Supplementary information for: More than news! Mapping the deliberative potential of a political online ecosystem with digital trace data

Lisa Oswald
Center for Adaptive Rationality
Max Planck Institute for Human Development
Berlin
Germany

Appendix A
Methodological Details

Political website identification

In order to capture every potentially relevant website visit, an extensive dictionary of political topics, in particular topics within the public discourse in 2017, was compiled to classify the full sample of URLs. While most studies in the field of online media research operate on the domain or subdomain level (Guess, 2021), I use the full text of URLs for this classification step.

Keywords capturing potentially political topics were collected from a range of source websites, for example the federal center for political education (Federal Agency for Civic Education, 2017) and various news articles discussing relevant issues for the 2017 federal election in Germany (see SI D). I went through the full text of those articles and collected keywords until the dictionary appeared 'saturated' because with further keyword collection, majorly duplicates appeared[1]. Instead of limiting the analytical focus a-priorily, this procedure ensures that as many sites as possible that are potentially relevant for the public discourse are captured. The first version of the dictionary included 266 keywords[2]. Using this list, I ran the filter query to select the potentially political URLs. In a next step, grouped by domains, the output was checked manually for face validity on the website level (by visiting the URL). More specifically, I checked whether the websites were actually featuring political topics or discussions and if not, I tried to identify the keyword that mistakenly selected this website. I then excluded a considerable set of keywords from the dictionary, to sharpen the query. I also added a layer of exclusions to the query, for example I include URLs with the keyword 'petition' but exclude 'competition'. I repeated this man-

[1] I also tried scraping the full text of websites and processing it with various automated text mining approaches to extract keywords in a semi-automated way. However, those resulting keyword sets appeared much less useful which I attribute to substantive parts of text being embedded in graphics which cannot be scraped easily with an html-based text extraction process. Therefore, and facing a small number of dense websites, I opted for a manual approach of keyword extraction.

[2] Words including German Umlaute like 'ä', 'ö' and 'ü', I included as is, as 'ae', 'oe' and 'ue', and in UTF-8 hexadecimal encoding.

ual cross validation step repeatedly, until I did not find any more systematic mismatches (false positives). The final dictionary and the list of exclusions can be found in SI (D).

Furthermore, since many keywords (such as family, rent and finance) also appear in many non-political contexts, all website visits that were previously classified as irrelevant category[3] were excluded from the set of relevant URLs. Search engines were kept in the dataset as information-providers when people searched for political topics, even though they often only work as mediators and do not provide information themselves. However, political information seeking is largely considered as highly important digital feature of the overall democratic process, to the extent that it can be used to predict elections (Salem & Stephany, 2020).

Latent Class Analysis

The latent class model probabilistically assigns each website to a 'latent class', which in turn produces expectations about how that website will score on each of the manifest deliberative criteria that constituate the six dimensions of deliberative potential (Linzer & Lewis, 2011). Accordingly, websites within the same latent class are similar on certain dimensions, while those in different latent classes are dissimilar from each other in some significant way.

The parameters of the latent class model, p_r and π_{jrk} as prior and class-conditional outcome probabilities that each individual belongs to each class, are iteratively estimated by an EM-Algorithm (estimation maximization). In a first step, class-membership probabilities are estimated, while in a second step those estimates are altered to maximize the likelihood-function

$$\ln L = \sum_{i=1}^{N} ln \sum_{r=1}^{R} p_r \prod_{j=1}^{J} \prod_{k=1}^{K_j} (\pi_{jrk})^{Y_{ijk}}. \tag{A1}$$

With N representing the number of observations (websites) and J representing the manifest variables (deliberative criteria) with K possible outcomes (present or not present). R represents the outcome probabilities which are fixed prior to estimation.

Both steps are iteratively repeated until the global maximum is found (Ohlsen, 2015). I fitted the latent class models with the poLCA::poLCA() R function (Linzer & Lewis, 2011), which suits the categorical assessment of the criteria for deliberative potential. In this case, the number of classes in the sample is not theoretically defined but identified through the consideration of the smallest Bayesian Information Criterion (BIC). Ideally, the identified model should have the smallest number of classes possible while maintaining a low classification error Díaz and Koutra (2013).

I had no prior theoretical expectations about the number of latent classes and therefore, started to fit an 'independence' model with one latent class (Linzer & Lewis, 2011) and then iteratively increased the number of latent classes by one up to a number of 10 latent classes.

[3]irrelevant domain categories identified in the first round of top 1000 domain coding: astrology, banking, cashback, cooking, dating, file-hosting, function, gambling, mail, micro-job, porn, shopping, sport, survey, tracking, transport, travel, uni, work

Considering various model fit criteria a model with three latent classes was selected (Lin & Dayton, 1997). The model fit criteria were very similar between a two-class and a three-class solution. I therefore also present results for the two class solution in SI but stick with the three-class solution in the main body of the text. I validated the model (a) using a split-sample approach, randomly cutting 50% of the sample with no considerable effects to the latent class solution (see SI C10) and (b) by making changes to the input criteria (Bacher & Vermunt, 2010). Dropping criteria from the model input that do not differ substantially within a class (in this case, age and educational inclusivity and party preference heterogeneity), does also not significantly change the results (see SI C9). The classes still show the same characteristics and relative sizes, with only minor changes in the individual class memberships of websites.

As another robustness check, the supply side criteria (information, communication and participation) and the demand side criteria (connectivity, inclusivity and heterogeneity) were also considered separately in two LCA models, as they form two natural, methodological groups. For the computationally coded, demand side criteria, a simple two factor solution was suggested with one class including all websites with high probabilities of fulfilling each criterion and one class with overall very low scores for connectivity, inclusivity and heterogeneity in other words, high and low engagement websites. The model including only the manually coded infrastructural criteria of information, communication and participation possibilities suggested a more interesting pattern that is in line with the findings from the main model including all criteria. A first class contains websites with a strong information profile, including all public broadcasting pages. A second class contains websites with an especially strong forum component or communication profile with pages that also enable participation to some extent. The last class is rather a residual class including websites with overall low probabilities of fulfilling any criteria.

Shannon-Wiener Index

Entropy-based diversity indices such as the here applied Shannon-Wiener index (Dixon, 2003; Grafton et al., 2012; Kiernan, 2014; Oksanen, 2013) are regularly used in the context of bio-diversity assessments, combining measures of richness (the number of categories) and evenness (similar frequencies of categories).

$$H = -\sum_{i=1}^{S} p_i ln p_i \tag{A2}$$

with p_i being the proportion of category i, S being the number of categories so that $\sum_{i=1}^{S} p_i = 1$ (Oksanen, 2013).

Appendix B
Information on the Survey and Tracking Data

Sampling

The survey was administrated by YouGov combining purposive sampling with a multi-stage sample-matching and weighting procedure (Rivers, 2006). First, the target population, the German online population, was defined. For the panel, data on demographic marginals (gender, age and educational attainment) from Best for Planning (2017) were used, who conducted 30,000 face-to-face interviews to evaluate the German online population. A stratified sample was drawn from this frame and matched as closely as possible to YouGov's longstanding panel (with over a million members) (Munzert et al., 2021, same data).

The resulting target sample constitutes a representative set of respondents in terms of traditional sampling theory. However, respondents might be hard to contact because they either have never reported their contact details or do not agree to the terms of the survey. Hence, multi-stage matching was applied, combining the representative target sample with YouGov's longstanding panel of reliable respondents. From this panel, a sample of individuals was selected that matches as closely as possible the distribution of the target sample and had opted in to providing website visit data (see below). Through this procedure, YouGov guarantees not only a minimum of 1,000 respondents in the survey, but also the inclusion of hard-to-reach population subgroups (Munzert et al., 2021).

Survey Design and Fielding

All data was gathered by YouGov from July 1 to December 9, 2017. The panel was made up of five waves. The survey covered a wide range of topics such as people's political preferences, political knowledge on several domains, their general attitudes towards politics, opinions on particular parties, and what people think of the election campaign (the federal elections were held on September 24, 2017) (Munzert et al., 2021, same survey).

