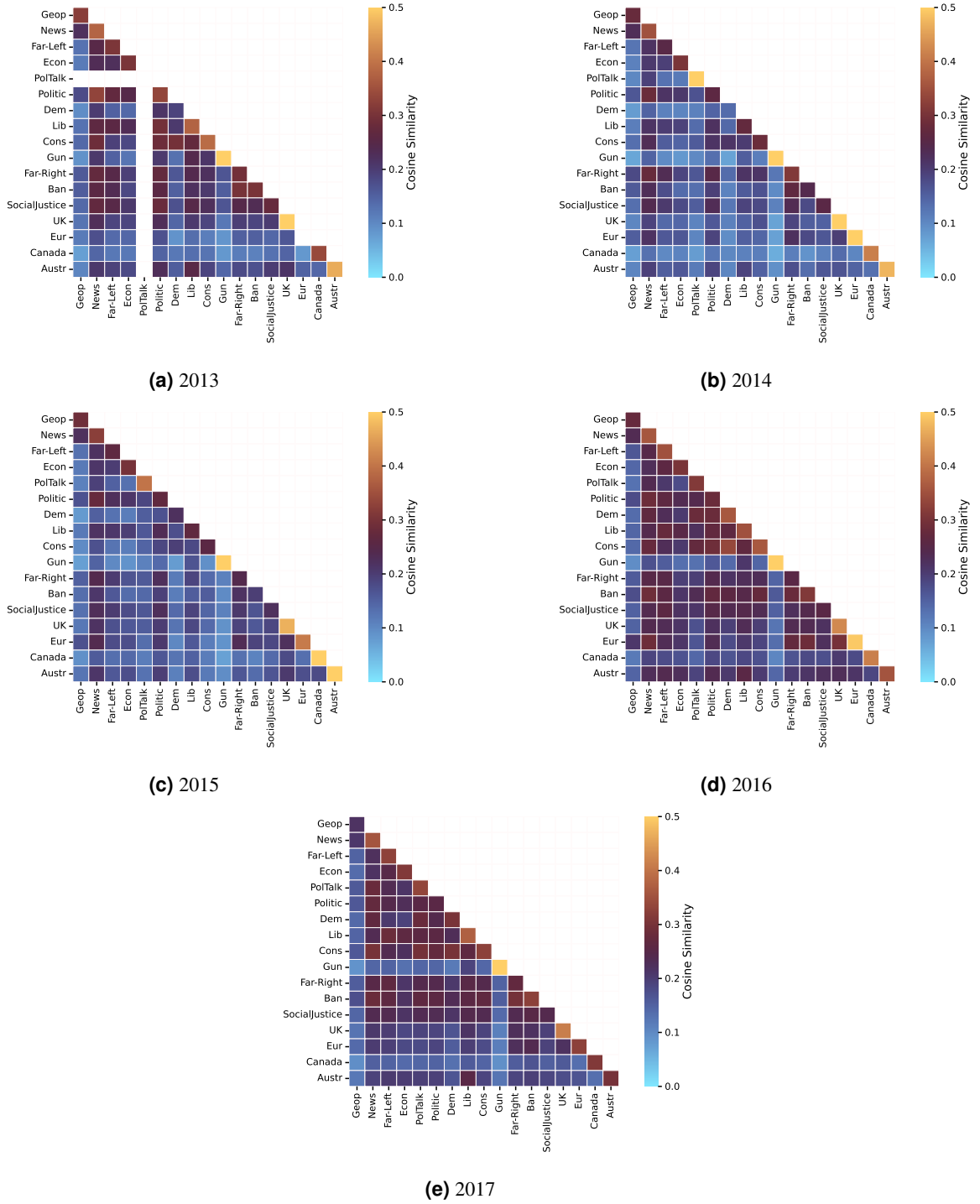


Figure 17. Heatmaps of cosine similarity between selected tags, 2013–2017. Each panel shows pairwise similarities among tag-based communities, with lighter colors indicating lower similarity.

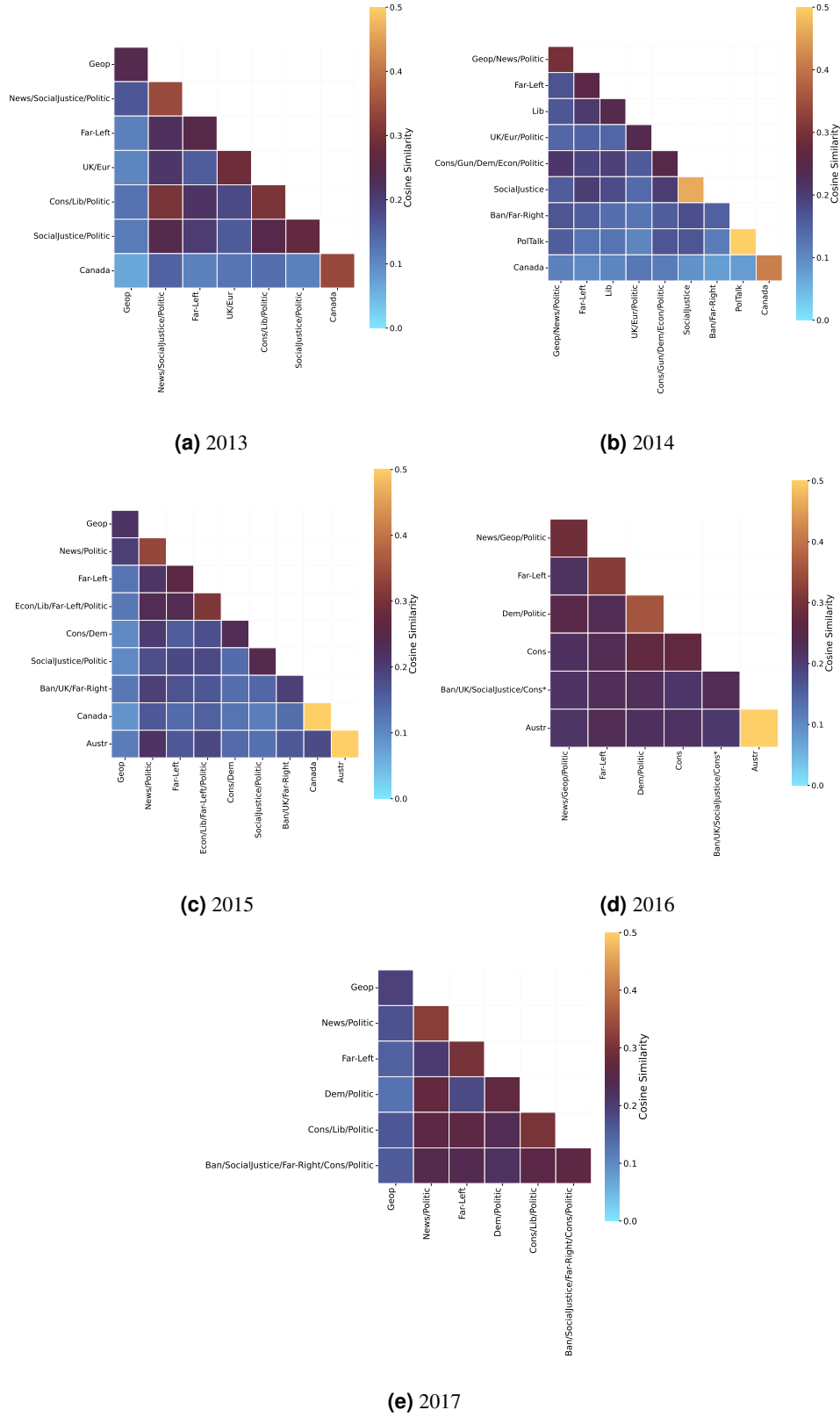


Figure 18. Heatmaps of cosine similarity between validated subreddit communities, 2013–2017. Each panel reports pairwise similarities among network-derived communities. For 2016, the label *Ban/UK/SocialJustice/Cons/Far-Right/Eur/Politics* has been shortened for readability.

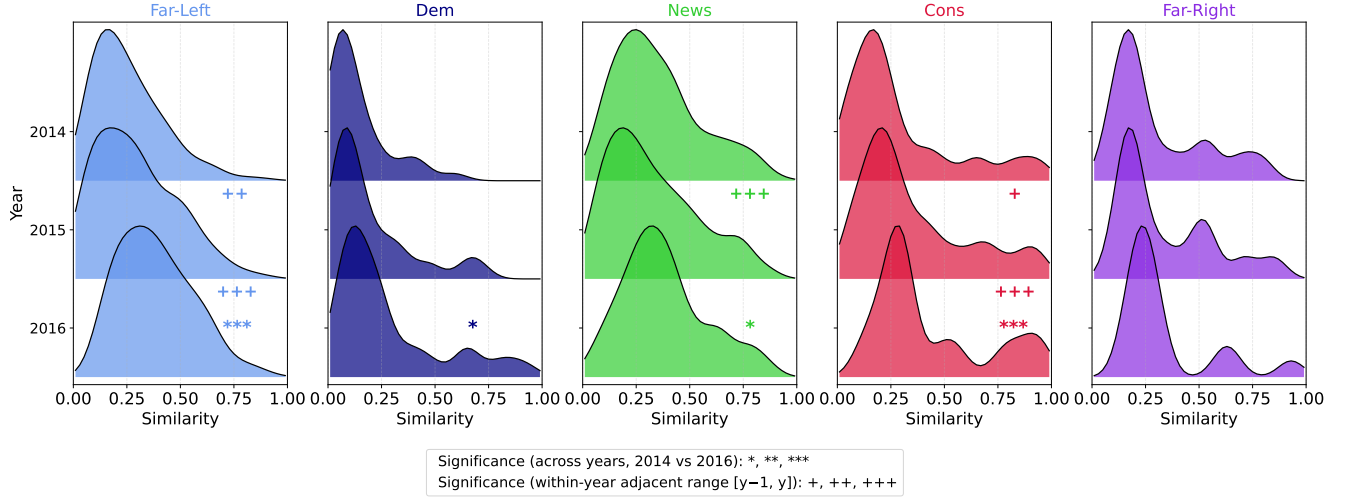


Figure 19. Joyplots showing distributional shifts in cosine similarity across years, restricted to subreddits active throughout 2014–2016. Two-sample Kolmogorov–Smirnov tests confirm significant upward shifts in textual similarity, consistent with results from Fig. 5 (main text).

Tag	KS p -value 2014 $\rightarrow$ 2015	KS p -value 2015 $\rightarrow$ 2016	KS p -value 2014 $\rightarrow$ 2016
Far-Left	(0.0013) ++	(< 10^{-14}) +++	(< 10^{-28}) ***
Dem	(0.51)	(0.21)	(0.018) *
News	(0.09)	(0.00015) +++	(0.045) *
Cons	(0.041) +	(0.0081) ++	(0.0000042) ***
Far-Right	(1.00)	(0.18)	(0.18)

Table 7. Significance of pairwise comparisons across years for each community group. p -values are truncated to two decimals. Significance thresholds: 0.05, 0.01, 0.001.

7 Domain sharing across subreddits

In this section, we examine the subreddit–domain bipartite networks to study how political orientations propagate through news sharing and how domains contribute to the overall polarization structure. Community tags assigned to subreddits are propagated to the domains they share, allowing us to characterize the political composition of the information ecosystem.

Figure 20e illustrates the distribution of subreddit tags across domains in 2017 through donut charts, highlighting how news outlets reflect the partisan alignment of the communities that circulate them. The overall patterns broadly mirror those observed in the interaction-based analysis: Far-Right participation within Banned communities is visible before the elections but then declines, while Banned participation in Conservative domains grows during the electoral period, though less sharply than in the interaction analysis.

Based on these mappings, we recalculated the polarization index (see **Methods** in the main text) for selected domain categories. As shown in Figure 21, polarization levels are generally lower than in the interaction-based analysis. A notable exception is represented by Far-Right subreddits, which display a marked decline in polarization when examined in terms of their shared domains over time, while Banned subreddits exhibit a more moderate increase compared to the stronger polarization observed in the interaction-based analysis.

Table 8 complements these results by reporting, for each year, the top 30 news domains most frequently shared by subreddits labeled as Far-Left, Democratic, Conservative, Far-Right, or Banned. Some domains clearly align with the propagated labels, reflecting the expected partisan orientation of the communities.

A few widely shared platforms, such as `youtube.com` or `reddit.com`, are instead labeled as Far-Right after normalization. This may occur because these mainstream domains are broadly present across multiple large communities, but their relative weight can appear amplified within smaller groups (particularly Far-Right subreddits), once normalization is applied during the label propagation process (see **Methods**). This does not substantially distort the partisan composition of Far-Right

domains, which remains consistent with the patterns observed in the interaction-based analysis. This overview provides a concise longitudinal perspective on domain-level sharing behavior across 2013–2017.

Table 8. Top 30 news domains shared across subreddit communities from 2013 to 2017. Labels are assigned through label propagation from subreddits to domains. The table reports the most frequently shared domains associated with Far-Left, Democratic, Conservative, Far-Right, and Banned communities.

Far-Left	Democrats	Conservatives	Far-Right	Banned
2013				
libcom.org anarchistnews.org news.infoshop.org 325.nostate.net climateandcapitalism.com anarkismo.net crimethinc.com theralmovement.wordpress.com indybay.org spiritofcontradiction.eu anti-imperialism.com readingisfornobs.com fightbacknews.org rosswolfe.wordpress.com labornotes.org marxist.com gonzotimes.com kasamaproject.org signalfire.org voluntaryvirtues.com wearmany.org propagandalalaland.blogspot.com bqbrew.com libertarian-labyrinth.blogspot.com ashevillefm.org theredplebian.wordpress.com moufawad-paul.blogspot.com litotpublishing.org prisoncensorship.info democracyatwork.info	boldprogressives.org obamacare.healthinsuranceexchangeenvoy.com corentewire.com au.org genelalor.com egbertowillies.com thepoliticalpragmatic.blogspot.com wapo.st fdlaction.firedoglake.com pensstoreview.com politi.co cir.ca imageshack.us poy.time.com truthnews.wordpress.com wegoted.com 4wheeledlefty.com boompopmedia.com act.boldprogressives.org allthingsdemocrat.com bobcesca.thedailybanter.com elections.firedoglake.com posting.org ezkool.com images.politi.co act.credodaction.com illuminate.newsvine.com nashvillescene.com eaglerising.com theobamadiary.com	hotair.com frugal-cafe.com redstate.com newser.com unitedliberty.org fireandreamitchell.com cagle.com lifefews.com powerlineblog.com cleveland.com blogs.rollcall.com volokh.com rollcall.com randpaulreview.com fivethirtyeight.blogs.nytimes.com yidwithlid.blogspot.com jammiewf.com foxbusiness.com rare.us publicpolicy polling.com people-press.org us.cnn.com politicalwire.com newslineusa.com fox19.com mainwashed.com youngcons.com whitehousedossier.com variety.com freedomworks.org	youtube.com reddit.com nytimes.com i.imgur.com huffingtonpost.com washingtonpost.com imgur.com bbc.co.uk reuters.com cnn.com guardian.co.uk news.yahoo.com thinkprogress.org rawstory.com youtu.be theguardian.com alternet.org politico.com salon.com rt.com foxnews.com cbc.ca usatoday.com bloomberg.com npr.org abcnews.go.com breitbart.com telegraph.co.uk latimes.com dailymail.co.uk	revolutionarycommunist.org bluevirginia.us atheism.about.com news.outlookindia.com themuslimsue.wordpress.com mobile.bbc.co.uk thebiglead.com everydayfeminism.com peopleslawoffice.com tealeafnation.com wlrn.org kleinonline.wnd.com careandwashingofthebrain.blogspot.com patriotaction.net freedomportal.net ozconservative.blogspot.com rinocracy.com anepigone.blogspot.com k0nsl.org sunnyisright.com yadadarcyada.wordpress.com alchemyoftheword.net asstr.org crisisrepublic.com streetcarnage.com coastalcoed.wordpress.com gettingworse.co.uk nationstates.net returntothepit.com buzzle.com
2014				
marxists.org anarchistnews.org labornotes.org bayareaintifada.wordpress.com pslweb.org ashevillefm.org submedia.tv sooperarticles.com 325.nostate.net hq-law.com socialistworld.net en.internationalism.org apsense.com climateandcapitalism.com thepiratebay.se dwardmac.pitzer.edu blog.designs.codes thetechcult.com litotpublishing.org detroitinquiry.org en.contrainfo.espiv.net ic.pics.livejournal.com isreview.org kasamaproject.org fightimperialism.org rdwolff.com thenorthstar.info signalfire.org avtonom.org revolutionarycommunist.org	pensstoreview.com snopes.com ezkool.com killingthebreeze.com colbertnation.com aquidneckinquirer.typepad.com messagapresident.com greencarreports.com acasignups.net pharmpro.com barackobama.com globalnewsnetwork.us mtcowgirl.com postcrescent.com progresscentral.x10.mx theconservativepundit.net allthingsdemocrat.com datab.us palmerreport.com codebluepolitics.com enewspf.com greencarcongress.com healthinsurance.org music.yahoo.com blog.pfaw.org consumerreports.org globegazette.com jaybookman.blog.ajc.com politicaloutcast.com softballpolitics.com	msnbc.com hotair.com mediaite.com newsbusters.org politicaltucker.blogs.cnn.com newsmax.com powerlineblog.com tampabay.com therightscoop.com ace.mu.nu twinsopinion.com timesdispatch.com appalachianareanews.com blog.heritage.org reviewjournal.com finance.townhall.com m.townhall.com unitedliberty.org host.madison.com conventionofstates.com video.foxnews.com minx.cc media.townhall.com libertymindsbreakfree.com nbcnewyork.com m.washingtonexaminer.com jammiewf.com buzzpo.com stevegoddard.wordpress.com joannenova.com.au	youtube.com nytimes.com reddit.com bbc.co.uk washingtonpost.com reuters.com en.itar-tass.com i.imgur.com huffingtonpost.com npr.org rt.com online.wsj.com cnn.com imgur.com ibtimes.com aljazeera.com np.reddit.com news.yahoo.com breitbart.com bbc.com youtu.be foxnews.com twitter.com salon.com thinkprogress.org telegraph.co.uk dailymail.co.uk hosted.ap.org theatlantic.com	robhoey127.blogspot.ca mrc.org jsmstateofmind.com onourselvesandothers.com grizzom.blogspot.com.br morningsstarnews.org premiumtimesng.com contextflorida.com dnj.com peoplebranch.wordpress.com conservativeinfidel.com kwiksurveys.com modiforpm.org themedicalbag.com thenewage.co.za androidhippo.com m.france24.com news.siteintelgroup.com static3.businessinsider.com atlantamuslim.com montgomerynews.com newsweekpakistan.com noliesradio.org pushbacknow.net vho.org culturewars.com oi61.tinypic.com pi-news.net ushmm.org winstonsmithministryoftruth.blogspot.co.uk
2015				
libcom.org marxists.org earthfirstjournal.org theanarchistlibrary.org fireworksbyarea.com crimethinc.com workers.org new-compass.net revolutionaryrds.wordpress.com jw.org en.squat.net havanatimes.org socialistworld.net blackrosefed.org anti-imperialism.com thenib.com interkkt.gr ashevillefm.org leftvoice.org indybay.org blog.designs.codes mer-rsm.ca fightimperialism.org	ezkool.com samuel-warde.com liberalvaluesblog.com climatecrops.com politi.co nationalmemo.com vtdigger.org eoinhiggins.blogspot.com i2.cdn.turner.com democraticunderground.com reachoutjobsearch.com dandygoat.com monmouth.edu store.bernieanders.com bernie2016events.org electbernieanders2016.blogspot.com.br n.pr ontheissues.org 2016.democracyforamerica.com pac.petitions.moveon.org politicickernj.com wikipedia-sucks-badly.blogspot.com act.credodaction.com	cnn.com msnbc.com hotair.com politifact.com realeclearpolitics.com usnews.com yahoo.com newsbusters.org redstate.com desmoinesregister.com twinsopinion.com imgflip.com onpolitics.usatoday.com bostonherald.com lifefews.com publicpolicy polling.com video.foxnews.com conservativeview.com thefiscaltimes.com spectator.org media.townhall.com hughewitt.com wmur.com	youtube.com bbc.co.uk i.imgur.com twitter.com np.reddit.com breitbart.com rt.com thehill.com news.yahoo.com independent.co.uk zerohedge.com vox.com thinkprogress.org foxnews.com hosted.ap.org washingtontimes.com vdare.com abcnews.go.com slate.com nationalreview.com usatoday.com nbcnews.com bigstory.ap.org	imgur.com israelandstuff.com truthrevolt.org elections.huffingtonpost.com dcwhispers.com trunews.com forum.codoh.com ulstermanbooks.com nrconservative.com en.shafaqna.com theconservativepundit.net electoral-vote.com gospelherald.com theendofzion.com langerresearch.com desertpeace.files.wordpress.com indiafacts.co.in iop.harvard.edu profootballtalk.nbcsports.com en.metapedia.org nbc.com writing.wikinut.com talkoakland.org