Passive Metering Technology

Wakoopa, the tracking software used by YouGov in this study, ran in the background of panelists' devices and collected anonymized visit data. There are no technological limits to the types of websites that can be included in the data. Moreover, the software tracks web traffic (passwords and financial transactions are ignored) for all browsers installed on a user's computer. The technology does not slow the performance of users' computers and is transparent about the data that is being sent: Panelists can see a list of the last several captured URLs and can also pause tracking for 15 minutes. Of course, they can also uninstall the software at any time. YouGov encourages its panelists to install the software on as many devices as possible, including laptops, mobile phones, and tablets. The capabilities for mobile tracking are somewhat more limited for privacy reasons, but data on domain-level visits and app use are collected (Munzert et al., 2021, same data).

Panelists were recruited from YouGov's traditional participant pool via incentives. The company reports fairly strong incentives: 4,000 "points" for signing up and downloading the Wakoopa software—roughly 8 times the number offered for a typical survey—and 1,000 additional points every month. Participants in online surveys can redeem these points for clothing, prepaid gift cards, and other merchandise. One consequence of this recruitment strategy is that YouGov Pulse users are a subset of the overall panel, making sampling somewhat more challenging (Munzert et al., 2021).

Privacy and Ethical Considerations of Data

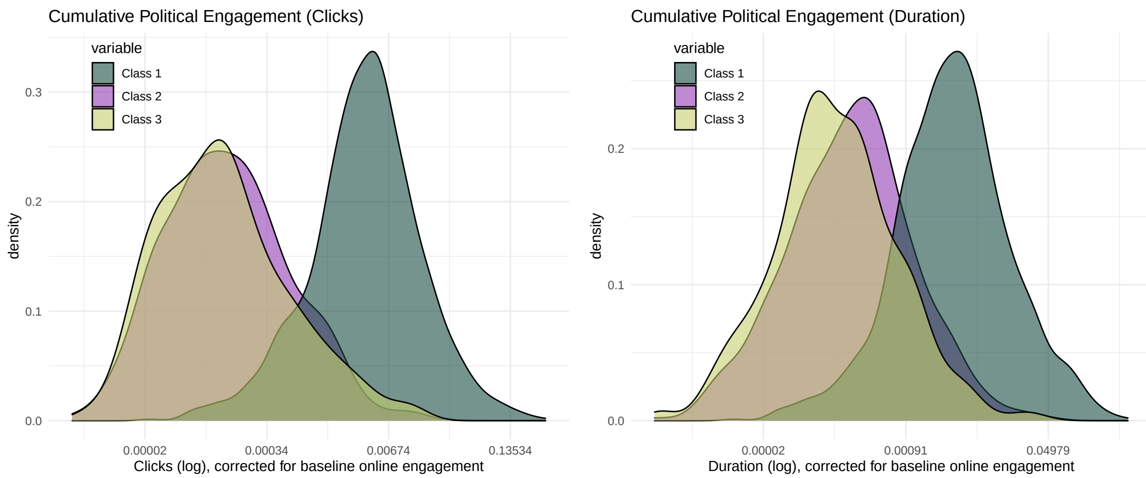
Combining survey data and digital trace data of the same respondents has substantive merits to understand the effects of online exposure on people's attitudes and behavior. However, it entails challenging tasks for the protection of privacy and raises ethical questions, as users may not be aware of how their data are being used. Even with the consent of the participants, it still could be problematic because the account names and meta-information of social media accounts can be identifiable and linked to their survey responses (Stier et al., 2020). Thus, it is important to communicate these concerns as clearly as possible when collecting data (Menchen–Trevino, 2013). In every step of data collection, participants were informed about the scope of data collection, data management, confidentiality, and research purpose. Explicit and informed consent was obtained from all participants whose data was collected (Munzert et al., 2021, same considerations).

Regarding the web-tracking data, YouGov received the consent from the panel that their web browsing data can be linked to other survey items they have participated. They highlighted that participants have complete control over which data they share for research purposes. Participants can choose which information they want to share, pause the tracking app when they want, and withdraw their consent anytime. After data collection, YouGov removed any personally identifying information and sensitive data (e.g., financial transaction) and stripped-out geocoding information that is too specific before delivering the data to researchers. The deliverables are de-identified and anonymized and fully comply with the EU's General Data Protection Regulation (GDPR) requirements (Munzert et al., 2021).

Appendix C
Supporting Tables and Figures

Figure C1

Engagement with different classes of sites

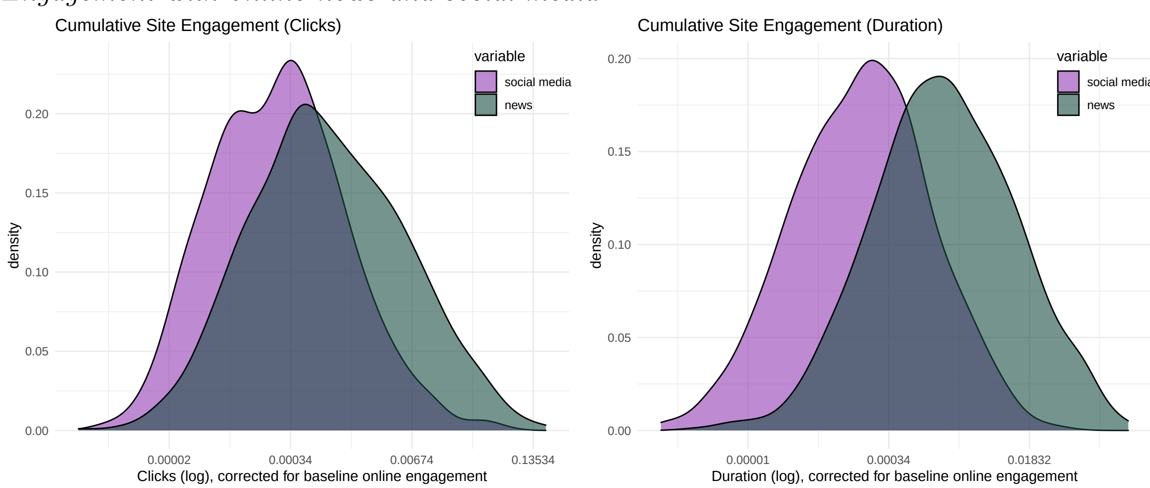


(a) *Number of website visits, baseline corrected.* (b) *Duration, baseline corrected.*

Note. Based on response probability patterns and class membership, class 1 was named ‘mainstream hubs’, class 2 was named ‘quality information providers’ and class 3 was named ‘niche forums’.

Figure C2

Engagement with online news and social media



(a) *Number of website visits, baseline corrected.* (b) *Duration, baseline corrected.*

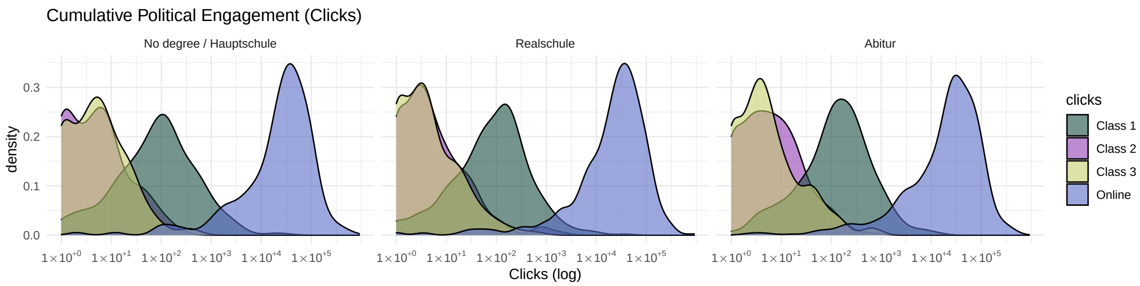


Figure C3

Engagement with different classes of sites by level of formal education. Class 1: 'mainstream hubs', class 2: 'quality information providers' and class 3: 'niche forums'.

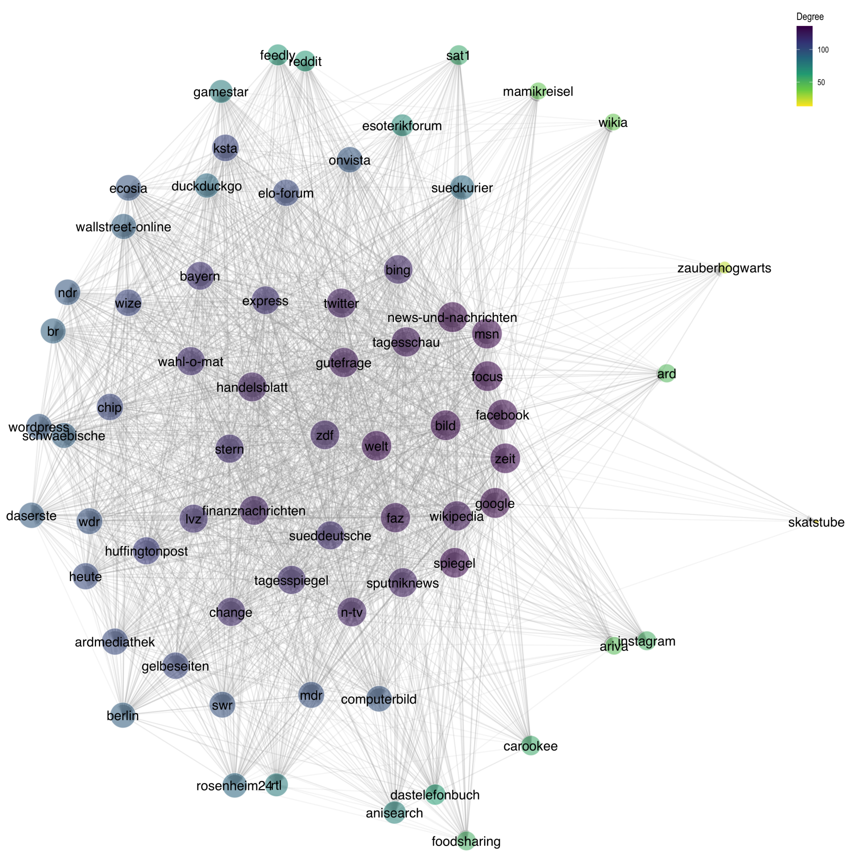


Figure C4

Connectivity - quantitative measure of ingoing and outgoing traffic from other relevant sites.

Table C1

Class Memberships

C3 domains	clicks	**C1 domains**	clicks	**C2 domains**	clicks
elo-forum	1208	google	102926	heute	1179
computerbild	1151	news-und-nachrichten	49265	swr	1129
duckduckgo	845	bild	36771	suedkurier	1068
ecosia	807	facebook	31592	ksta	1040
wordpress	785	wahl-o-mat	30046	schwaebische	912
wallstreet-online	711	msn	29460	daserste	896
onvista	646	focus	27413	berlin	894
gamestar	595	welt	18819	wdr	883
esoterikforum	253	twitter	17566	ndr	831
instagram	228	zeit	16609	mdr	767
dastelefonbuch	210	spiegel	11999	rosenheim24	662
reddit	209	gutefrage	10758	br	607
feedly	133	handelsblatt	9698	rtl	437
carookee	124	wikipedia	9287	sat1	156
wikia	114	faz	8810	ard	99
foodsharing	105	finanznachrichten	7279		
ariva	104	tagesschau	6712		
mamikreisel	83	bayern	5848		
zauberhogwarts	29	zdf	5544		
skatstube	7	n-tv	5451		
		bing	4957		
		sputniknews	3993		
		change	3437		
		sueddeutsche	2820		
		stern	2403		
		tagesspiegel	2237		
		lvz	2078		
		express	1812		
		gelbeseiten	1712		
		chip	1356		
		huffingtonpost	1345		
		wize	1282		
		anisearch	1280		
		ardmediathek	1242		

Note. Classes of domains according to predicted probabilities of class membership, sorted by clicks in the sample.

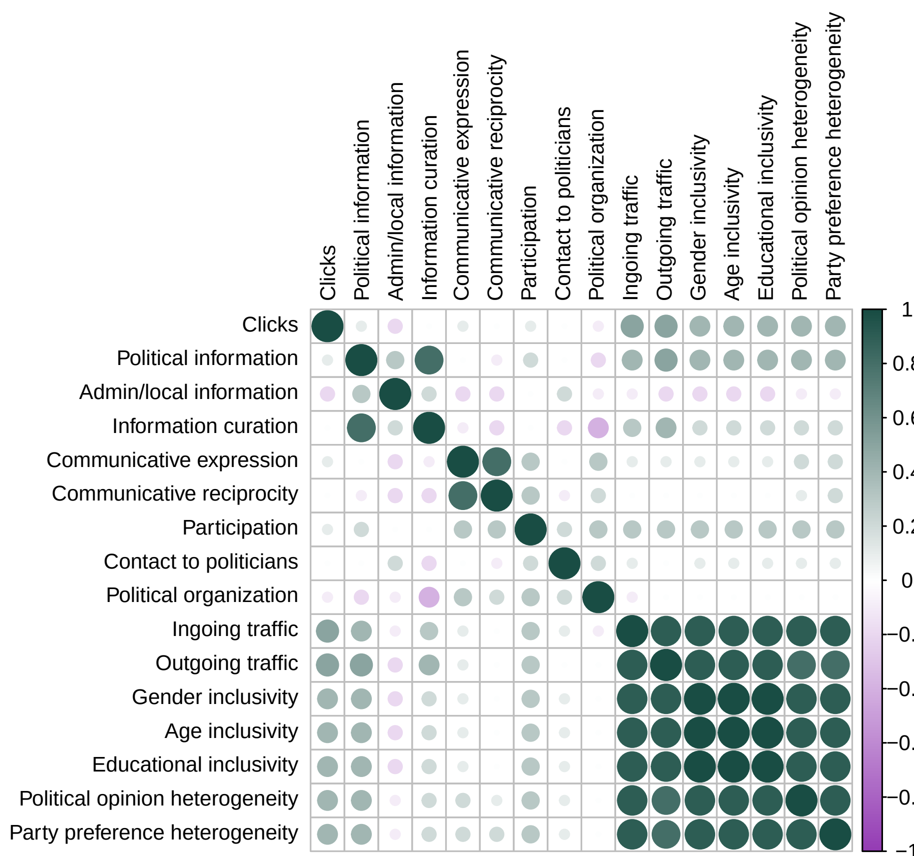


Figure C5

Correlation table between different Deliberative Potential criteria. Black squares indicate criteria that belong to one dimension.

Table C2

Latent Class Model Comparison

Model	Log Likelihoods	Df	BIC	aBIC	AIC	Likelihood Ratios
model 1	-651	54	1365	1318	1380	833
model 2	-408	38	948	850	979	348
model 3	-376	22	951	803	998	283
model 4	-353	6	972	774	1035	237
model 5	-337	-10	1009	760	1088	206
model 6	-321	-26	1045	746	1140	174
model 7	-312	-42	1094	744	1205	155
model 8	-301	-58	1139	739	1266	133
model 9	-295	-74	1195	744	1338	121
model 10	-285	-90	1244	743	1403	102

Note. Model number represents number of latent classes. Values rounded to integers.