Far-Left	Democrats	Conservatives	Far-Right	Banned
isreview.org thenorthstar.info pflp.ps insurrectionnewsworldwide.wordpress.com ocap.ca chuangen.org existentialcomics.com	gravismarketing.com kspr.com pplswar.wordpress.com suffolk.edu bet.com declass3.com nationalnursesunited.org	republicandojo.com politicsandfinance.blogspot.com quinnipiac.edu laprogressive.com us.redd.it rushlimbaugh.com steynonline.com	dailycaller.com reason.com forbes.com arstechnica.com washingtonexaminer.com americanthinker.com buzzfeed.com	asadmahmudexposingislam.blogspot.ie fundrazr.com projectveritas.com dailynewsservice.co.uk humansofjudaism.com vibe.com aish.com

2016

jornada.unam.mx libcom.org marxists.org theanarchistlibrary.org rodong.rep.kp submedia.tv thenorthstar.info anarchistnews.org cpp.ph crimethinc.com communismgr.blogspot.gr youcaring.com earthfirstjournal.org insurrectionnewsworldwide.com links.org.au existentialcomics.com en.granma.cu social-ecology.org cdn.thedailybeast.com xenagoguevicene.com bunkermag.org cpusa.org new-compass.net ndfp.org anarchism.pageabode.com chomsky.info prisoncensorship.info s1.webmshare.com anti-imperialism.com blackrosefed.org	thehill.com commondreams.org inquisitr.com alternet.org democracynow.org usuncut.com truthdig.com celebritybabies.people.com elespectador.com berniesanders.com us.blastingnews.com go.berniesanders.com stylenews.people.com rightwingwatch.org rollicall.com patheos.com m.dailykos.com scribd.com elections.huffingtonpost.com pastemagazine.com reverpress.com trofire.com greatideas.people.com dailynewsb.in bluenationreview.com secure.actblue.com sfchronicle.com caucus99percent.com opendns.com cc.com	tampabay.com themoralofthistory.us therightscoop.com thepoliticalinsider.com dcwhispers.com m.washingtontimes.com theresurgent.com 270towin.com spectator.org rushlimbaugh.com newsninja2012.com endingthefed.com youngcons.com trumping.com weaselzippers.us americafans.com legitgov.org nationalmemo.com americasfreedomfighters.com ijreview.com natmonitor.com downstreampolitics.com politistick.com rightwingnews.com fox6now.com wapo.st conservativelead.com gop.com 100percentfedup.com ace.mu.nu	youtube.com reddit.com i.sli.me i.redd.it independent.co.uk i.imgur.com youtu.be cnn.com sli.me foxnews.com np.reddit.com sputniknews.com usatoday.com nbcnews.com archive.is cbsnews.com nypost.com zerohedge.com bbc.com dailycaller.com rt.com latimes.com wikileaks.org hosted.ap.org wsj.com yahoo.com thegatewaypundit.com thedailybeast.com espn.com washingtonexaminer.com	video.foxnews.com dailywire.com westernjournalism.com redflagnews.com ibankcoin.com refugeeresttlementwatch.wordpress.com conservativedailypost.com theruthdivision.com thedailysheep.com christianpost.com savenmysweden.com nowtheendbegins.com thecommonsenseshow.com trump-conservative.com milo.yiannopoulos.net rense.com christiantoday.com superstition95.com articles.latimes.com newswithviews.com joeforamerica.com tammybruce.com barstoolsports.com jookos.com discoverthenetworks.org theburningplatform.com americanfreepress.net beltwaytimes.com supload.com openmedianews.com
--	--	--	---	--

2017

libcom.org earthfirstjournal.org itsgoingdown.org anti-imperialism.org theanarchistlibrary.org muraselon.com crimethinc.com antifascistnews.net pics.me.me socialistalternative.org nunonline.org wsn.ie internationalist.org viewpointmag.com 78.media.tumblr.com philippinerevolution.info uniconriot.ninja dsausa.org linksunten.indymedia.org anarchistnews.org kuow.org monthlyreview.org redd.it leftcom.org borderedbyisolation.noblogs.org redguardsaustin.wordpress.com redspark.nu pics.mclatchyinteractive.com existentialcomics.com countercurrents.org	thirtyeight.com medialite.com politicususa.com aol.com medpagetoday.com thinkadvisor.com justice.gov esquire.com c-span.org mic.com 84.com mediamatters.org civilbeat.org billmoyers.com georgiapol.com amp.cnn.com epa.gov socialistworker.org people.com sierraclub.org dailydot.com biologicaldiversity.org palmerreport.com bustle.com extranewsfeed.com thenevadaindependent.com us.blastingnews.com climatecentral.org wnyc.org washingtonmonthly.com	patriotpost.us therightscoop.com mynewsgru.com theblackosphere.net conservativestories.com ibledredwhiteblue.com pacificpundit.com arcamax.com politicshicks.com theodysseyonline.com trump-conservative.com comicinconnect.wpengine.netdna-cdn.com trumpmovements.com toledoblade.com lifeandabout.com chicksongheright.com stream.org prageru.com westernfreepress.com patriotretort.com thedailynewscycle.com hollywoodintoto.com freshmedianews.com thehayride.com lidblog.com foxsports.com usainfonow.com donsurber.blogspot.com energy.gov floppingaces.net	i.redd.it reddit.com youtube.com twitter.com miamiherald.com reuters.com abcnews.go.com independent.co.uk cnn.com i.imgur.com youtu.be forbes.com nbcnews.com hosted.ap.org philly.com telegraph.co.uk washingtonexaminer.com smh.com.au archive.is dailycaller.com japantimes.co.jp bloomberg.com sciencedaily.com yahoo.com nationalreview.com facebook.com zerohedge.com apnews.com wsj.com pbs.twimg.com	theguardian.com nytimes.com foxnews.com imgur.com washingtontimes.com nypost.com breitbart.com npr.org i.redd.it imgur.com i.magaing.net oann.com bangkokpost.com rt.com dailymail.co.uk israelnationalnews.com iol.co.za timesofisrael.com thegatewaypundit.com vnd.com newsweek.com dailywire.com time.com google.com liferette.com infowars.com nydailynews.com i.sli.me freebeacon.com insider.foxnews.com vshare.com
--	--	---	---	---

8 Communities and subcommunities in subreddit networks

In this section, we provide an overview of the validated subreddit networks, starting with basic network statistics. Tables 9 and 10 report the structural properties of the user-based and domain-based subreddit networks (2013–2017), including their bipartite origins and the corresponding validated projections. In both cases, validation leads to higher modularity and lower density than the non-validated networks, consistently sharpening community structure across years (see also SI2).

To further characterize the identified communities, Figures 22 and 23 display radar plots summarizing their internal composition. Each polygon corresponds to one community, with axes representing topical categories. These visualizations highlight differences in purity across communities: while some clusters are dominated by a single topical label—such as the Far-Left community in the user-based networks—others exhibit a more heterogeneous composition, for instance the mixed clusters combining Democratic and Conservative subreddits together with Gun-rights or general Politics forums. The radar plots derived from user-based and domain-based networks show broadly similar patterns, reinforcing the robustness of the detected community structures.

Looking at specific cases, in the early years Democratic and Conservative subreddits often appear within the same

community, although Democrats are sometimes closer to News-oriented forums, while Conservatives co-occur more frequently with subreddits on guns, libertarianism, or economics—a pattern especially clear in 2015. Starting from 2016 in the user-based networks, partisan blocs are more sharply defined: one community groups Democrats and several Conservative subreddits, including candidate-specific forums such as `r/KasichForPresident` and libertarian pages, while another cluster brings together Banned forums and Trump-related subreddits, most notably `r/The_Donald`, alongside extremist spaces such as `r/NationalSocialism`, `r/WhiteRights`, and European nationalist forums like `r/Le_Pen`. Among such cases, an illustrative hybrid is `r/RepublicansForSanders`, which consistently clusters with Democratic communities despite its partisan label, highlighting clear overlaps in user activity across partisan boundaries.

In the domain-based networks, Democrats and Conservatives remain within the same large community during the electoral years, albeit positioned at opposite ends of it; they eventually split into distinct clusters in 2017. Several subreddits, such as `r/prolife`, `r/AltRightChristian`, or `r/climateskeptics`, appear within Conservative or Banned communities across both user- and domain-based networks, underlining their bridging role between mainstream conservative and more extremist discourse.

following the approach outlined in the **Results** section, we applied the Louvain algorithm a second time to the induced subgraphs of each primary community in order to uncover fine-grained subcommunities for the echo-chamber analysis (see **Results** and **SI8**). Figure 34 highlights two representative cases from the user-based networks in 2016 (density = 0.139, average modularity = $0.44622 \pm 1.9 \times 10^{-4}$) and 2017 (density = 0.180, average modularity = $0.35132 \pm 1.2 \times 10^{-4}$) both characterized by high modularity and strong internal cohesion. These partitions reveal distinct Conservative/Banned and Far-Right/Banned components, as well as SocialJustice-related clusters. Notably, in both years the forum `r/SargonOfAkkad` emerges as a bridging node, consistently linking otherwise distant communities. These higher-resolution partitions illustrate the presence of well-defined substructures nested within broader partisan clusters.

Table 9. Network statistics for the bipartite subreddit–user networks and their subreddit projections (2013–2017).

Year	Bipartite (Subreddit–User)				Projection (unvalidated)			Projection (validated)		
	Subr. nodes	User nodes	Links	Density	Nodes	Density	Avg. Modularity	Nodes	Density	Avg. Modularity
2013	214	387 691	562 139	7.47e-06	214	0.549	0.048950 ($\pm 9.2e-05$)	167	0.101	0.37234 ($\pm 2.5e-04$)
2014	246	308 382	480 322	1.01e-05	246	0.574	0.048520 ($\pm 7.0e-05$)	203	0.0806	0.43829 ($\pm 2.7e-04$)
2015	296	424 172	673 957	7.48e-06	296	0.681	0.043548 ($\pm 3.2e-05$)	246	0.0780	0.45288 ($\pm 1.6e-04$)
2016	413	888 444	1 941 907	4.92e-06	413	0.728	0.037757 ($\pm 3.9e-05$)	363	0.0868	0.42752 ($\pm 3.7e-04$)
2017	481	1 027 388	2 129 530	4.03e-06	479	0.696	0.043539 ($\pm 3.1e-05$)	417	0.0788	0.41670 ($\pm 3.1e-04$)

Table 10. Network statistics for the bipartite subreddit–domain networks and their subreddit projections (2013–2017).

Year	Bipartite (Subreddit–Domain)				Projection (unvalidated)			Projection (validated)		
	Subr. nodes	Domain nodes	Links	Density	Nodes	Density	Avg. Modularity	Nodes	Density	Avg. Modularity
2013	216	68 809	122 284	5.13e-05	216	0.749	0.045021 ($\pm 7.4e-05$)	170	0.0731	0.41534 ($\pm 5.2e-04$)
2014	248	45 594	112 438	1.07e-04	248	0.738	0.051909 ($\pm 1.5e-05$)	197	0.0711	0.47419 ($\pm 3.5e-04$)
2015	284	45 443	128 874	1.23e-04	284	0.827	0.034889 ($\pm 2.9e-05$)	229	0.0868	0.36036 ($\pm 3.2e-04$)
2016	399	67 837	200 068	8.59e-05	399	0.786	0.041765 ($\pm 1.7e-05$)	342	0.0806	0.35139 ($\pm 2.5e-04$)
2017	461	65 877	200 643	9.12e-05	461	0.829	0.027294 ($\pm 5.8e-05$)	350	0.0656	0.35983 ($\pm 2.8e-04$)

9 Echo-chamber structures and statistical patterns

As detailed in the Results, we evaluated the alignment between subreddit communities identified in interaction-based and news-sharing networks by constructing overlap matrices, where each entry records the number of subreddits shared between a pair of partitions. Figures 35–39 display chord diagrams of community-overlap networks. Flows are proportional to the number of subreddits shared between partitions; source and target tiles are colored according to their communities, and each flow is shaded with the average of the two. Most flows closely match their endpoints, indicating that overlaps predominantly occur between communities with similar topical or partisan orientations.

To further investigate echo-chamber dynamics, we weighted community matches by the number of users active in both partitions, producing the edge-contribution (EC) matrices. These matrices quantify how much each pair of communities contributes to cross-participation, thus capturing echo chambers through user engagement patterns. We validated these matrices using the BiWCM null model [18] (see Methods in the main text), which filters out random noise and retains only statistically significant matches. Figures 40–44 report the validated overlap matrices together with the corresponding *p*-values. Each panel shows, on the left, the weighted bipartite network of interactions between user communities and domains, and on the right, the corresponding statistical validation through *p*-values obtained with the test.

Although community matches often emerge in correspondence with similar topical areas—even at very low p -values—few entries remain statistically significant once corrections for multiple hypothesis testing are applied using the false discovery rate ($\alpha = 0.05$).

This limitation points to a loss of resolution in the analysis. At the same time, many of the subgraphs identified through validated community matches display high modularity (see Fig. 34) and emerge as near-complete, fully connected components within the interaction-derived subreddit network. This combination of internal structure and external cohesion indicates that the large communities identified in the first partition are meaningful units. It also suggests that further resolution can uncover well-defined substructures nested within them. To address these resolution limits, we conducted a finer-grained study by applying community detection within the validated communities themselves, thereby uncovering subcommunities nested within the larger partisan clusters.

The results are shown in Figures 45–49, which present the analysis at the subcommunity level through adjacency matrices of statistically validated community matches. These figures also highlight the links that remain significant under FDR correction—corresponding to effective echo chambers—and indicate that most validated matches follow shared topical themes. In these plots, subcommunities have been reordered to maximize consistency across years, making topical correspondences clearer. For readability, multi-tag communities are abbreviated with their main components, and full names are provided in Tab. 11.