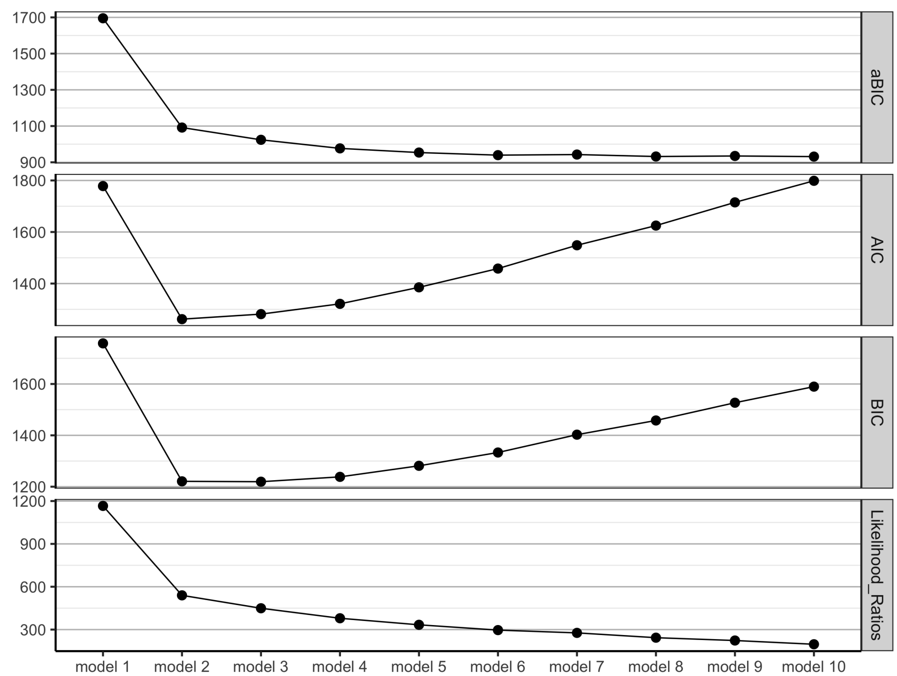
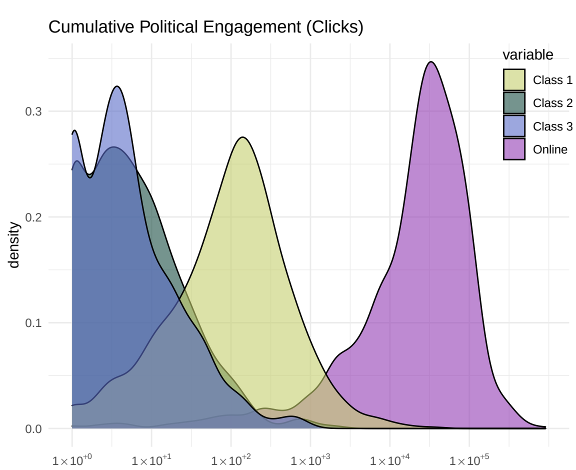
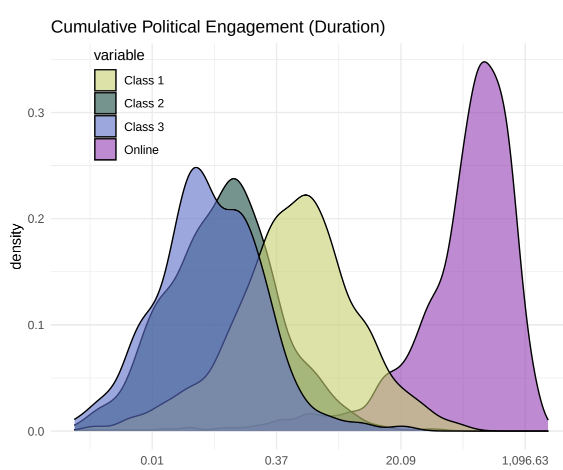


Figure C6

Elbow plot - Latent Class Analysis Model Fit Comparison



(a) *Engagement with classes of sites (clicks, raw)*



(b) *Engagement with classes of sites (duration, raw)*

Table C3

Class Memberships - Two-class solution

C1_domains	clicks	C2_domains	clicks
google	102926	elo-forum	1208
news-und-nachrichten	49265	heute	1179
bild	36771	computerbild	1151
facebook	31592	swr	1129
wahl-o-mat	30046	suedkurier	1068
msn	29460	ksta	1040
focus	27413	schwaebische	912
welt	18819	daserste	896
twitter	17566	berlin	894
zeit	16609	wdr	883
spiegel	11999	duckduckgo	845
gutefrage	10758	ndr	831
handelsblatt	9698	ecosia	807
wikipedia	9287	wordpress	785
faz	8810	mdr	767
finanznachrichten	7279	wallstreet-online	711
tagesschau	6712	rosenheim24	662
bayern	5848	onvista	646
zdf	5544	br	607
n-tv	5451	gamestar	595
bing	4957	rtl	437
sputniknews	3993	esoterikforum	253
change	3437	instagram	228
sueddeutsche	2820	dastelefonbuch	210
stern	2403	reddit	209
tagesspiegel	2237	sat1	156
lvz	2078	feedly	133
express	1812	carookee	124
gelbeseiten	1712	wikia	114
chip	1356	foodsharing	105
huffingtonpost	1345	ariva	104
wize	1282	ard	99
anisearch	1280	mamikreisel	83
ardmediathek	1242	zauberhogwarts	29
		skatstube	7

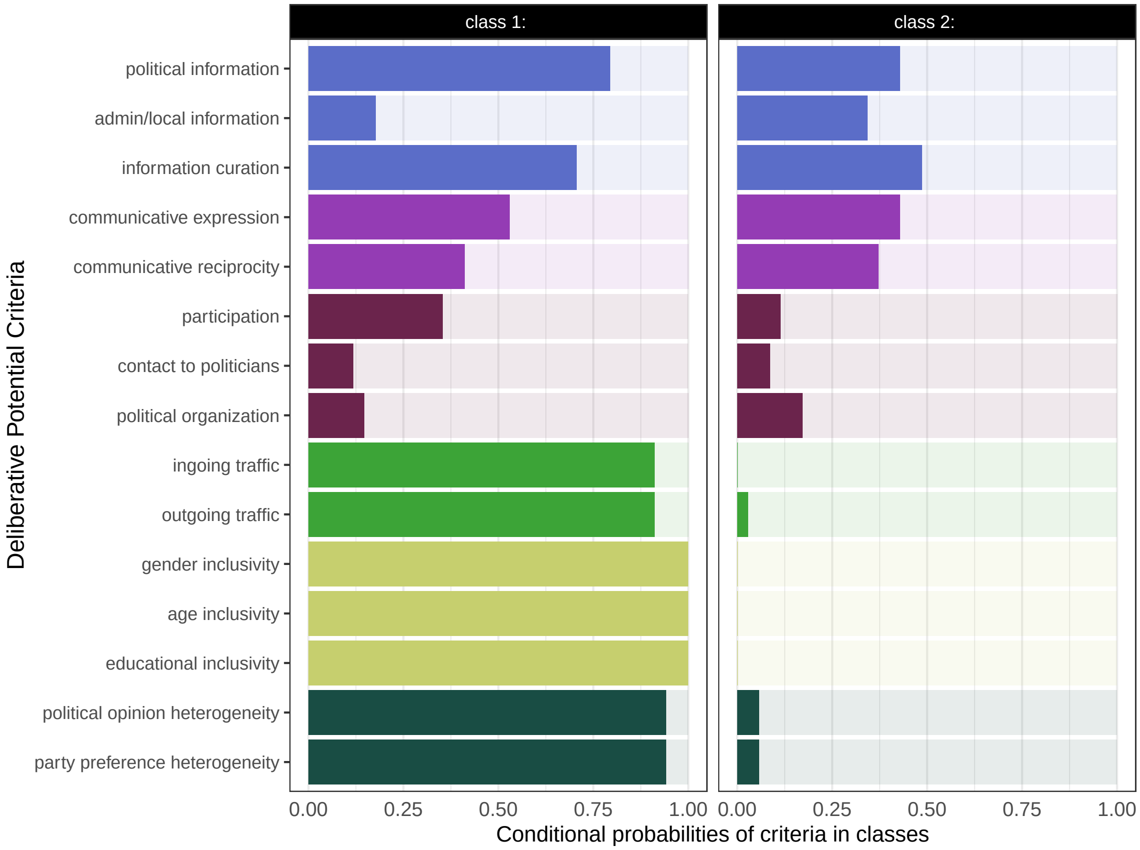


Figure C8

Conditional response probabilities, by deliberative potential criterion, of belonging to each latent class - Two-class solution.

Table C4

Fit criteria - Hand coded infrastructure only

Model	Log_Likelihoods	Df	BIC	aBIC	AIC	Likelihood_Ratios
model 1	-317	61	667	642	675	228
model 2	-276	52	625	571	642	147
model 3	-255	43	620	538	646	105
model 4	-238	34	624	514	659	71
model 5	-227	25	640	501	684	48
model 6	-221	16	666	499	719	36
model 7	-216	7	694	499	756	27
model 8	-213	-2	727	503	798	21
model 9	-210	-11	758	507	838	15
model 10	-207	-20	792	511	881	10

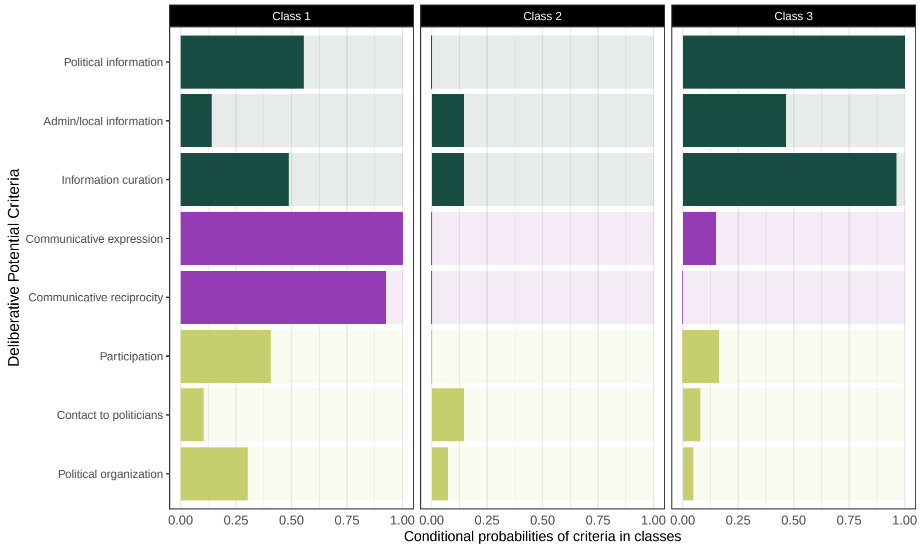


Figure C9

Hand coded criteria only.

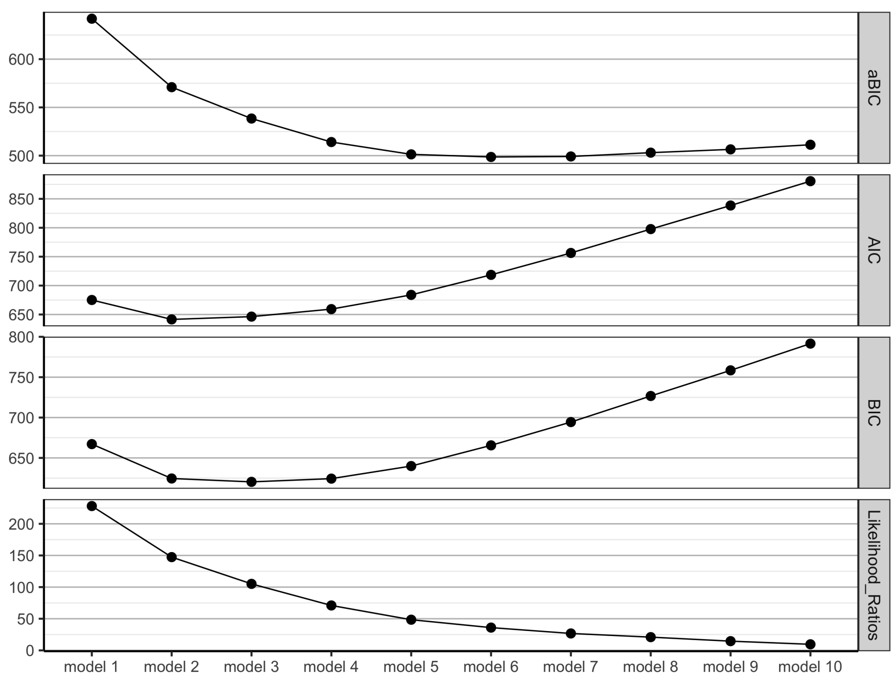


Figure C10

Elbow plot - Latent Class Analysis Model Fit Comparison - Hand coded criteria only

Table C5

Class Memberships - Hand coded criteria only

C1_domains	clicks	C2_domains	clicks	C3_domains	clicks
wahl-o-mat	30046	google	102926	bild	36771
msn	29460	finanznachrichten	7279	facebook	31592
wikipedia	9287	bing	4957	focus	27413
tagesschau	6712	gelbeseiten	1712	twitter	17566
bayern	5848	chip	1356	zeit	16609
zdf	5544	computerbild	1151	spiegel	11999
n-tv	5451	duckduckgo	845	gutefrage	10758
stern	2403	ecosia	807	faz	8810
lvz	2078	wordpress	785	sputniknews	3993
express	1812	onvista	646	sueddeutsche	2820
ardmediathek	1242	dastelefonbuch	210	tagesspiegel	2237
swr	1129	feedly	133	huffingtonpost	1345
suedkurier	1068	foodsharing	105	wize	1282
ksta	1040	ariva	104	anisearch	1280
daserste	896			elo-forum	1208
berlin	894			heute	1179
wdr	883			wallstreet-online	711
ndr	831			rosenheim24	662
mdr	767			br	607
rtl	437			gamestar	595
sat1	156			esoterikforum	253
ard	99			instagram	228
				reddit	209
				carookee	124
				wikia	114
				mamikreisel	83
				zauberhogwarts	29
				skatstube	7

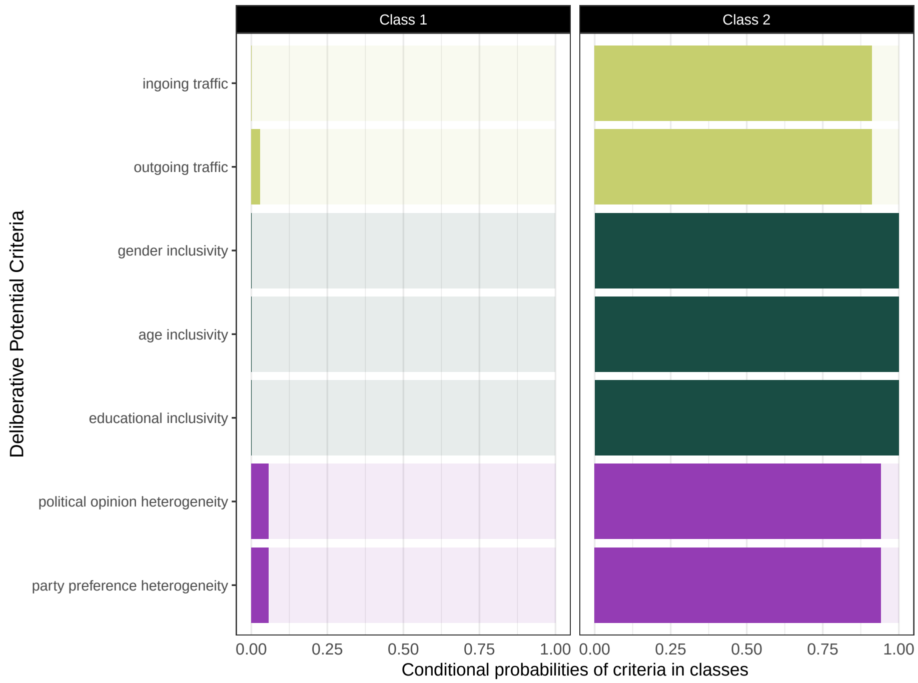


Figure C11

Computationally coded criteria only.