Table 12 complements this structural evidence by quantifying user participation in validated echo chambers. For each year, it reports the absolute number of users active in communities identified as echo chambers, their fraction relative to the total active users in the same year, and their distribution across partisan subcommunities. In particular, we highlight the share of EC users participating in Democratic/Conservative or Banned communities, expressed both in absolute counts and as fractions relative to all EC users of that year.

Year	Abbreviation	Full name
2013	Mixed Dem/Gun/Econ *	Dem/Gun/Econ/Lib/Politic
	Mixed Far-Left/Dem/Lib *	Far-Left/Dem/Lib/Politic
	Mixed Far-Left/News/Econ *	Far-Left/News/Econ/Eur/Politic
	Mixed Far-Left/PolTalk/Dem *	Far-Left/PolTalk/Dem/Ban/Politic
	Mixed Gun/Dem/News *	Gun/Dem/News/Politic
2014	Mixed Geop/News/Far-Left *	Geop/News/Far-Left/Econ/PolTalk/Dem/Lib/Cons/Gun/Far-Right/Ban/SocialJustice/UK/Eur/Canada/Austr
	Mixed Econ/Far-Left/Dem *	Econ/Far-Left/Dem/Politic
	Mixed Lib/Dem/Far-Left *	Lib/Dem/Far-Left/Econ
2015	Mixed Geop/News/Far-Left *	Geop/News/Far-Left/Econ/PolTalk/Dem/Lib/Cons/Gun/Far-Right/Ban/SocialJustice/UK/Eur/Canada/Austr
	Mixed Far-Left/SocialJustice/Econ *	Far-Left/SocialJustice/Econ/Politic
	Mixed SocialJustice/Far-Right/UK *	SocialJustice/Far-Right/UK/Politic
2016	Mixed Far-Left/News/Cons *	Far-Left/News/Cons/Ban/Politic
	Mixed Dem/Ban/News *	Dem/Ban/News/Far-Left/Cons/Politic
	Mixed Gun/Dem/News *	Gun/Dem/News/Lib
	Mixed Far-Right/Far-Left/Dem *	Far-Right/Far-Left/Dem/SocialJustice/Ban/Politic
	Mixed SocialJustice/News/Far-Left *	SocialJustice/News/Far-Left/PolTalk/Dem/Lib/Politic
2017	Mixed Geop/News/Far-Left *	Geop/News/Far-Left/Econ/PolTalk/Dem/Lib/Cons/Gun/Far-Right/Ban/SocialJustice/UK/Eur/Canada/Austr
	Mixed Dem/Cons/Geop *	Dem/Cons/Geop/Politic
	Mixed Far-Left/Dem/Cons *	Far-Left/Dem/Cons/Politic
	Mixed Far-Left/Cons/SocialJustice *	Far-Left/Cons/SocialJustice/Dem/Politic
	Mixed Lib/SocialJustice/Geop *	Lib/SocialJustice/Geop/News/Politic
	Mixed SocialJustice/Far-Left/Dem *	SocialJustice/Far-Left/Dem/News/Politic

Table 11. Full names of multi-tag communities abbreviated in the heatmaps of echo chambers at the sub-community level. Entries marked with an asterisk (*) correspond to the abbreviated forms displayed in the figures.

10 Insights on Democrats, Conservatives, and Banned communities

To further qualify the findings reported in the main text, we examined user activity within the Democratic, Conservative, and Banned-aligned communities. As shown in Fig. 50, panel (a) highlights a progressive increase in exclusive participation across partisan subreddits: Democratic commenters become increasingly concentrated in Democratic-labeled forums, and the same holds for Conservatives, while the fraction of users active in both spheres steadily decreases over time. Panel (b) shows that, despite this rise in partisan activity, the average score of comments in these communities declines sharply compared to 2013 levels, with decreases of 93.4% for Democratic subreddits, 85.6% for Conservative ones, and 97.7% for Banned forums. The decline is especially pronounced in candidate-specific subreddits: Sanders (99.9%), Clinton (100.0%), and Trump (98.6% compared to 2015, when his forum first appeared), all exhibit substantially lower average scores than their corresponding

Year	Valid. Users	Frac. Users / Tot.	Dem/Cons Users in EC	Frac. Dem/Cons / EC	Ban Users in EC	Frac. Ban / EC
2013	76 864	0.1983	34 395	0.4475	2 564	0.0334
2014	61 429	0.1992	31 550	0.5136	2 256	0.0367
2015	152 571	0.3597	16 612	0.1089	4 852	0.0318
2016	247 804	0.2789	88 026	0.3552	62 465	0.2521
2017	348 533	0.3392	177 617	0.5096	40 336	0.1157

Table 12. Validated users in echo chambers (EC). For each year, the table reports the total number of users active in validated echo chambers, their fraction relative to the total active users in the same year (Frac. Users / Tot.), as well as the number and fraction of EC users participating in Democratic/Conservative (Frac. Dem/Cons / EC) and Banned (Frac. Ban / EC) communities.

partisan communities, underscoring the increasingly polarizing dynamics of electoral discussions.

Figure 51 complements this analysis by focusing on the mutual distances between Democratic, Conservative, and Banned communities in the statistically validated subreddit projections. Distances are reported both at the user level (left) and domain level (right), in analogy with Fig. 5 of the main text. Beyond the general trends, here we explicitly test whether yearly changes are statistically significant by comparing each distribution of distances with that of the previous year through Kolmogorov–Smirnov tests. The results highlight significant shifts especially during electoral years, indicating that partisan and banned communities became more clearly separated as elections approached.

The previous analyses were carried out at the level of subreddits, meaning that measures such as comment scores could not disentangle whether low evaluations arose from interactions between supporters and opponents engaging in cross-posting activity, or from exchanges occurring within the same partisan group. To address this limitation and move to a finer resolution, we conducted a targeted analysis of Democratic-, Conservative-, and Banned-labeled users, as introduced in the main text and partially illustrated in Fig. 5. Specifically, we filtered the original dataset of comments by assigning labels to users through propagation from subreddit to user, and then retaining only comments authored by users tagged as Democrats, Conservatives, or Banned. The size of the resulting dataset is reported in Tab. 13.

Within this filtered set, we further identified comments directed specifically toward posts and comments authored by members of each faction. Figures 52 and 53 report, respectively, the normalized number of comments exchanged between groups and the average reception (scores) of these interactions. The row-normalized comment matrices reveal how communication volumes evolved across factions, while the score heatmaps capture systematic asymmetries in the evaluation of partisan and banned contributions across the 2013–2017 period.

Overall, the analysis shows that cross-group interactions increased over time, even though the majority of comments still occurred within each partisan community. At the same time, while average scores declined across all groups, cross-group exchanges were evaluated even more negatively, a pattern that became especially pronounced between 2015 and 2017.

Year	Democrats	Conservatives	Banned
2013	208 327	645 817	43 402
2014	219 528	493 419	40 672
2015	1 558 935	400 348	148 262
2016	9 326 035	6 946 677	8 095 653
2017	9 947 710	3 481 146	11 511 021

Table 13. Number of comments in the filtered dataset authored by users labeled as *Democrats*, *Conservatives*, and *Banned* between 2013 and 2017. Labels were assigned through label propagation from subreddits to users, and the total dataset was filtered according to these tags.

11 Robustness to temporal resolution

To probe the robustness of our findings at finer temporal scales, we partitioned the dataset into three four-month windows per year: y_1 (January–April), y_2 (May–August), and y_3 (September–December). The third window systematically captures the electoral cycles, including campaigning, election day, and immediate aftermath for both the 2014 midterms and the 2016 presidential election.

As shown in Fig. 54, the communities identified in four-month windows largely mirror those found in the annual analysis, with clusters emerging around the same topics and political orientations. In particular, we observe the formation of large

opposing communities, such as Democratic- and Conservative-aligned clusters, especially pronounced during electoral periods, as well as communities dominated by banned and far-right subreddits. Some differences arise due to the finer resolution: for instance, in the early years Democratic-aligned subreddits appear less distinct and are often embedded in larger clusters dominated by other labels, such as News. Around electoral periods, however, the polarization between opposing communities becomes sharper.

Nevertheless, polarization indices and user label compositions (Fig. 55) confirm that the core polarization and echo-chamber patterns are consistent with those observed in the annual analysis, while also revealing finer-grained shifts in the alignment of banned users across electoral periods. In particular, the 2016 windows reveal a decrease in overall polarization but a stronger association of banned users with Conservative subreddits. Echo-chamber analysis further replicates the alignment patterns observed in the annual setting.

Finally, inter-group distance measures (Fig. 56) reinforce these findings. Mean distances across validated networks show consistent trends with the annual analysis: Democratic–Banned distances remain the largest, decreasing slightly before elections, while Conservative and Banned communities remain close. Democratic–Conservative distances increase during electoral periods, with statistically significant shifts observed in selected four-month windows including 2014_3, and 2016_3. By contrast, peaks in Conservative–Banned distances at the end of 2015 and 2016 do not appear to correspond to particularly significant changes in the distance distributions relative to the rest of the network.

12 GPT TAG-assignment validation

We manually annotated subreddit tags by extending those in the Politosphere dataset and validated them using GPT-4 Turbo [19]. Initially, we prompted GPT-4 with the full list of subreddit names to suggest a minimal set of representative categories, and its outputs closely matched our manual scheme. Next, each subreddit—along with its public description—was submitted to GPT-4 to assign one or more of the chosen tags, allowing mixed labels. The resulting tag distribution was then refined by overriding assignments based on external metadata: subreddits known to have been prohibited were tagged “Banned,” and those with Democratic or Republican metadata were labeled accordingly. Tag scores were normalized per subreddit to account for multiple labels.

We observed an overrepresentation of the general “Politics” category, prompting a secondary reclassification in which subreddits labeled exclusively as “Banned” or “Politics” were re-evaluated and relabeled. Performance was then assessed through two complementary measures. At the subreddit level, we computed *weighted recall-like and precision-like scores*, which quantify, respectively, the fraction of true tags correctly recovered and the fraction of predicted tags that match the manual annotation, with weights ensuring that partially correct multi-label cases contributed proportionally. At the aggregate level, we measured the overall distributional agreement between manual and GPT labels using the *coefficient of determination* (R^2), which captures how well GPT reproduces the variance of the real distribution. Under this refinement, weighted precision and recall increased from 0.61 and 0.75 to 0.73 and 0.81, while R^2 improved markedly from 0.51 to 0.85.

In both classification rounds, despite a slight mismatch in the distribution of Democrats before the elections and of Far Right afterwards, the overall community labels and their population structure remain highly consistent with the patterns reported in the main text (Fig. 2 and Fig. 3). This consistency is further illustrated in Fig. 57, where panels (a)–(c) compare hand-labeled and GPT-corrected tag distributions, show donut charts of 2016 polarization at the user level, and trace the yearly formation of subreddit communities.

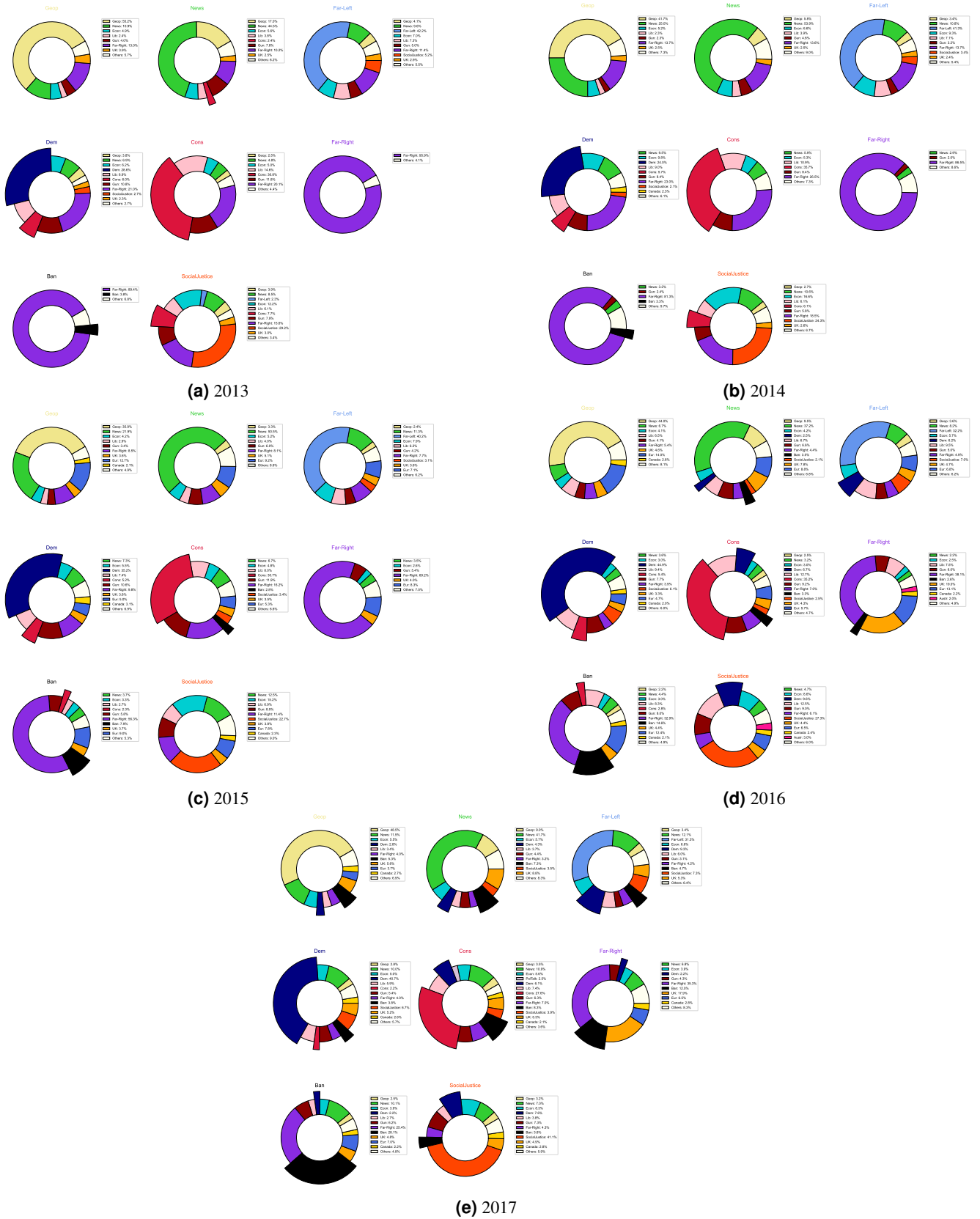


Figure 20. Donut plots showing the distribution of news domains across communities from 2013 to 2017. Each panel represents one year and highlights the relative contribution of domains linked to different political or thematic communities.

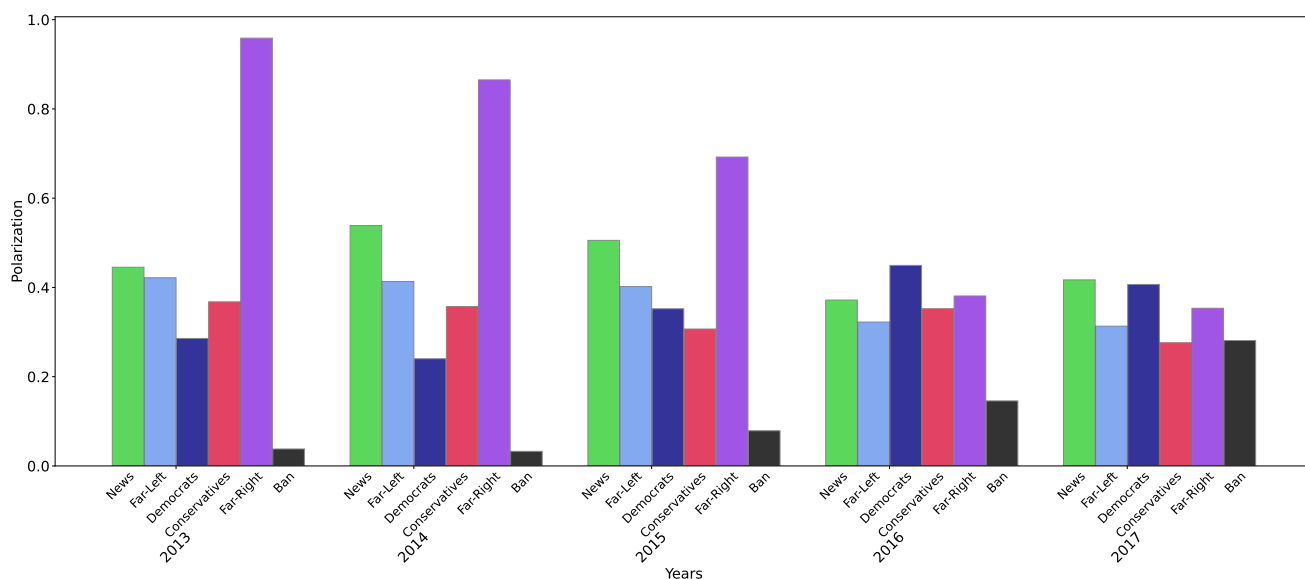


Figure 21. Polarization indices of news domains from 2013 to 2017. Each panel reports the diagonal elements of the polarization matrix (see Methods), capturing the extent to which domains inherit partisan alignment from the subreddits that share them.

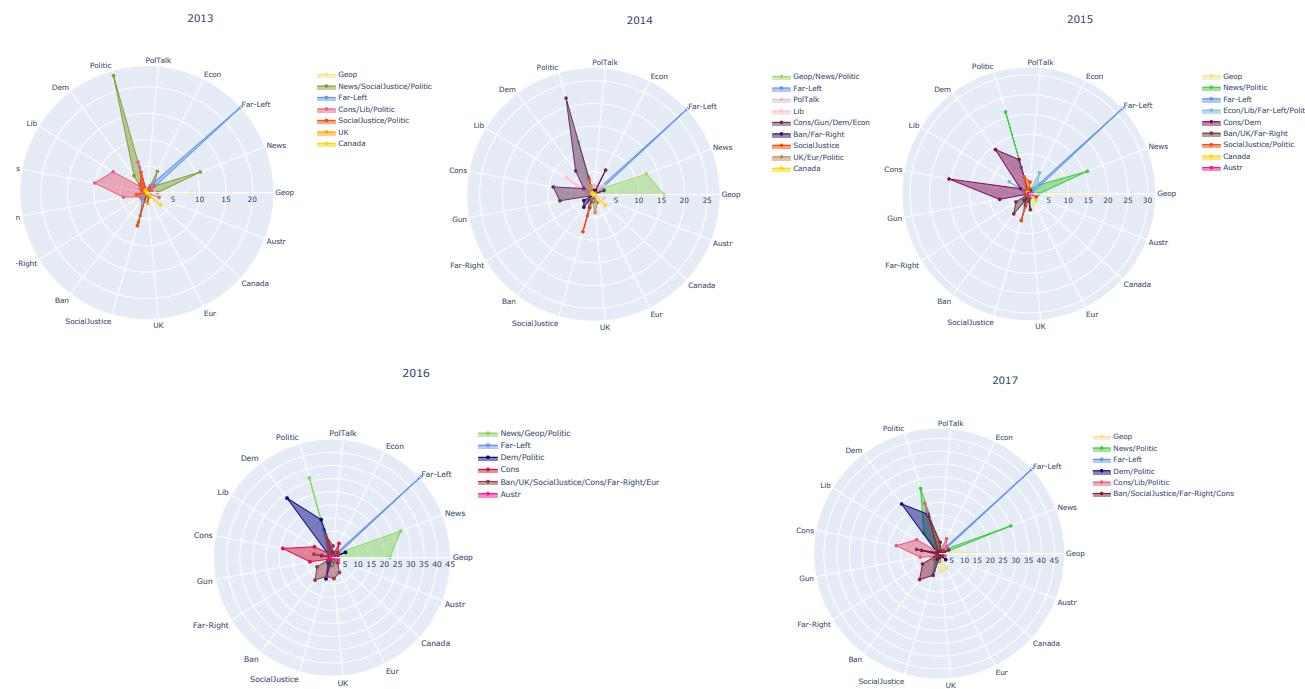


Figure 22. Radar plots of validated community structures from user-based subreddit networks (2013–2017). Each polygon corresponds to a community, with axes representing topical categories. The plots illustrate the degree of topical purity or heterogeneity across communities.

- Geop/News/Politic
- Far-Left
- PolTalk
- Lib
- Cons/Gun/Dem/Econ/Politic
- Ban/Far-Right
- SocialJustice
- UK/Eur/Politic
- Canada

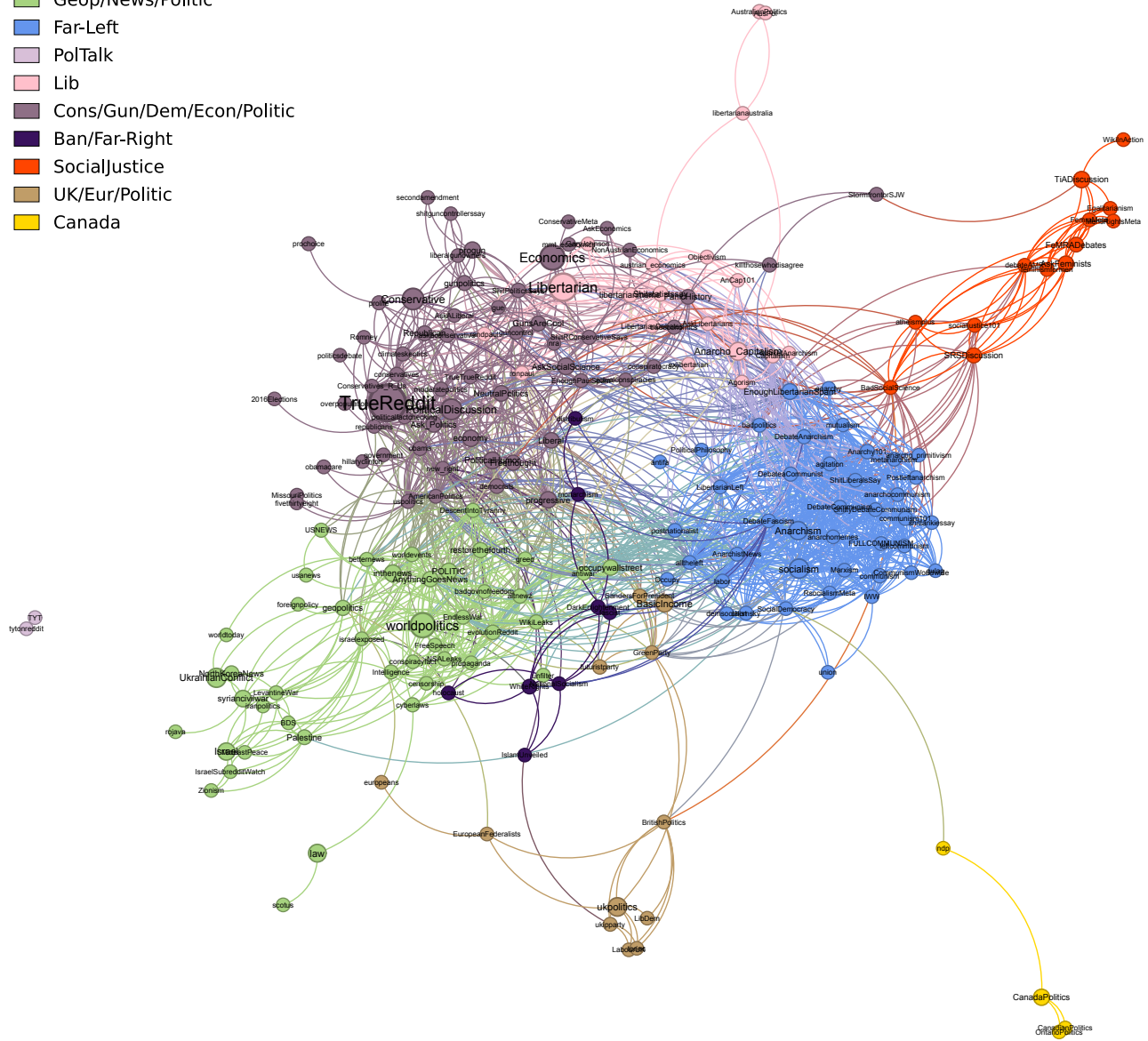


Figure 25. Validated community structure in the user-based subreddit network, 2014.

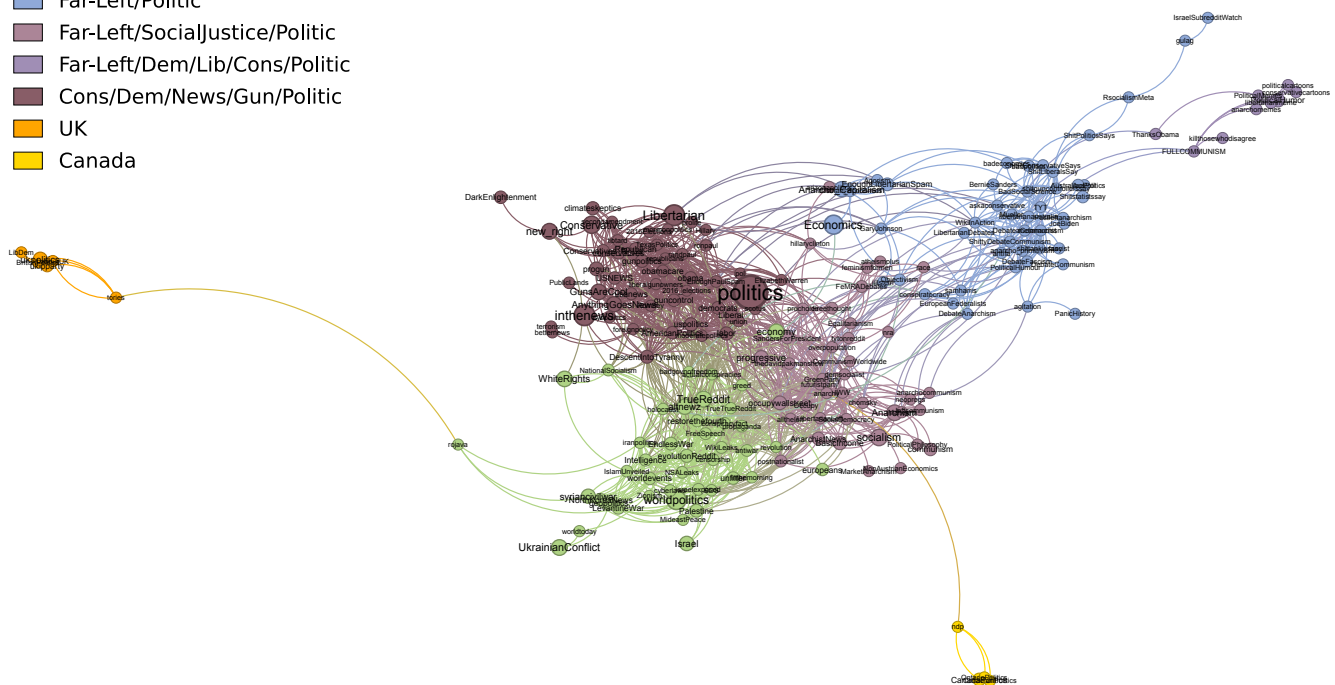


Figure 30. Validated community structure in the domain-based subreddit network, 2014.

- Geop/Politic
- Far-Left
- Far-Left/Politic
- Cons/Dem/Politic
- SocialJustice/News/Politic
- UK/Ban/News/Far-Right
- Austr

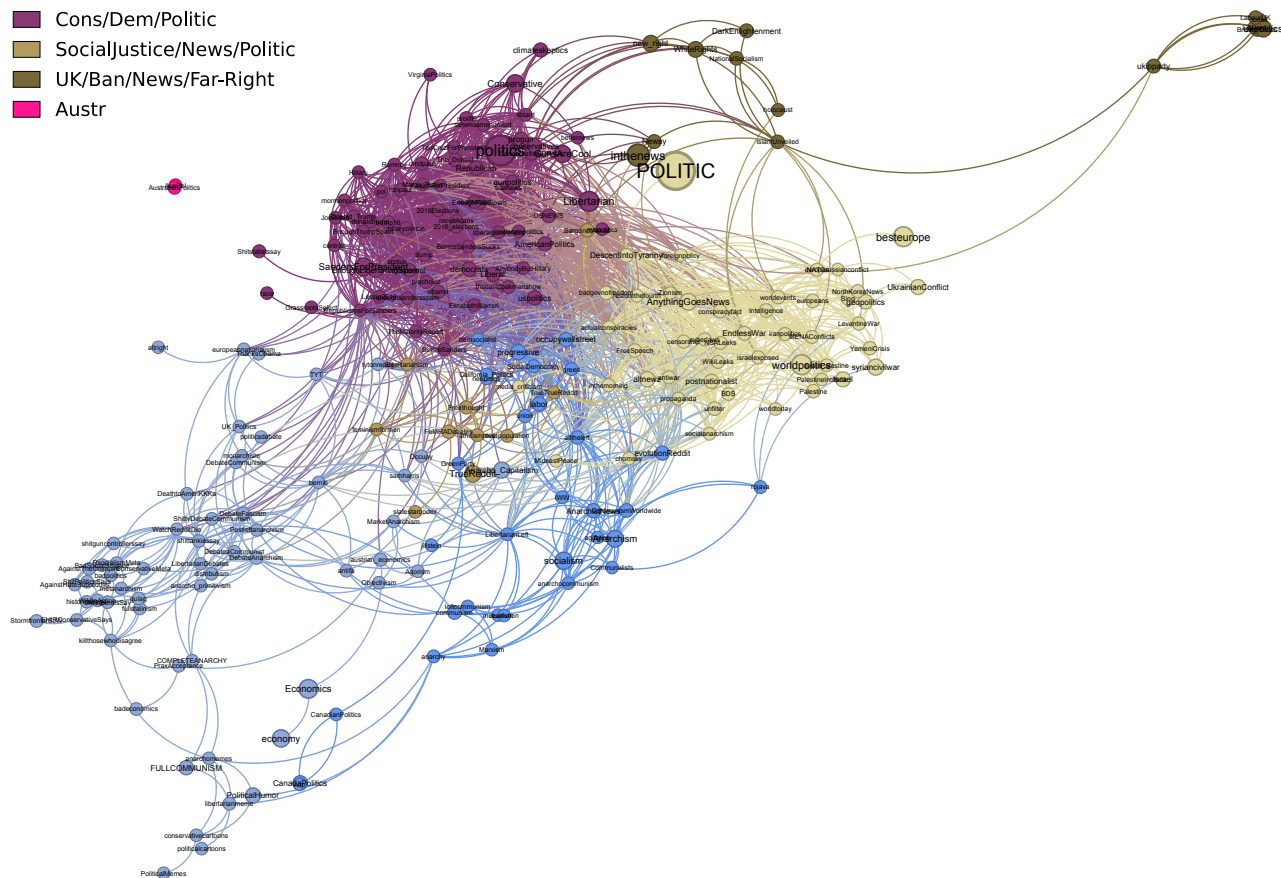


Figure 31. Validated community structure in the domain-based subreddit network, 2015.