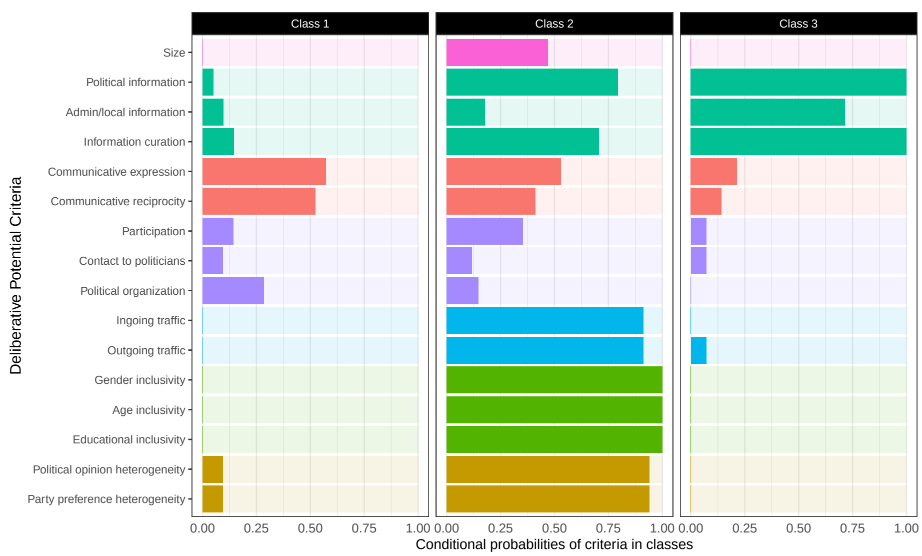


Figure C12

Website size (measured in clicks) added as additional criterion.

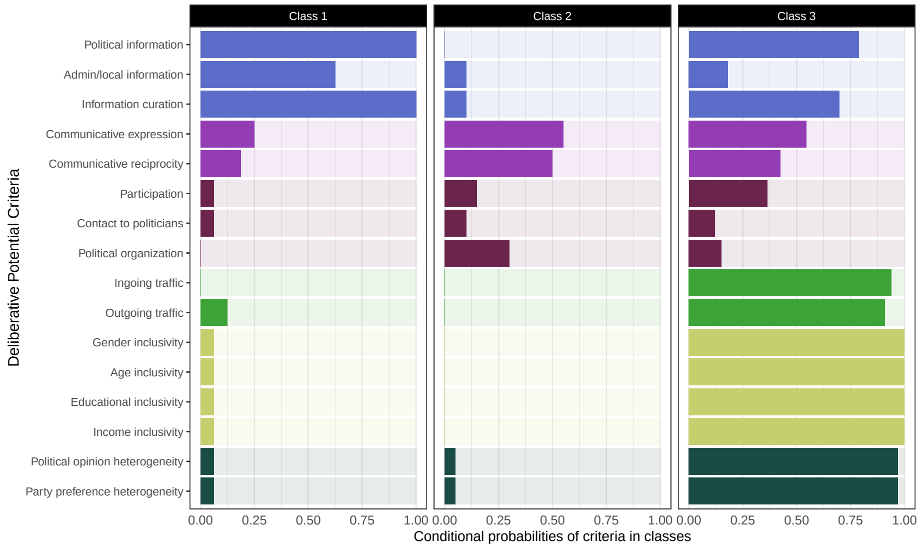


Figure C13

Household income added as additional criterion for inclusivity.

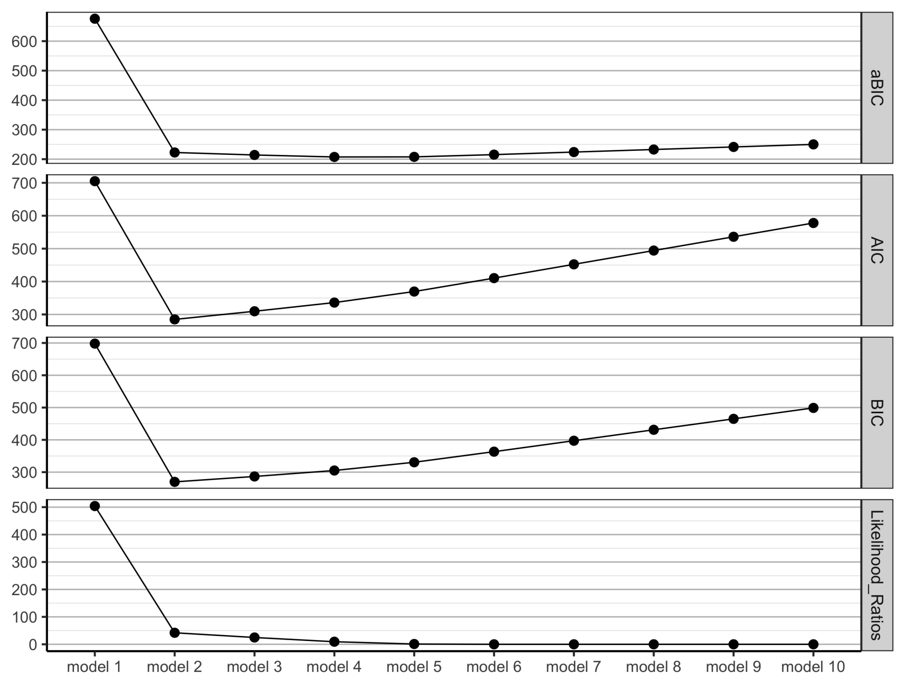


Figure C14

Elbow plot - Latent Class Analysis Model Fit Comparison - Demand criteria only

Table C6

Fit criteria - Computationally coded engagement based criteria only

Model	Log_Likelihoods	Df	BIC	aBIC	AIC	Likelihood_Ratios
model 1	-334	62	698	676	705	504
model 2	-103	54	270	223	285	42
model 3	-95	46	287	214	310	25
model 4	-87	38	305	207	336	9
model 5	-83	30	331	208	370	1
model 6	-82	22	363	215	410	0
model 7	-82	14	397	224	452	0
model 8	-82	6	431	233	494	0
model 9	-82	-2	465	241	536	0
model 10	-82	-10	499	250	578	0

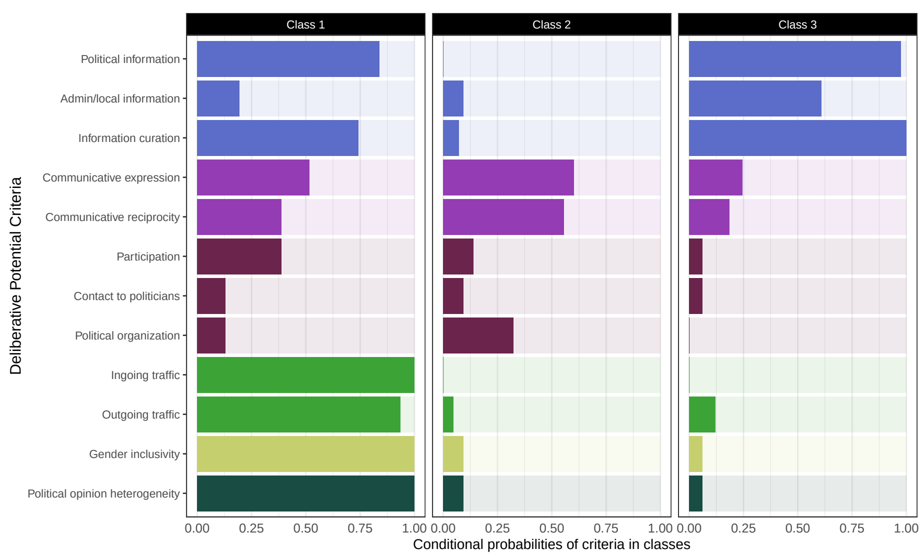


Figure C15

Solution with 3 criteria dropped from the sample.