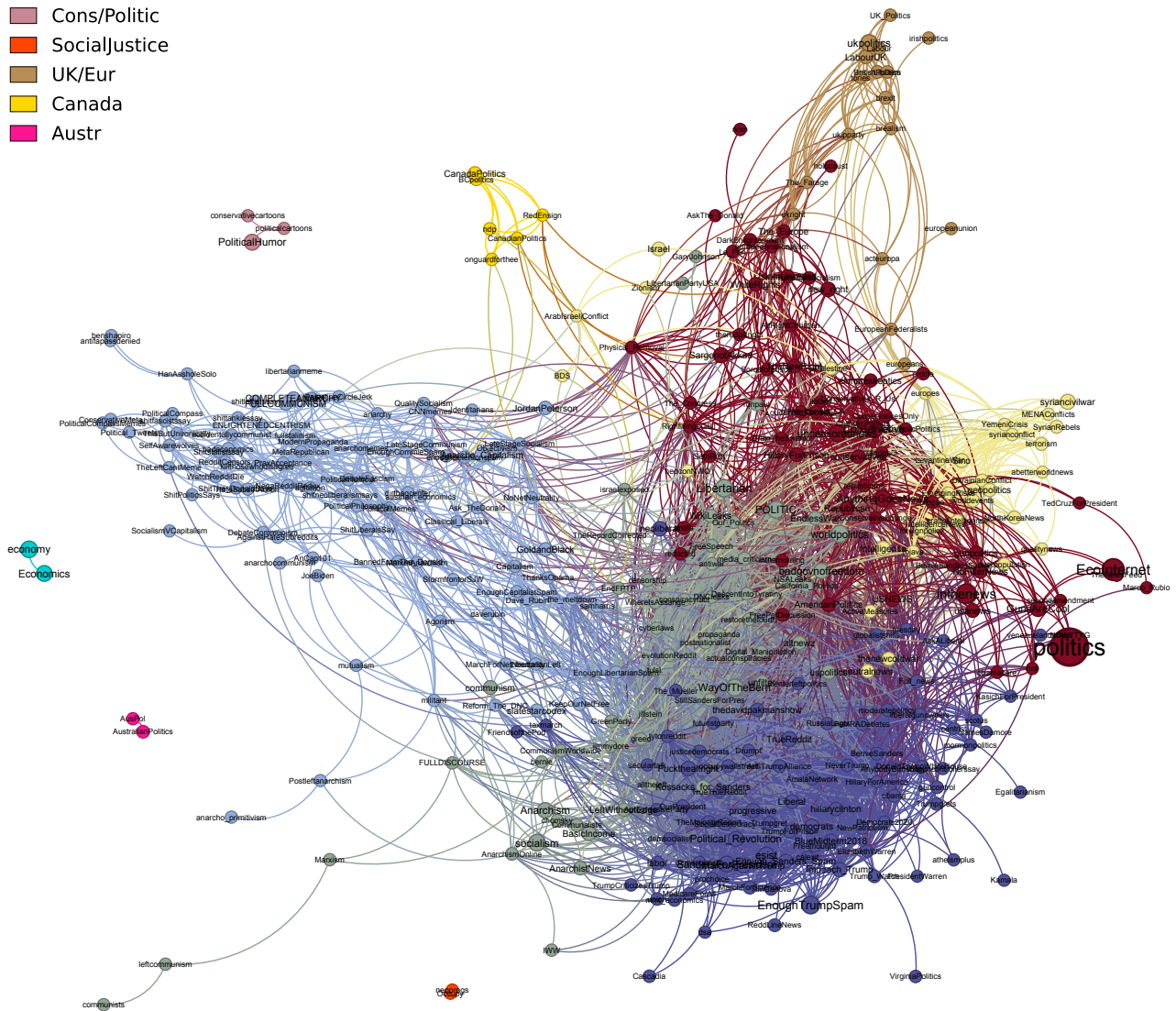
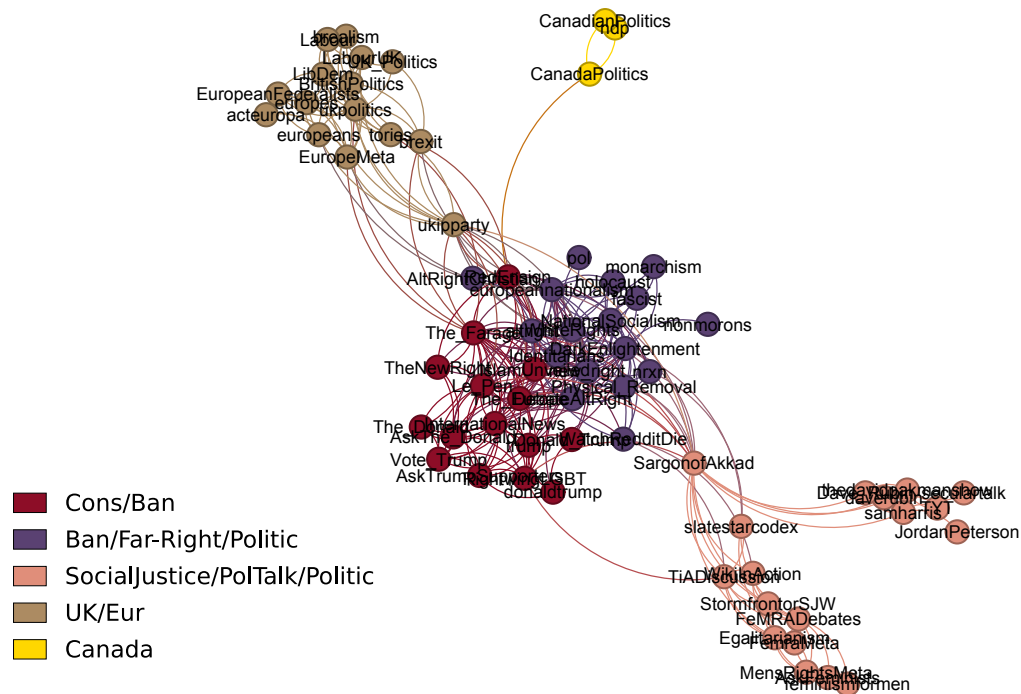
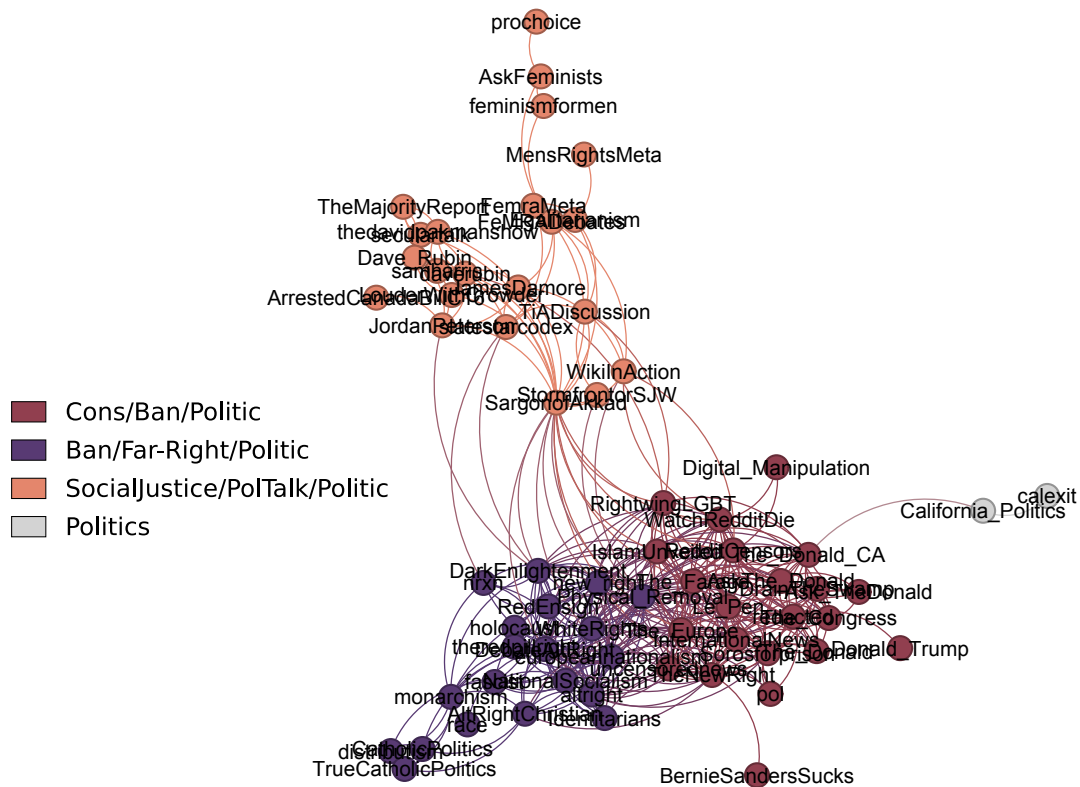


Figure 33. Validated community structure in the domain-based subreddit network, 2017.



(a) 2016.



(b) 2017.

Figure 34. Subcommunities of the mixed community (predominantly including banned subreddits) identified in the user-based subreddit network. The panels show the higher-resolution partitions obtained in (a) 2016 and (b) 2017.

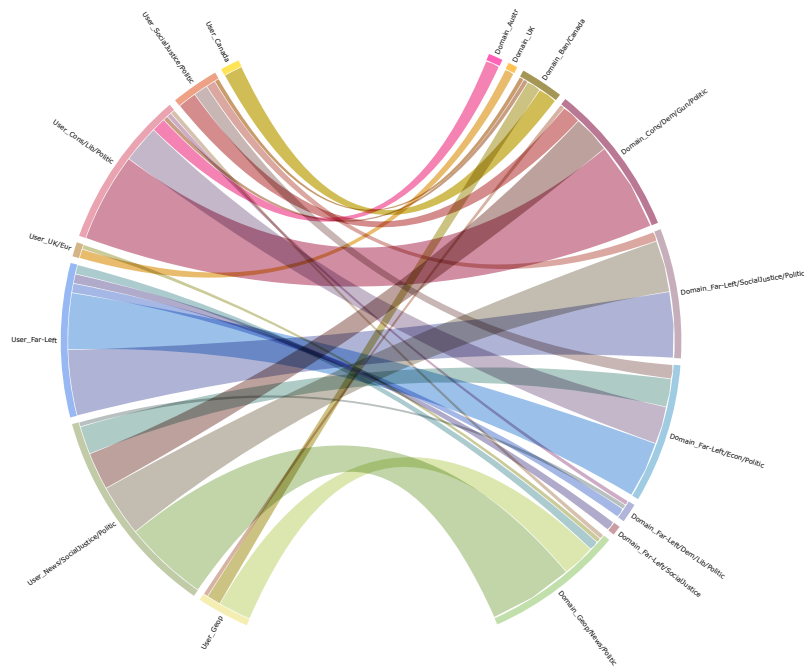


Figure 35. Chord diagram of community matches, 2013.

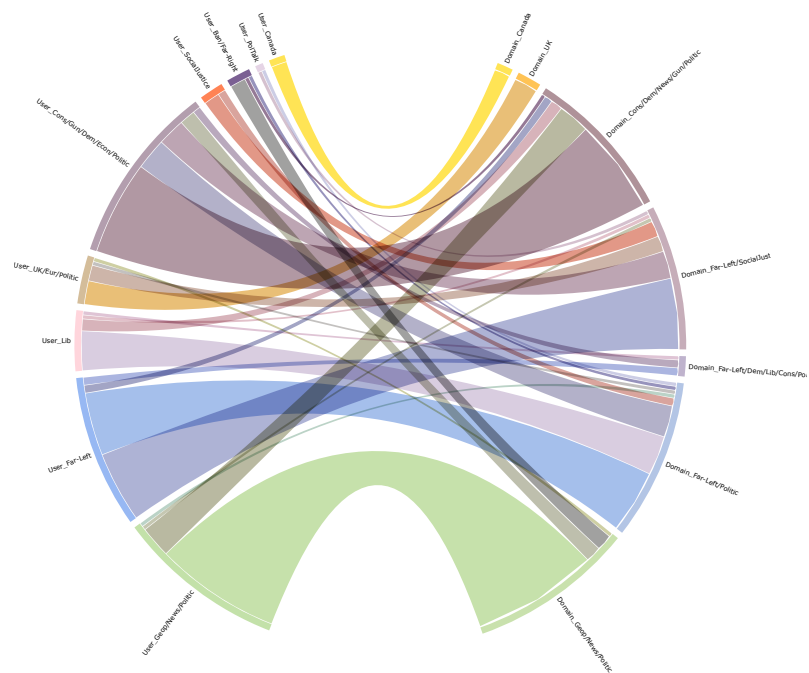


Figure 36. Chord diagram of community matches, 2014.

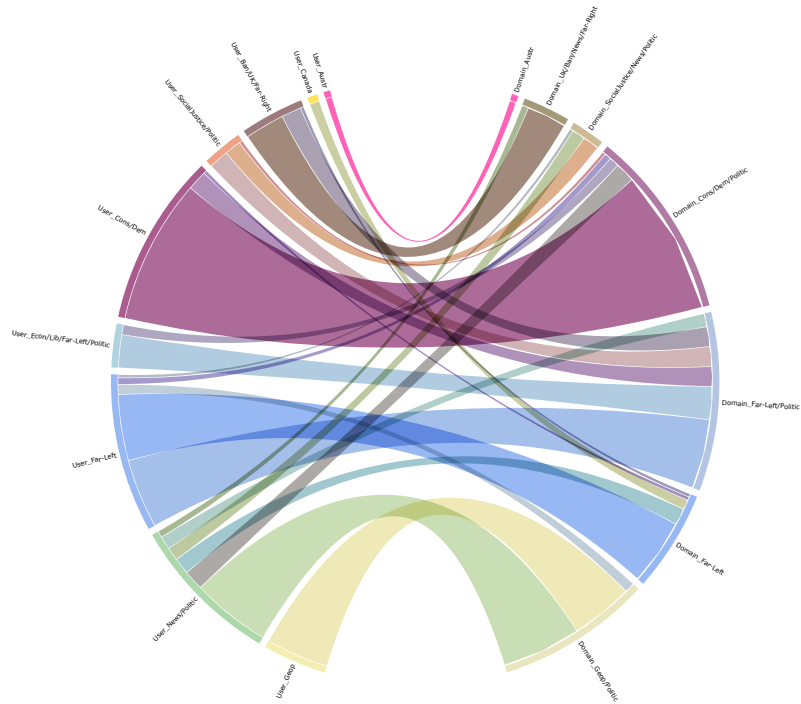


Figure 37. Chord diagram of community matches, 2015.

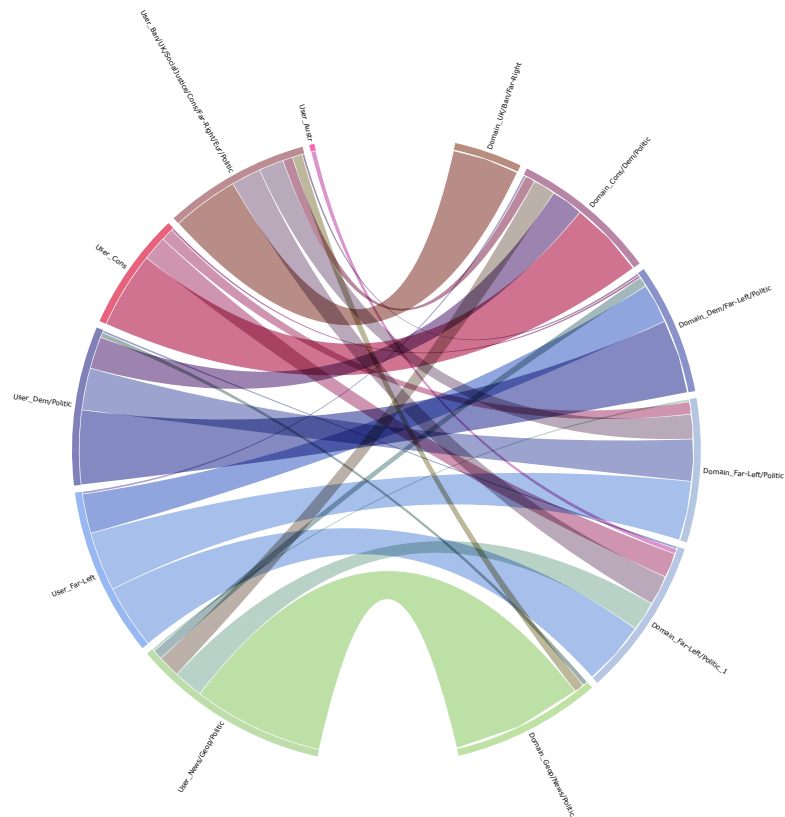


Figure 38. Chord diagram of community matches, 2016.

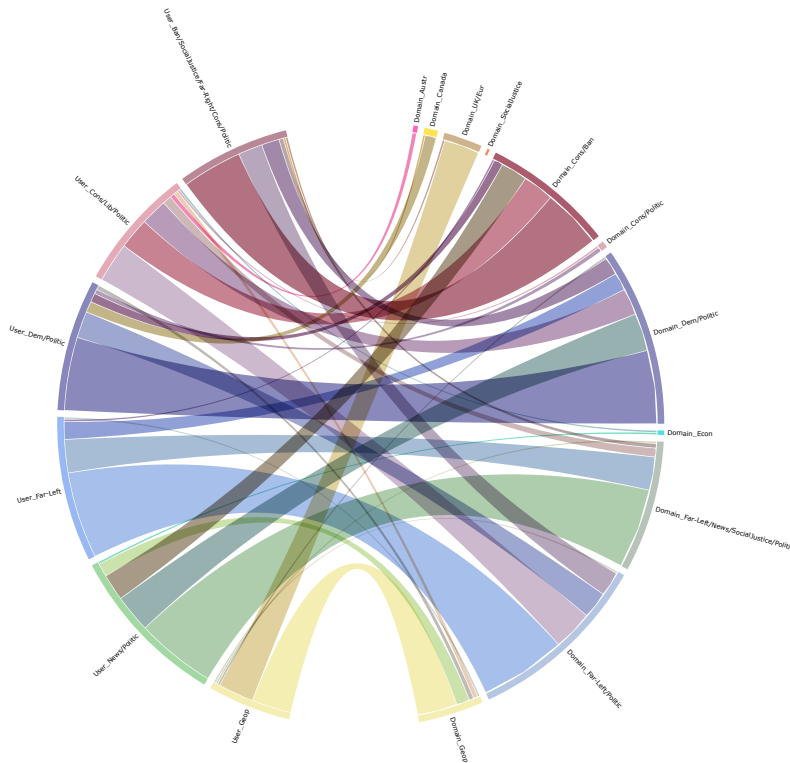


Figure 39. Chord diagram of community matches, 2017.

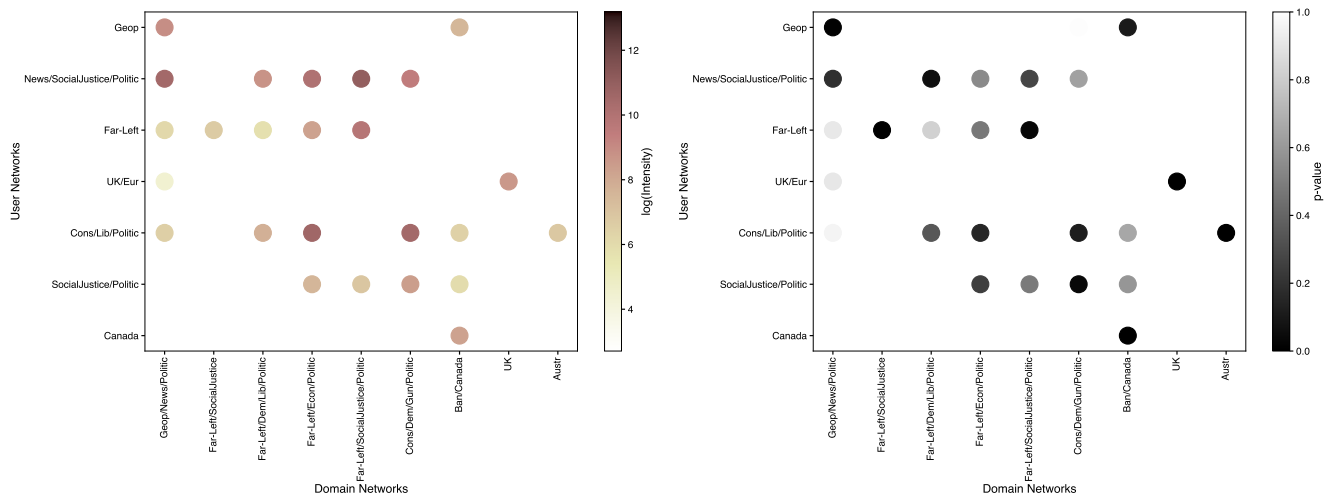


Figure 40. Echo chambers at the community level, 2013.

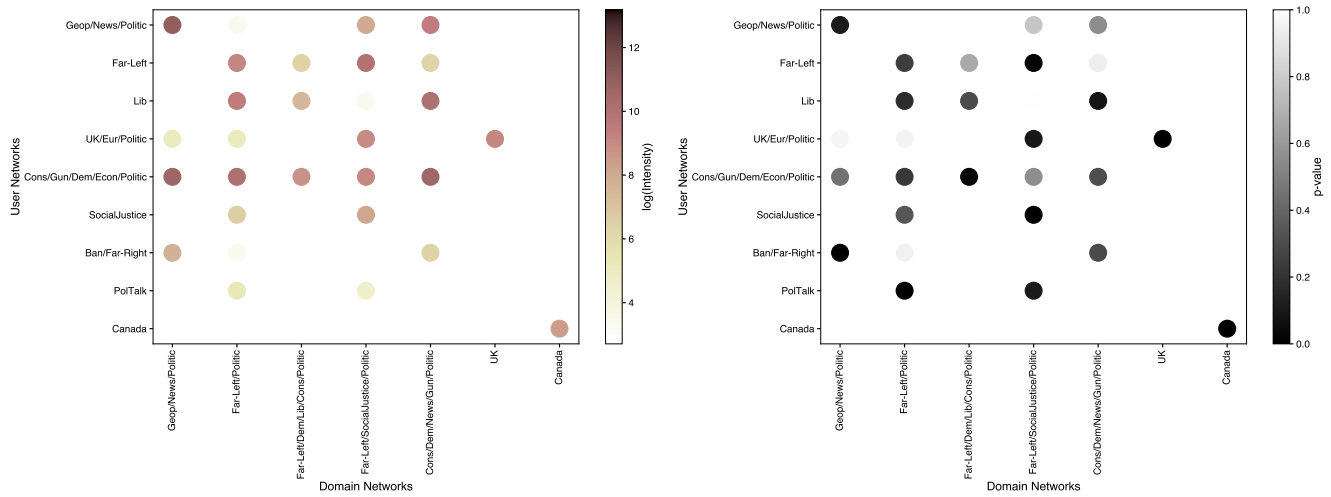


Figure 41. Echo chambers at the community level, 2014.

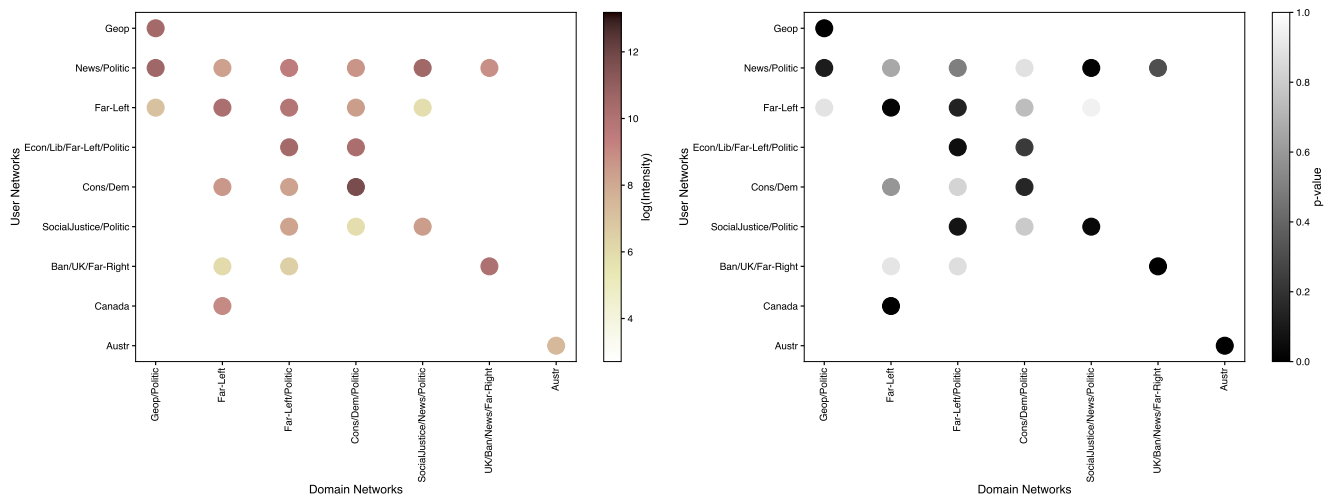


Figure 42. Echo chambers at the community level, 2015.

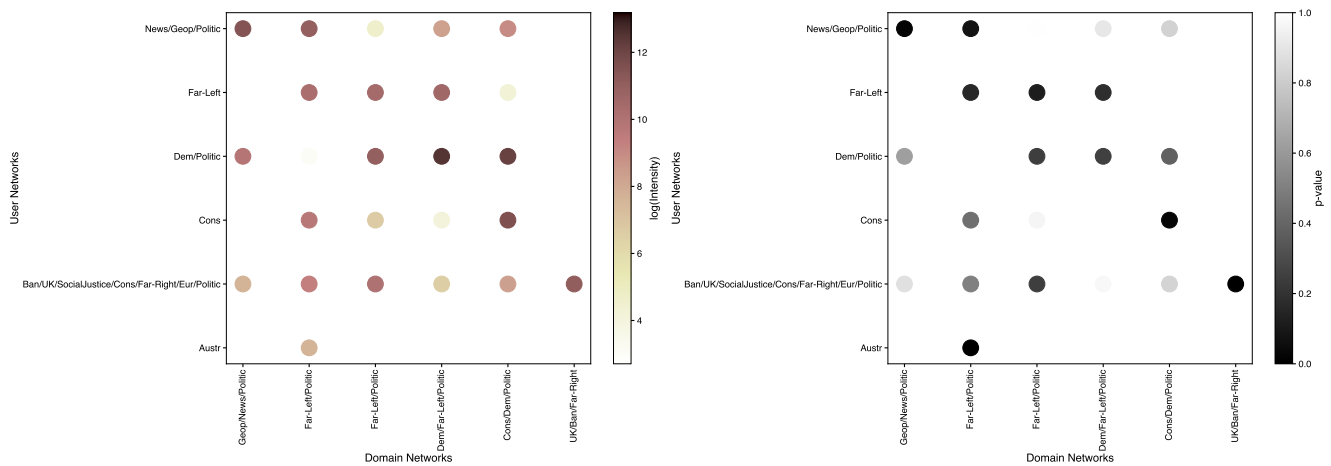


Figure 43. Echo chambers at the community level, 2016.

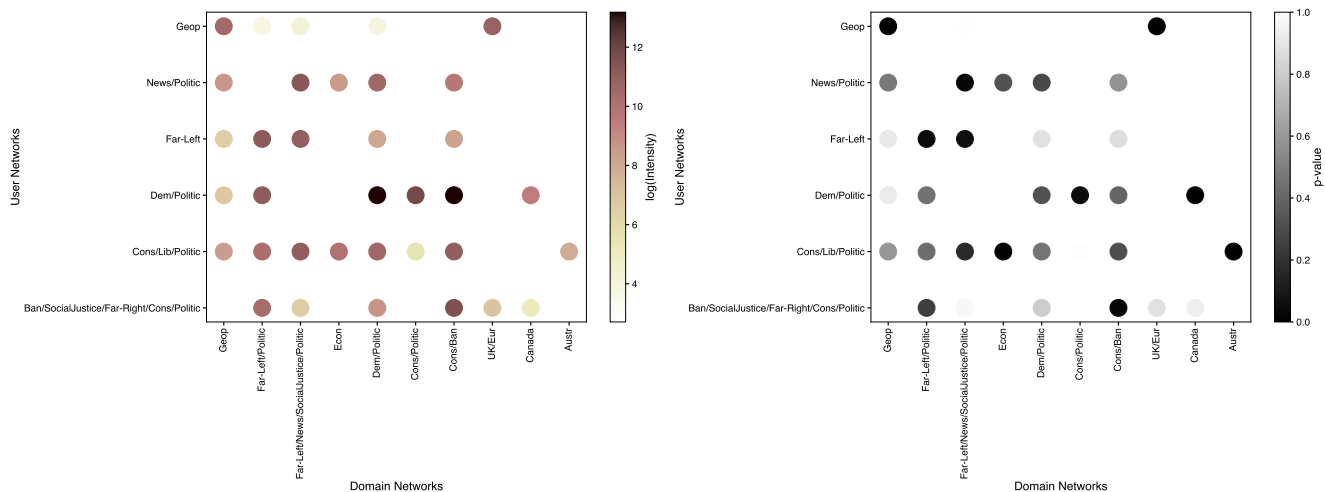


Figure 44. Echo chambers at the community level, 2017.

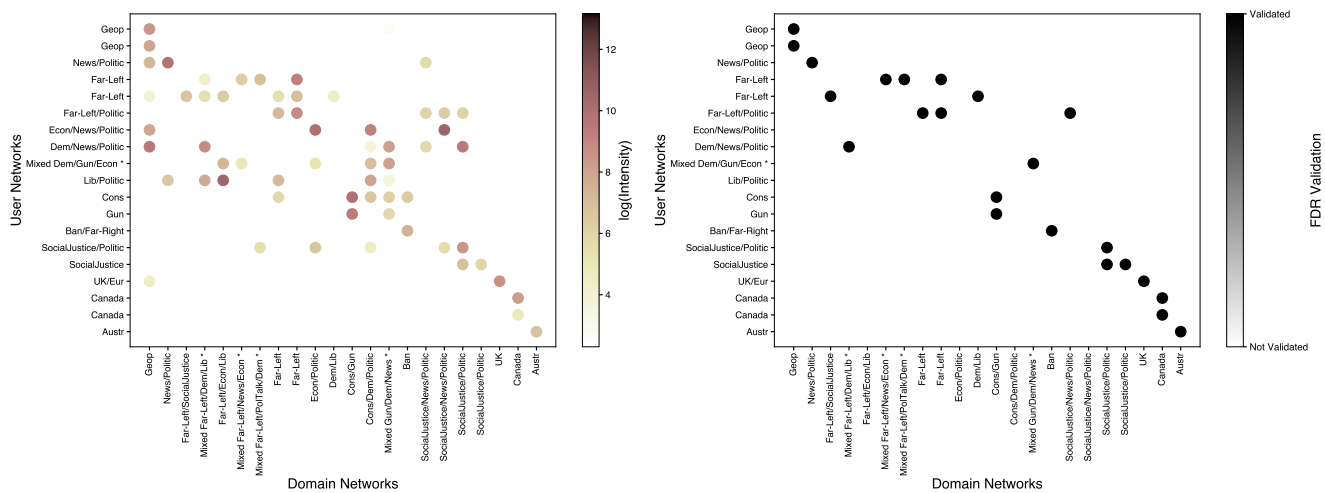


Figure 45. Echo chambers at the sub-community level, 2013.

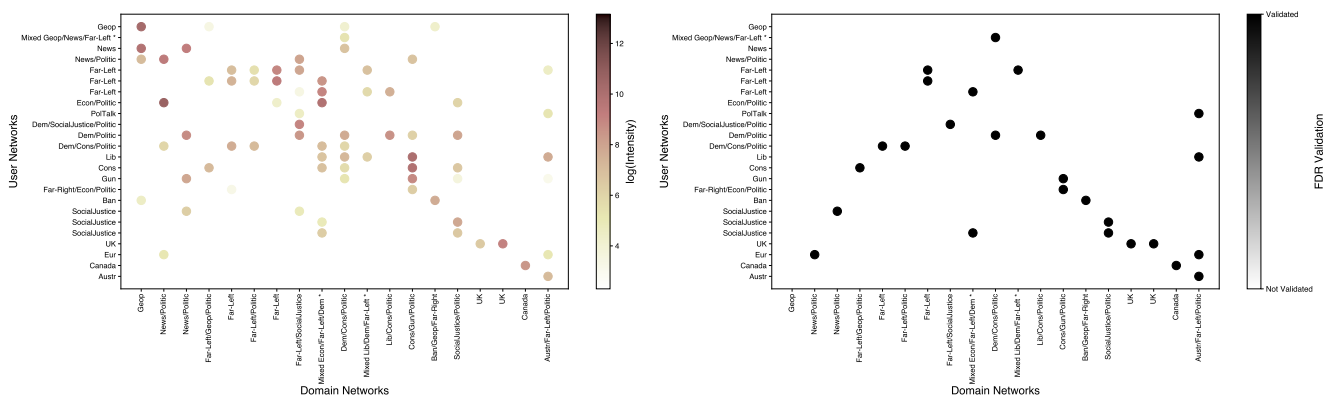


Figure 46. Echo chambers at the sub-community level, 2014.

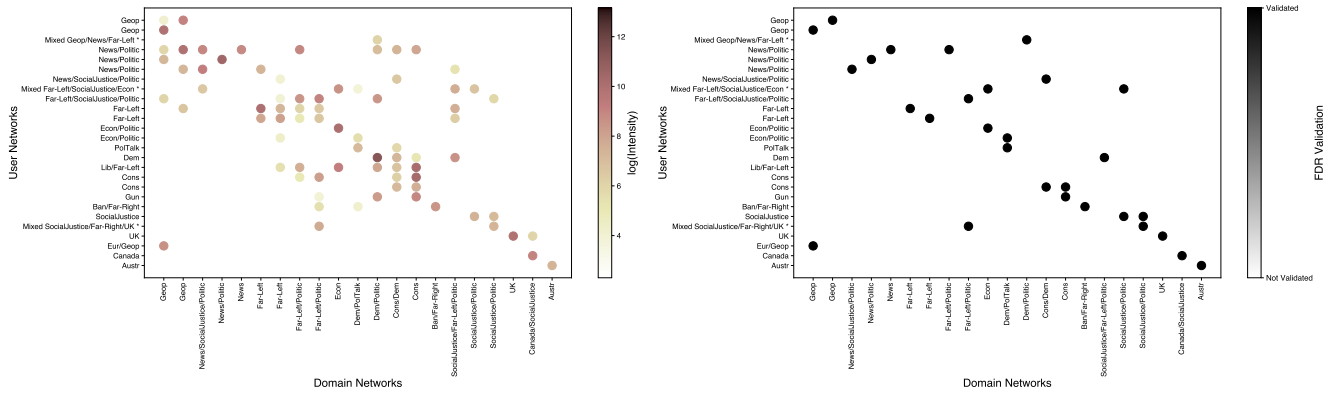


Figure 47. Echo chambers at the sub-community level, 2015.

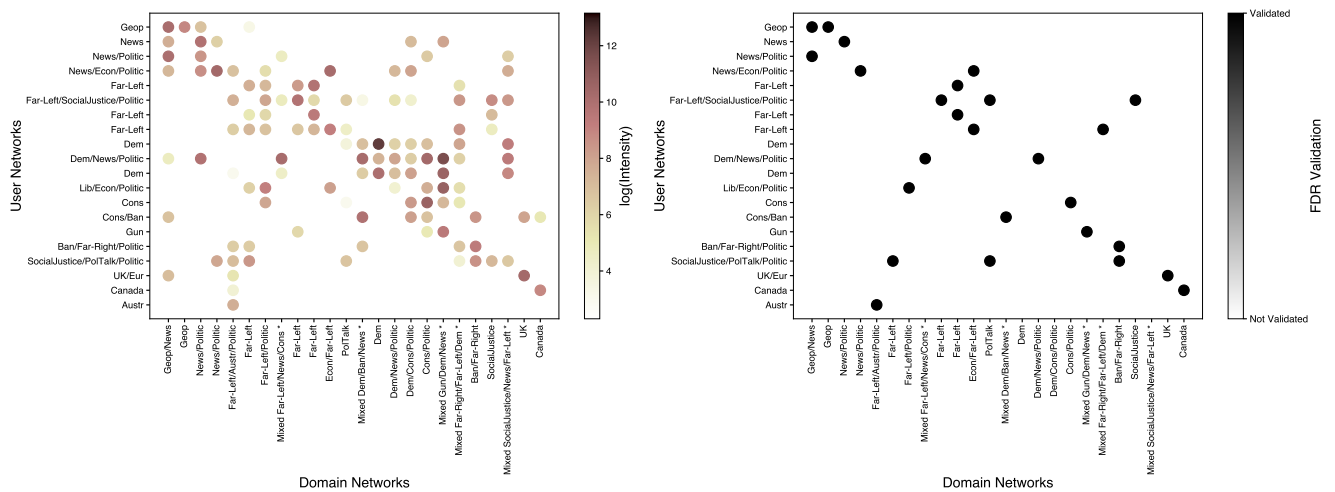


Figure 48. Echo chambers at the sub-community level, 2016.