Table C7

Class Memberships - Computationally coded crieria only

C1_domains	clicks	C2_domains	clicks
google	102926	elo-forum	1208
news-und-nachrichten	49265	heute	1179
bild	36771	computerbild	1151
facebook	31592	swr	1129
wahl-o-mat	30046	suedkurier	1068
msn	29460	ksta	1040
focus	27413	schwaebische	912
welt	18819	daserste	896
twitter	17566	berlin	894
zeit	16609	wdr	883
spiegel	11999	duckduckgo	845
gutefrage	10758	ndr	831
handelsblatt	9698	ecosia	807
wikipedia	9287	wordpress	785
faz	8810	mdr	767
finanznachrichten	7279	wallstreet-online	711
tagesschau	6712	rosenheim24	662
bayern	5848	onvista	646
zdf	5544	br	607
n-tv	5451	gamestar	595
bing	4957	rtl	437
sputniknews	3993	esoterikforum	253
change	3437	instagram	228
sueddeutsche	2820	dastelefonbuch	210
stern	2403	reddit	209
tagesspiegel	2237	sat1	156
lvz	2078	feedly	133
express	1812	carookee	124
gelbeseiten	1712	wikia	114
chip	1356	foodsharing	105
huffingtonpost	1345	ariva	104
wize	1282	ard	99
anisearch	1280	mamikreisel	83
ardmediathek	1242	zauberhogwarts	29
		skatstube	7

Table C8

Model Fit - Three criteria dropped

Model	Log Likelihoods	Df	BIC	aBIC	AIC	Likelihood Ratios
model 1	-507	57	1066	1028	1078	546
model 2	-389	44	883	804	908	309
model 3	-352	31	866	746	904	236
model 4	-330	18	876	715	927	191
model 5	-313	5	896	695	960	157
model 6	-303	-8	931	689	1008	136
model 7	-294	-21	970	686	1060	120
model 8	-285	-34	1006	682	1109	102
model 9	-278	-47	1048	683	1164	88
model 10	-274	-60	1093	687	1222	79

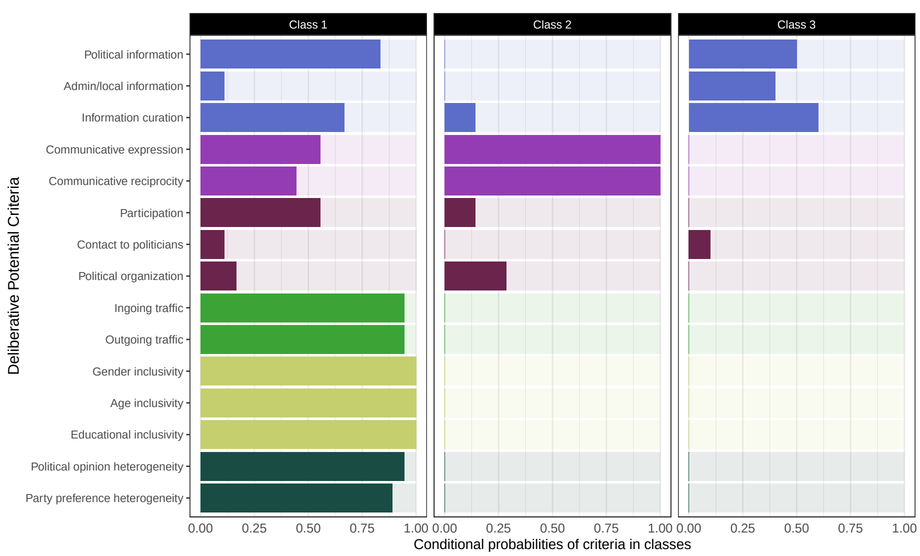


Figure C16

Solution with 50% split sample.

Table C9

Class Memberships - dropped 3 criteria

C1_domains	clicks	C2_domains	clicks	C3_domains	clicks
wize	1282	google	102926	ardmediathek	1242
anisearch	1280	news-und-nachrichten	49265	heute	1179
elo-forum	1208	bild	36771	swr	1129
computerbild	1151	facebook	31592	suedkurier	1068
duckduckgo	845	wahl-o-mat	30046	ksta	1040
ecosia	807	msn	29460	schwaebische	912
wordpress	785	focus	27413	daserste	896
onvista	646	welt	18819	berlin	894
gamestar	595	twitter	17566	wdr	883
esoterikforum	253	zeit	16609	ndr	831
instagram	228	spiegel	11999	mdr	767
dastelefonbuch	210	gutefrage	10758	rosenheim24	662
reddit	209	handelsblatt	9698	br	607
feedly	133	wikipedia	9287	rtl	437
carookee	124	faz	8810	sat1	156
wikia	114	finanznachrichten	7279	ard	99
foodsharing	105	tagesschau	6712		
mamikreisel	83	bayern	5848		
zauberhogwarts	29	zdf	5544		
skatstube	7	n-tv	5451		
		bing	4957		
		sputniknews	3993		
		change	3437		
		sueddeutsche	2820		
		stern	2403		
		tagesspiegel	2237		
		lvz	2078		
		express	1812		
		gelbeseiten	1712		
		chip	1356		
		huffingtonpost	1345		

Table C10

Model Fit - Split Sample (50%)

Model	Log Likelihoods	Df	BIC	aBIC	AIC	Likelihood Ratios
model 1	-312	20	678	631	693	425
model 2	-180	4	471	374	502	161
model 3	-164	-12	495	348	542	127
model 4	-149	-28	522	325	585	98
model 5	-140	-44	561	314	640	80
model 6	-131	-60	600	303	695	62
model 7	-124	-76	643	297	754	49
model 8	-119	-92	690	293	817	38
model 9	-115	-108	739	292	882	30
model 10	-112	-124	788	292	947	23

Table C11

Class Memberships - Split Sample (N = 35 domains)

C1_domains	clicks	C2_domains	clicks	C3_domains	clicks
google	102926	computerbild	1151	swr	1129
facebook	31592	duckduckgo	845	suedkurier	1068
wahl-o-mat	30046	ecosia	807	rosenheim24	662
msn	29460	onvista	646	br	607
welt	18819	esoterikforum	253	sat1	156
handelsblatt	9698	instagram	228	ard	99
wikipedia	9287	reddit	209		
finanznachrichten	7279	feedly	133		
tagesschau	6712	foodsharing	105		
bayern	5848	zauberhogwarts	29		
n-tv	5451				
bing	4957				
change	3437				
sueddeutsche	2820				
tagesspiegel	2237				
lvz	2078				
express	1812				
gelbeseiten	1712				
anisearch	1280				

Table C12

Top 1000 domains - categories

Domain Category	Clicks	Domain Category	Clicks
survey	10359976	tracking	392589
social_media	6016296	messanger	269634
mail	5525976	information	238403
shopping	5183965	networking	160659
search	4475112	tv	131309
micro_job	3752463	travel	120034
gaming	2244070	sport	109112
function	1418048	work	99610
news	1077675	participation	65411
streaming	945883	transport	60278
porn	779554	uni	46372
gambling	575798	cooking	43618
banking	563014	radio	38112
cashback	548676	blog	25754
forum	497458	file-hosting	25535
dating	440430	astrology	13759
hobby	395972	counceling	7484