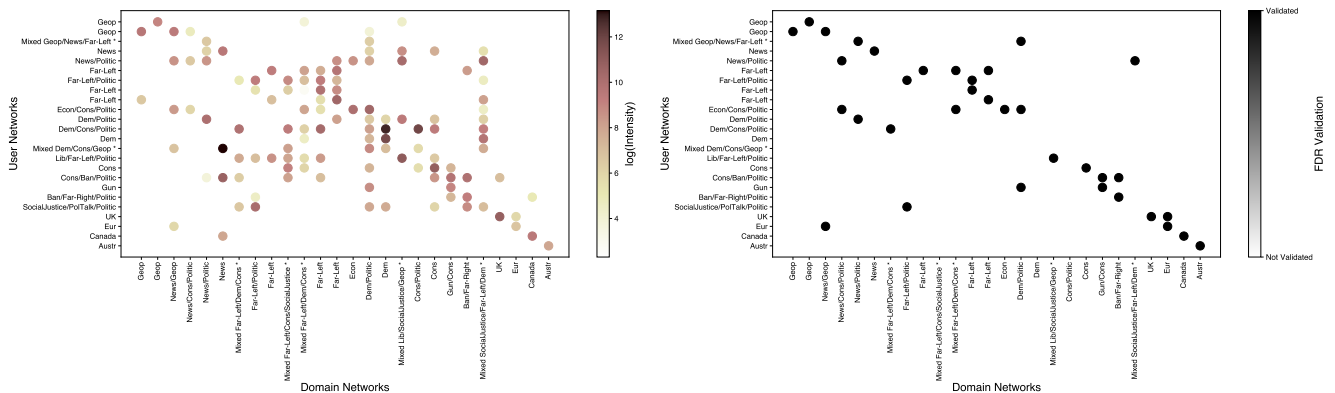
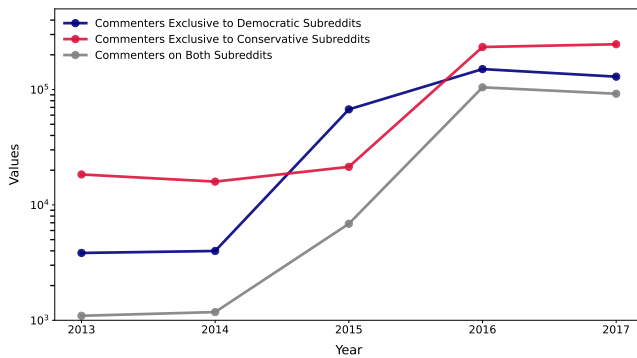
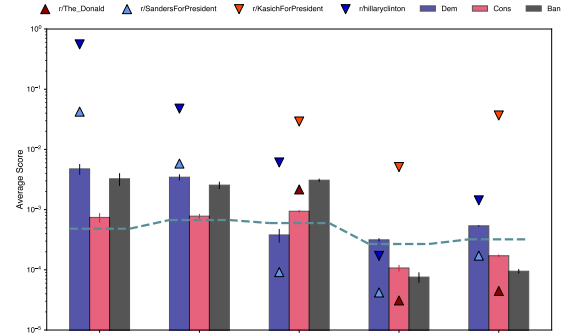


Figure 49. Echo chambers at the sub-community level, 2017.



(a) Number of commenters exclusive to Democratic subreddits, Conservative subreddits, or active in both.



(b) Average score of comments in Democratic, Conservative, and to-be-banned subreddits, with yearly averages and values for selected presidential candidates' subreddits.

Figure 50. User activity and scoring across partisan subreddits. Panel (a) reports the number of exclusive commenters by political alignment, showing increasing concentration within partisan communities. Panel (b) presents average comment scores across Democratic, Conservative, and Banned subreddits, with yearly trends and values for candidate-specific forums (Clinton, Sanders, Kasich, Trump). Results indicate a strong decline in average scores relative to 2013 baselines, especially within candidate subreddits.

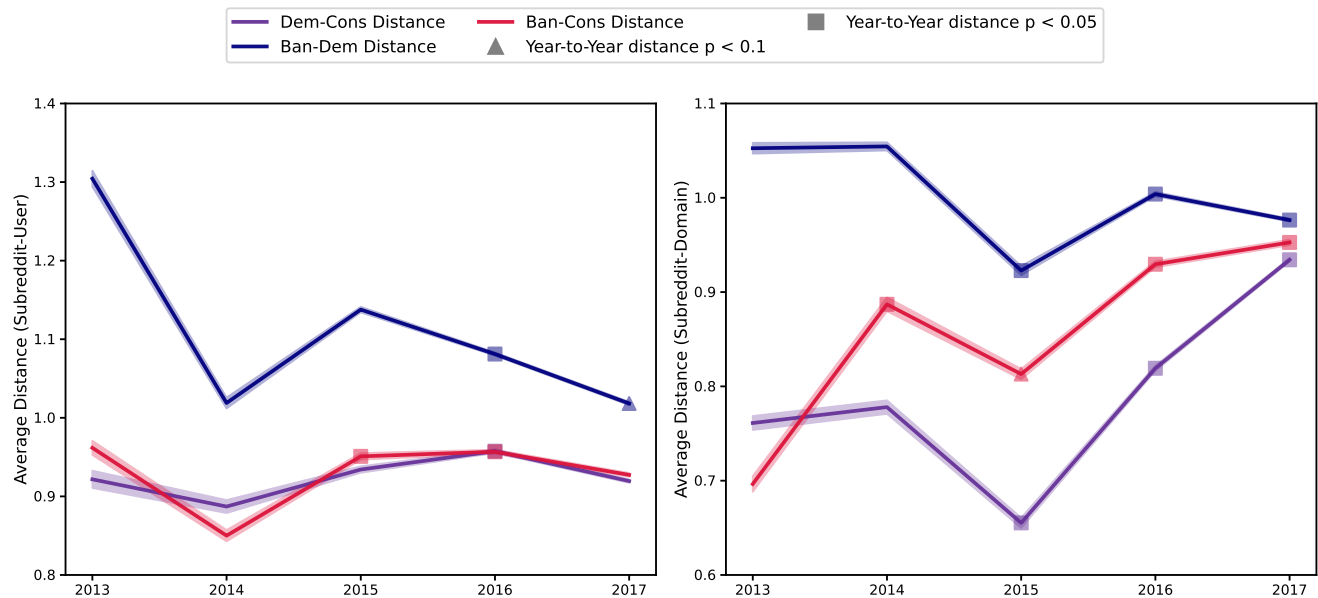


Figure 51. Average cosine distances between Democratic, Conservative, and Banned subreddits in the statistically validated projections. The left panel reports distances computed at the user level, while the right panel reports distances at the domain level. Shaded areas denote standard errors. Significance markers indicate the results of Kolmogorov–Smirnov tests, showing when distance distributions significantly diverge from those of the previous year, with marked effects during electoral periods.

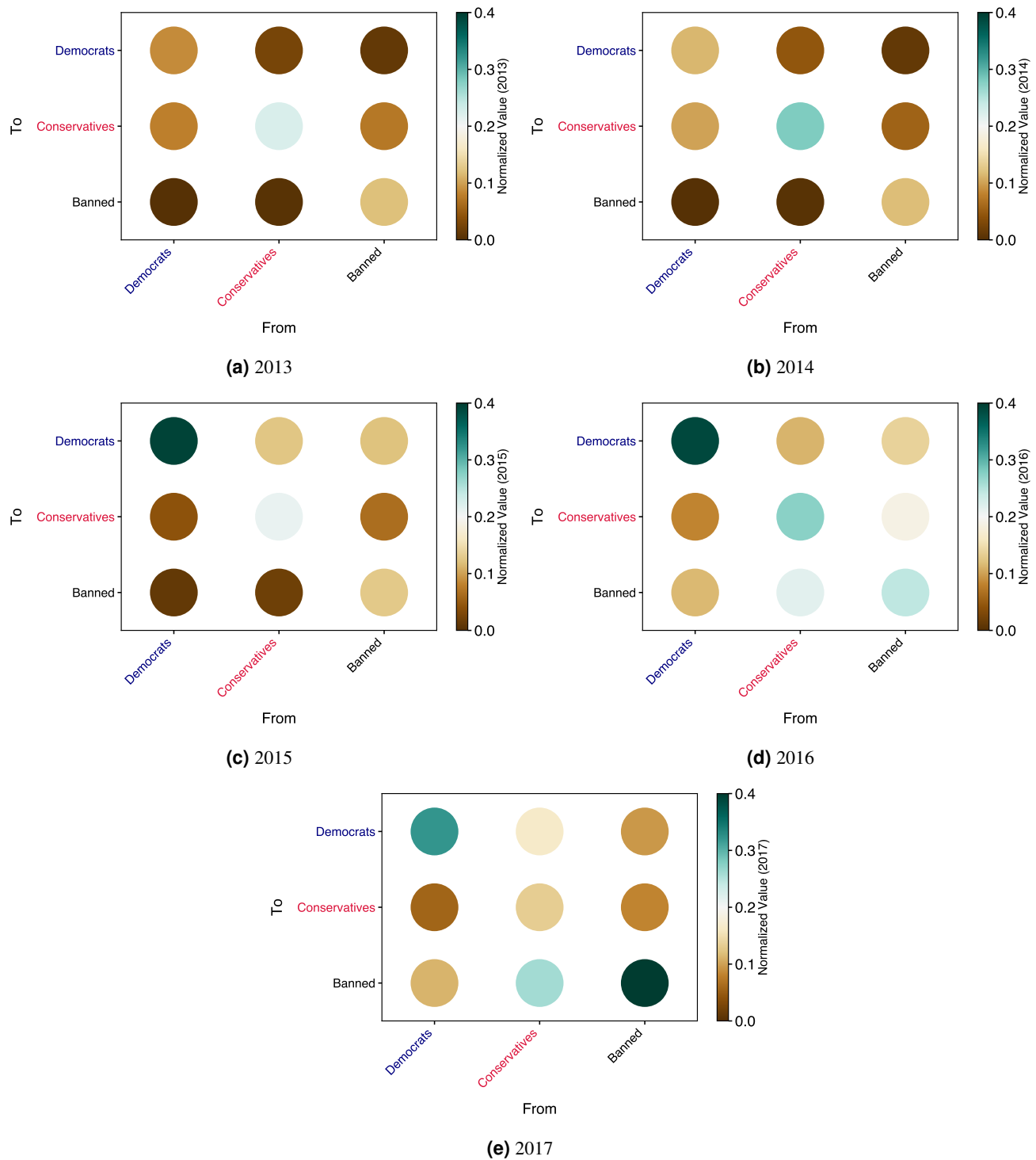


Figure 52. Normalized comment matrices from 2013 to 2017. Each panel shows the percentage of comments exchanged between factions, considering users labeled as Democrats, Conservatives, or Banned. Matrices are row-normalized so that each row represents the proportion of comments written by one faction and directed toward the others in the same year.

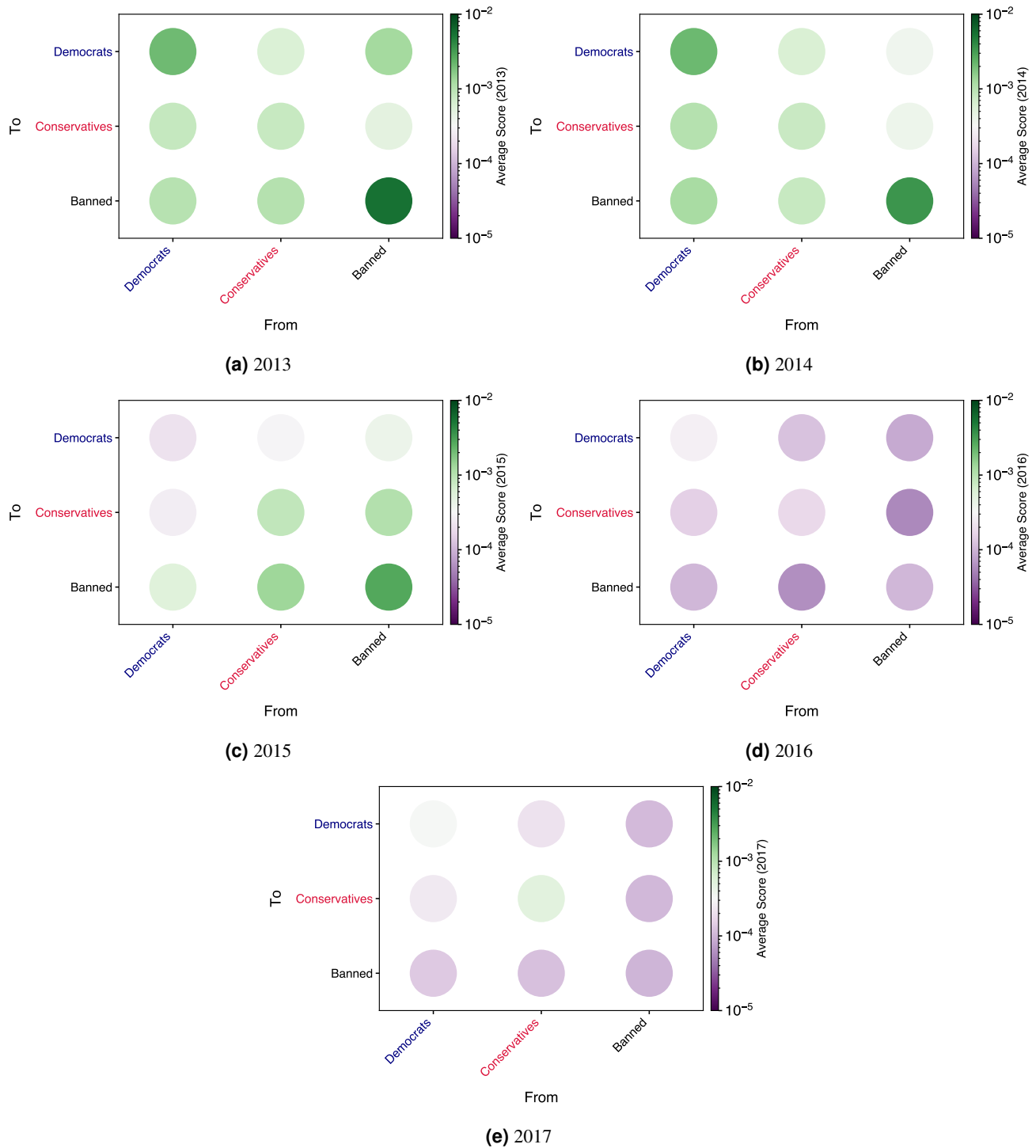


Figure 53. Average score heatmaps from 2013 to 2017. Each panel shows the mean score of comments exchanged between factions, considering users labeled as Democrats, Conservatives, or Banned. Matrices are row-wise, so that each row represents the average reception of comments produced by one faction and directed toward the others in the same year.

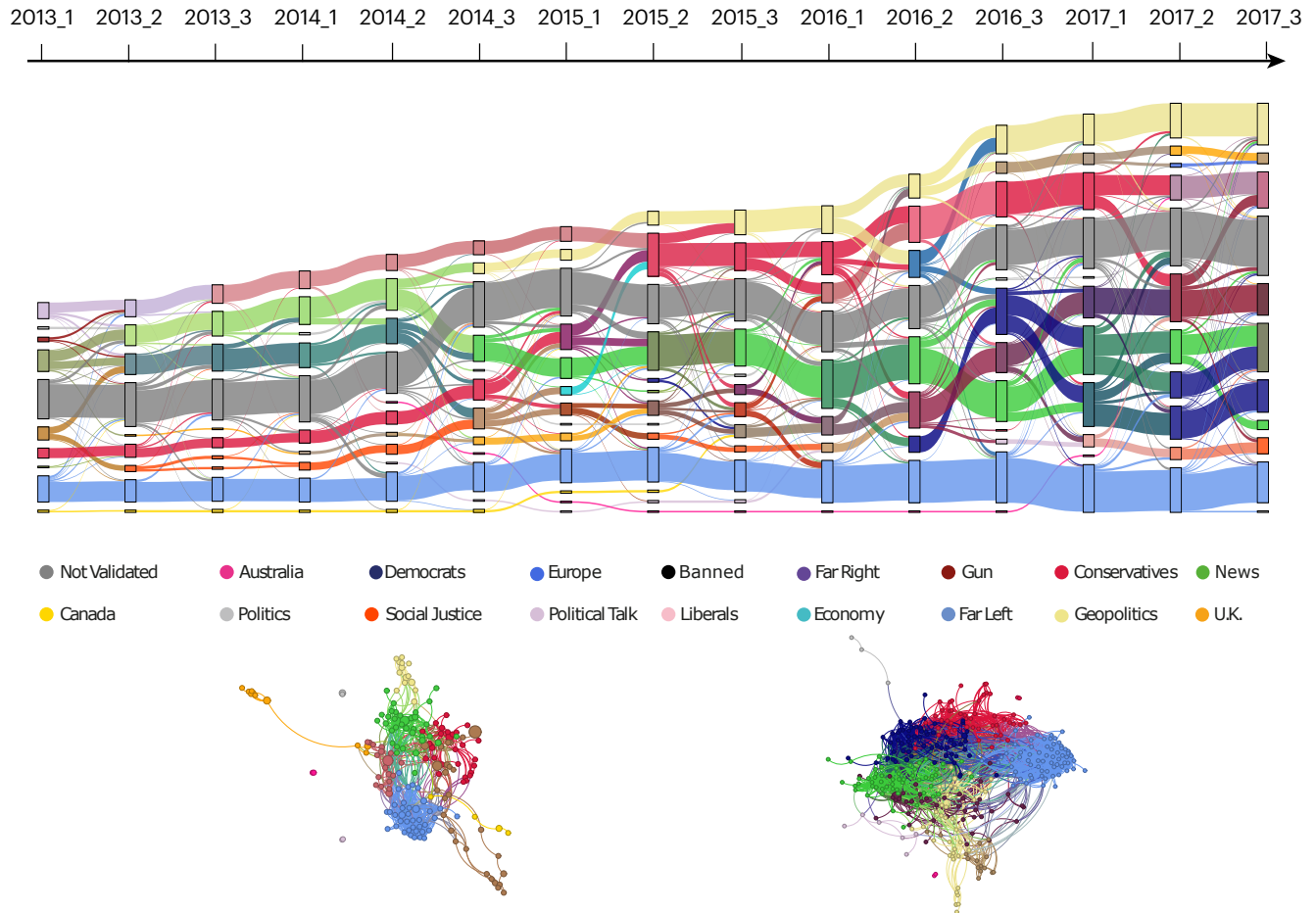


Figure 54. Temporal community flows in interaction-based networks. Flow diagrams illustrate the evolution of statistically validated subreddit communities at four-month resolution (y_1, y_2, y_3). Arrows indicate transitions of subreddits between communities across consecutive windows. Overall patterns broadly replicate the annual analysis.