Table C13

Political Topics

	Topic (group)	clicks
1	Politics and citizens	126542
2	News	76534
3	Election	46357
4	Employment	24141
5	Parties	23237
6	Finance	21699
7	Politicians	10548
8	Justice	9247
9	Science and Journalism	8708
10	Social Issues	7774
11	Governance	6904
12	Health and Pension	6547
13	Safety and Security	6018
14	Government	5114
15	International Politics	5094
16	Environmental Issues	5080
17	US Politics	4294
18	Migration	3468
19	Infrastructure	2950
20	Housing	1137
21	Brexit	657

Appendix D
Websites for political keyword selection

- https://de.wikipedia.org/wiki/Bundestagswahl_2017
- https://www.bpb.de/politik/hintergrund-aktuell/261923/rueckblick-2017
- https://www.sueddeutsche.de/politik/jahresrueckblick-die-politischen-momente-2017-1.3802865
- https://interaktiv.morgenpost.de/probleme-bundestagswahl-2017/
- https://de.statista.com/infografik/10787/wichtige-themen-bei-der-bundestagswahl/
- https://www.bpb.de/politik/hintergrund-aktuell/256110/bundestagswahl
- https://www.bertelsmann-stiftung.de/de/themen/aktuelle-meldungen/2017/oktober/bundestagswahl-2017-wahlergebnis-zeigt-neue-konfliktlinie-der-demokratie/
- https://web.de/magazine/politik/wahlen/bundestagswahl/bundestagswahl-2017-themen-entscheiden-wahlkampf-32540410

Political Keyword Dictionary

Inclusion = (abgabe,abgeordnete , abkommen , administrat , adoption , afd , altmaier , anteil , arbeit , armut , asyl , ausbau , auslaend , ausland , ausländ , auszählung , autorit , b%C3%BCndnis , b%C3%BCrger , bangladesch , bartsch , behoerd , berechtigt , beteilig , betreuung , betrug , bezirk , bildung , bombe , brandner , brexit , brüssel , buendnis , buerger , bundes , bündnis , bürger , cdu , christlich , clinton , delegiert , demokrat , digitalisier , diplomat , divers , einwander , energiewende , erdogan , erwerbslos , europäisch , feindlich , feminis , finanz , flücht , fluecht , fraktion , freiheit , gauland , gaza , gesellschaft , gesetz , gleichberechtigung , glyphosat , göring , gr%C3%BCne , greenpeace , groko , gruene , grundeinkommen , grundordnung , grüne , hagedorn , hamas , handel , hartz , hochrechnung , homoehe , infrastruktur , integration , investigativ , israel , jamaika , jerulasem , journalis , jugoslawien , justiz , kabinett , kampagne , kandidat , kanzler , katalan , katalon , kipping , kita , klausel , klima , koalition , komission , kompromiss , konflikt , kongress , konservativ , konzern , kriminal , kriminell , krise , kritiker , l%C3%B6hne , leyen , liberal , lindner , linke , loehne , lohn , löhne , maas , macron , maiziere , maizière , mecklemburg , mehrheit , merkel , mietpreis , migrant , migration , milieu , militär , minister , mobilisier , modernisier , mugabe , muslim , myanmar , nahles , nordkorea , nordrhein , notlage , notstand , oezdemir , %C3%B6zdemir , opposition , organisat , özdemir , pal%C3%A4stin , palaestin , palästin , parlament , partei , petition , petry , plakat , politbarometer , politi , populist , pr%C3%A4sident , praesident , präsident , präsidial , prek%C3%A4r , prekaer , prekär , president , propaganda , protektionis , protest , putsch , radikal , rakete , rassismus , recht , referendum , reform , regier , regional , rente , reporter , repression , republik , rheinland , richtlinie , rohingya , rutte , sanktion , sch%C3%A4uble , schaeuble , schäuble , schulz , seehofer , sicherheit , simbabwe , skepti , sondierung , sonntagsfrage , sozial , spaltung , spd , sperrklausel , staat , statistik , steinmeier , steuer , stimme , storch , terror , trump , umwelt , unabh%C3%A4ngig , unabhaengig , unabhängig , ungerecht , unterhaus

, unterschrift , untersuch , v%C3%B6lker , verantwort , verbrechen , verfassung , verhalten , verhandlung , verkehrswende , verordnung , versammlung , versorg , verteil , voelker , volk , völker , vorsitz , vorsorge , w%C3%A4hl , waehl , wagenknecht , wahl , wähl , wandel , washington , weidel , wende , wiedereinzug , wiedervereinigung , wilders , wirtschaft , wissenschaft , y%C3%BCcel , yücel , yuecel , zulassung)

Exclusion = (bundesliga, competition, auswahl, steuerberat, reitbeteiligung, gravitationsgesetz, robotergesetz, zickenkrieg, ausbildung, bearbeiten, schulze, kitarou)

Appendix E

Software Statement

The entire analysis was run under OS X 10.13.6 using R version 4.0.4 (2021-02-15) (R Core Team, 2018). In the empirical analysis, I made use of the following `R` software packages:

`dplyr` (Wickham & Francois, 2015),
`ggplot2` (Wickham, 2016),
`readr` (Wickham et al., 2017),
`readxl` (Wickham & Bryan, 2018) ,
`writexl` (Ooms, 2018) ,
`urltools` (Keyes et al., 2019),
`tidyverse` (Wickham et al., 2019),
`xtable`(Dahl, 2016) ,
`sf` (Pebesma et al., 2018),
`tidyLPA`(Rosenberg et al., 2019),
`vegan`(Dixon, 2003),
`poLCA`(Linzer & Lewis, 2011)
`gdata`(Warnes et al., 2015)
`knitr`(Xie, 2018)
`kableExtra`(Zhu, 2019)
`igraph`(Csardi & Nepusz, 2006)
`gridExtra`(Auguie & Antonov, 2017)
`panelr`(Long, 2021)
`jcolors`(Huling, 2019)

Code Book – Deliberative Potential Index

Unit of Analysis: Website / Domain

Dimension	Operational Definition	Criterion	Measurement	Reference
Information	Users can find political information on this website. Such sites serve as resource for opinion and will-formation.	1. Provides information on political issues, actors and institutions 2. Provides administrative or local information 3. Acts as primary or (journalistically) curated source of information[1]	Human rating (binary 0/1) for each criterion	Wiklund (2005), Esau et al. (2020)
Communication	The website provides users with the possibility to express and/or exchange political opinions with other users. Such sites serve as communicative spaces for (interactional) opinion and will-formation.	1. Enables commenting / political expression (and potentially ratings of comments) 2. Enables reciprocity / replies to comments of other users (open replies, not only ratings)	Human rating (binary 0/1) for each criterion	Wiklund (2005), Esau et al. (2020)
Participation	The website provides users with the possibility of online political participation or organization, implying a potential (direct) impact on political decision making.	1. Enables contact to political actors 2. Enables political participation (petitions, polls, etc.) 3. Enables political organization (events, groups, etc.)	Human rating (binary 0/1) for each criterion	Contact to politicians: Wiklund (2005), Polls: Richardson & Stanyer (2011), function for decision making/ aggregation of interests: Esau et al. (2020),
Isolation (Connectivity)	The website is connected to other relevant websites. This facilitates further research on political issues or the implementation of intentions of political participation.	1. Contains links to other politically relevant websites	a) Human rating (binary 0/1) b) Network centrality, in/outgoing relevant traffic	
Inclusivity[2]	The website is used by a comparably diverse set of individuals. This dimension serves as indicator for low barriers of access.	1. Educational backgrounds 2. Gender 3. Age	Quantitative assessment using web-tracking data in combination with survey information	Mansbridge & Parkinson (2012) / democratic function
Heterogeneity[2]	The website connects people holding a comparably wide range of political opinions. This stands in contrast to the notion of ideological online 'echo chambers'.	1. Political orientation 2. Party preferences	Quantitative assessment using web-tracking data in combination with survey information	Mansbridge & Parkinson (2012) / democratic function

[1] Curation representing a basic form of fact checking taking place (no obvious misinformation or strong opinion statements presented as factual statements)
[2] Optional category as only possible to assess with digital trace data + survey information)