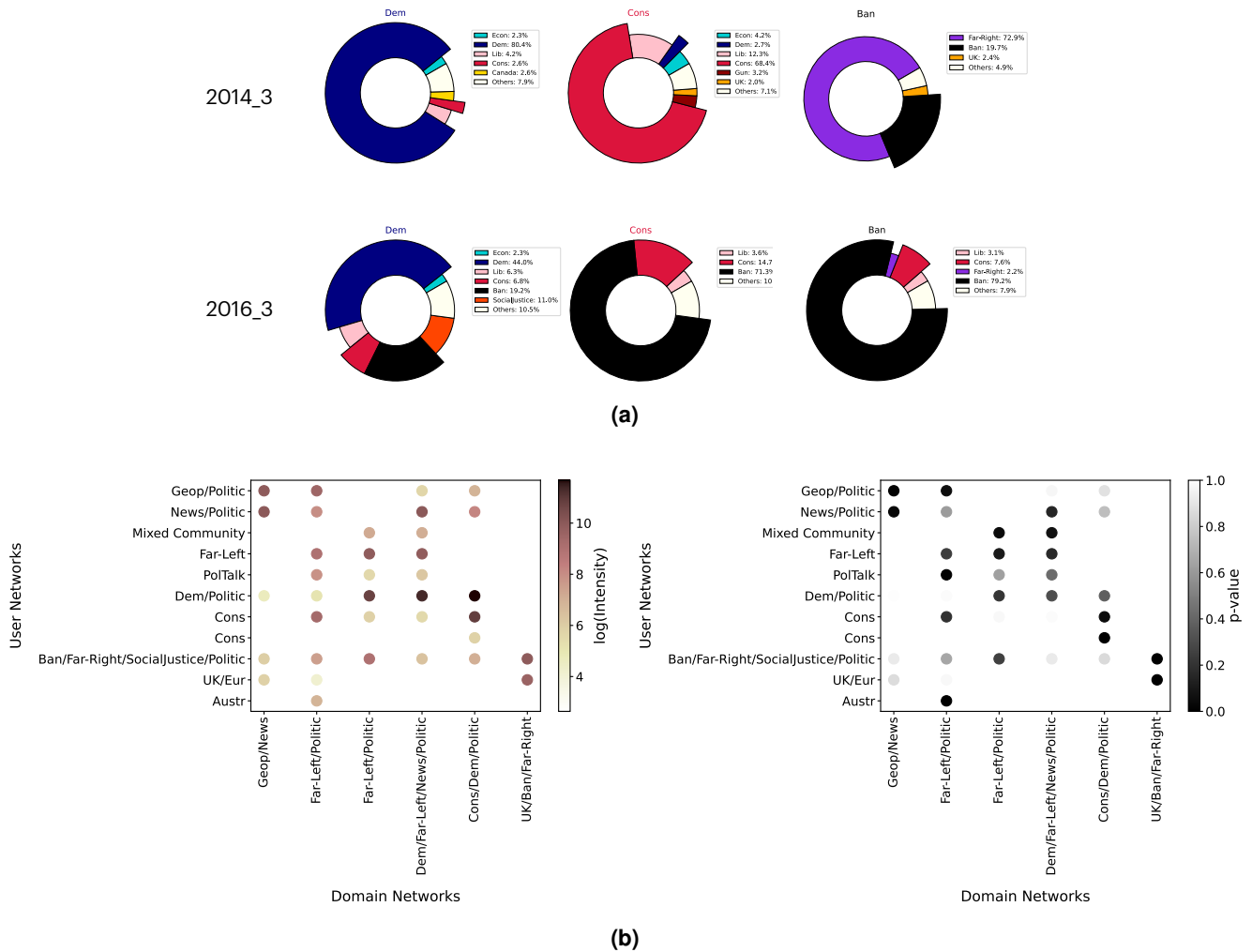
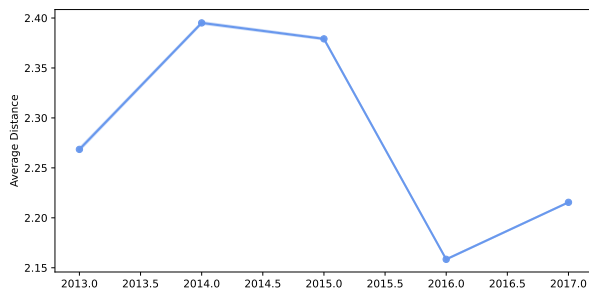
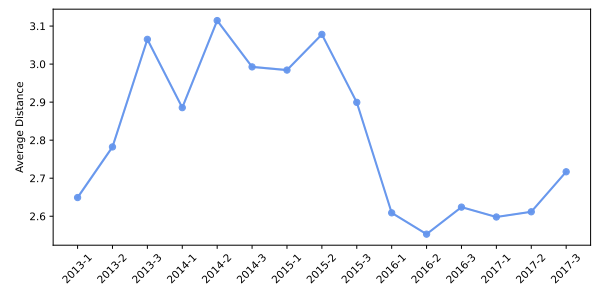


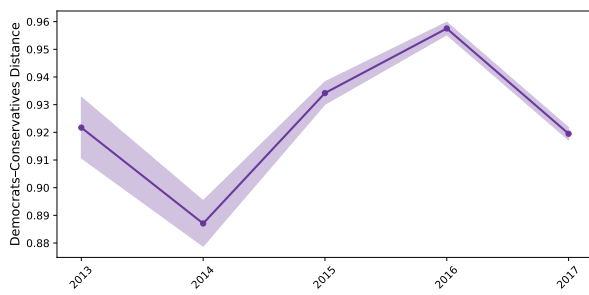
Figure 55. Polarization (a) and echo-chamber (b) analysis at finer resolution. (a) Donut charts showing user composition (Democratic, Conservative, Banned) from polarization matrices in the 2014_3 and 2016_3 windows. The 2016 window shows reduced overall polarization, yet a sharper alignment of banned users with Conservative subreddits. (b) Echo-chamber analysis for 2016_3, based on a weighted bipartite projection between subreddit communities from validated user-subreddit and subreddit-domain networks. Edge weights represent user overlap; BiWCM validation highlights significant alignments. Patterns are consistent with those observed in the annual analysis.



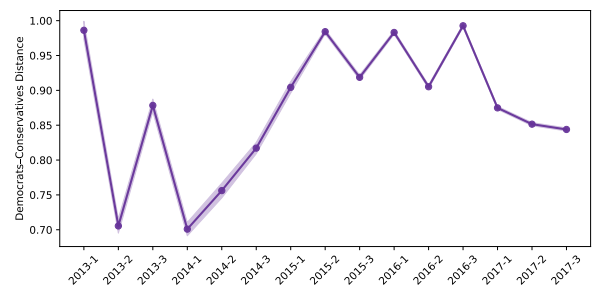
(a)



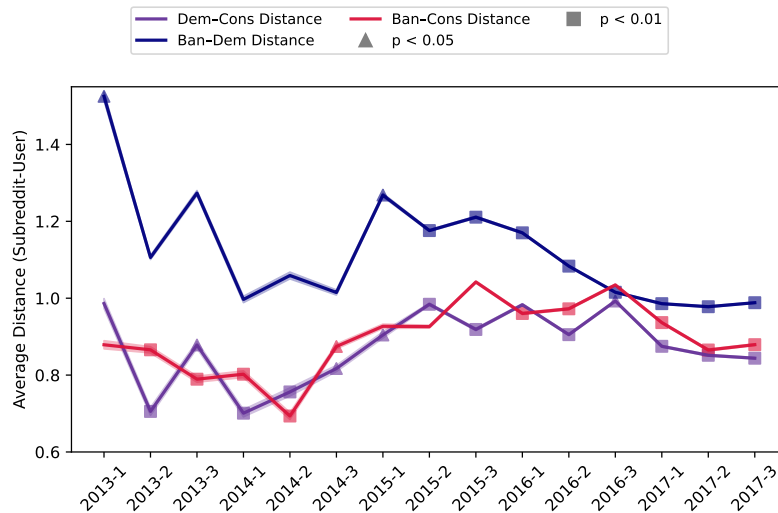
(b)



(c)



(d)



(e)

Figure 56. Inter-group distances in interaction-based networks. Mean distances across validated networks confirm the annual-level findings. Panels compare: (a) annual vs. (b) four-month averages, (c–d) distances between Democratic and Conservative communities, and (e) all pairwise distances (Democratic, Conservative, Banned). As in the annual analysis, Democratic–Banned distances remain the largest, decreasing slightly before elections, while Conservative and Banned communities remain close. Democratic–Conservative distances increase during electoral periods. Statistical tests highlight significant differences in selected windows, including the 2014 midterms (2014_3) and the 2016 presidential elections (2016_3).

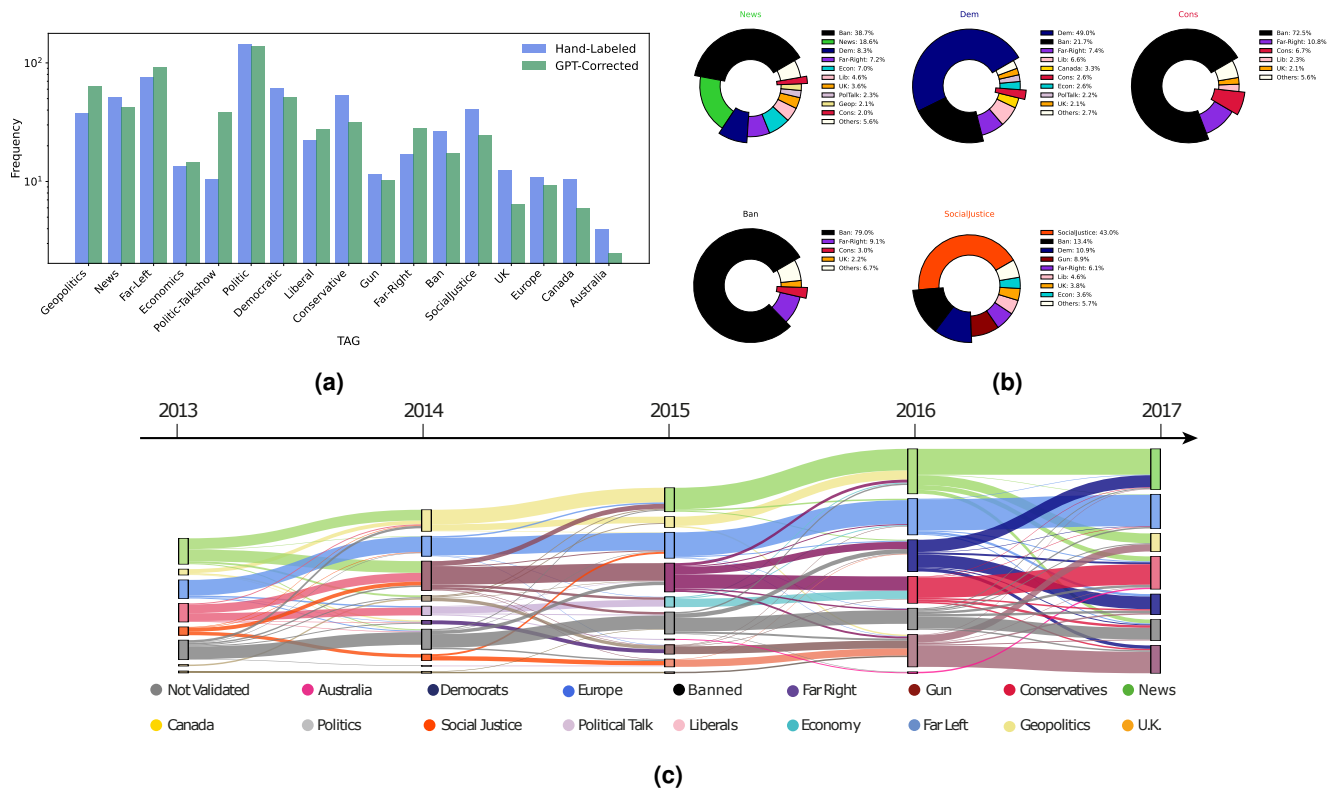


Figure 57. Comparison of GPT-corrected classification and community analysis. In panel (a) the distribution of hand-labeled and GPT-corrected tags is compared, (b) shows donut charts of 2016 polarization illustrating the composition of selected topic communities at the user level within validated annual subreddit networks, and (c) presents flowcharts tracing the yearly formation of subreddit communities within the same networks.

References

1. SimilarWeb Ltd. Similarweb. <https://www.similarweb.com> (2025). Accessed: 2025-07-08.
2. Hofmann, V., Schütze, H. & Pierrehumbert, J. B. The reddit politosphere: A large-scale text and network resource of online political discourse. *Proc. Int. AAAI Conf. on Web Soc. Media* **16**, 1259–1267, DOI: [10.1609/icwsm.v16i1.19377](https://doi.org/10.1609/icwsm.v16i1.19377) (2022).
3. Baumgartner, J., Zannettou, S., Keegan, B., Squire, M. & Blackburn, J. The pushshift reddit dataset. <https://arxiv.org/abs/2001.08435> (2020). [2001.08435](https://arxiv.org/abs/2001.08435).
4. Balassa, B. Trade liberalization and revealed comparative advantage. *The Manch. Sch. Econ. Soc. Stud.* **33**, 99–123, DOI: [10.1111/j.1467-9957.1965.tb00050.x](https://doi.org/10.1111/j.1467-9957.1965.tb00050.x) (1965).
5. Saracco, F. *et al.* Inferring monopartite projections of bipartite networks: an entropy-based approach. *New J. Phys.* **19**, 053022, DOI: [10.1088/1367-2630/aa6b38](https://doi.org/10.1088/1367-2630/aa6b38) (2017).
6. Vallarano, N. *et al.* Fast and scalable likelihood maximization for exponential random graph models with local constraints. *Sci. Reports* **11**, 15227, DOI: [10.1038/s41598-021-93830-4](https://doi.org/10.1038/s41598-021-93830-4) (2021).
7. Benjamini, Y. & Hochberg, Y. Controlling the false discovery rate: A practical and powerful approach to multiple testing. *J. Royal Stat. Soc. Ser. B (Methodological)* **57**, 289–300 (1995). Full publication date: 1995.
8. Shannon, C. E. A mathematical theory of communication. *The Bell Syst. Tech. J.* **27**, 379–423, DOI: [10.1002/j.1538-7305.1948.tb01338.x](https://doi.org/10.1002/j.1538-7305.1948.tb01338.x) (1948).
9. Cimini, G., Carra, A., Didomenicantonio, L. & Zaccaria, A. Meta-validation of bipartite network projections. *Commun. Phys.* **5**, 76, DOI: [10.1038/s42005-022-00856-9](https://doi.org/10.1038/s42005-022-00856-9) (2022).
10. Vinh, N. X., Epps, J. & Bailey, J. Information theoretic measures for clusterings comparison: Variants, properties, normalization and correction for chance. *J. Mach. Learn. Res.* **11**, 2837–2854 (2010).
11. Meilă, M. Comparing clusterings—an information based distance. *J. Multivar. Analysis* **98**, 873–895, DOI: <https://doi.org/10.1016/j.jmva.2006.11.013> (2007).
12. Blondel, V. D., Guillaume, J.-L., Lambiotte, R. & Lefebvre, E. Fast unfolding of communities in large networks. *J. Stat. Mech. Theory Exp.* **2008**, P10008, DOI: [10.1088/1742-5468/2008/10/P10008](https://doi.org/10.1088/1742-5468/2008/10/P10008) (2008).
13. Rosvall, M. & Bergstrom, C. T. Maps of random walks on complex networks reveal community structure. *Proc. Natl. Acad. Sci.* **105**, 1118–1123 (2008).
14. Rosvall, M., Axelsson, D. & Bergstrom, C. T. The map equation. *The Eur. Phys. J. Special Top.* **178**, 13–23, DOI: [10.1140/epjst/e2010-01179-1](https://doi.org/10.1140/epjst/e2010-01179-1) (2009).
15. Peixoto, T. P. The graph-tool python library (2014). <https://graph-tool.skewed.de>.
16. Peixoto, T. P. Hierarchical block structures and high-resolution model selection in large networks. *Phys. Rev. X* **4**, 011047 (2014).
17. Virtanen, P. *et al.* Scipy 1.0: fundamental algorithms for scientific computing in python. *Nat. Methods* **17**, 261–272, DOI: [10.1038/s41592-019-0686-2](https://doi.org/10.1038/s41592-019-0686-2) (2020).
18. Buffa, L. *et al.* Maximum entropy modeling of optimal transport: the sub-optimality regime and the transition from dense to sparse networks (2025). [2504.10444](https://arxiv.org/abs/2504.10444).
19. OpenAI. Gpt-4 technical report. <https://openai.com/research/gpt-4> (2023). Accessed via OpenAI API (<https://platform.openai.com>).