Population of X/Twitter users and web domains embedded in a multidimensional political opinion space

Antoine Vendeville^{1,2,3,†}, Jimena Royo-Letelier^{1,†}, Duncan Cassells^{4,1,3}, Jean-Philippe Cointet¹, Maxime Crépel¹, Tim Faverjon¹, Théophile Lenoir¹, Béatrice Mazoyer¹, Benjamin Ooghe-Tabanou¹, Armin Pournaki⁵, Hiroki Yamashita^{2,6}, and Pedro Ramaciotti^{2,1,3,*,†}



SciencesPo

data.sciencespo > Sciences Po, dépôt de données de recherche > médialab (MEDIALAB) >

Population of X/Twitter users and web domains embedded in a multidimensional political opinion space

Version 2.0

Ո	
111	
լլլ	

Antoine Vendeville; Jimena Royo-Letelier; Duncan Cassells; Jean-Philippe Cointet; Maxime Crépel; Tim Faverjon; Théophile Lenoir; Béatrice Mazoyer; Benjamin Ooghe-Tabanou; Armin Pournaki; Hiroki Yamashita; Pedro Ramaciotti, 2025, "Population of X/Twitter users and web domains embedded in a multidimensional political opinion space", <u>https://doi.org/10.21410/7E4/QPECFF</u>, data.sciencespo, V2, UNF:6:ppwt/IF80vKAIGr21iwpDw== [fileUNF]

Cite Dataset 🕶

Learn about Data Citation Standards.

Description 🕑

The undertaking of several studies of political phenomena in social media mandates the operationalization of the notion of political stance of users and contents involved. Relevant examples include the study of segregation and polarization online, the study of political diversity in content diets in social media, or AI explainability. While many research designs rely on operationalizations best suited for the US setting, few allow addressing more general design, in which users and content might take stances on multiple ideology and issue dimensions, going beyond traditional Liberal-Conservative or Left-Right scales. To advance the study of more general online ecosystems, we present a dataset of X/Twitter population of users in the French political Twittersphere and web domains embedded in a political space

Access Da	taset -
Contact Owner	Share
Make Data Count (MI since 2024-01-01 161 Views 🚱	DC) Metrics 🕑
27 Downloads 🕗	

Overview

A publicly available dataset containing political opinions of :

- 980K users in the French Twittersphere
- 883 French Members of Parliament (MPs)
- 400 media domains (lemonde.fr, lefigaro.fr, ...)

Overview

A publicly available dataset containing political opinions of :

- 980K users in the French Twittersphere
- 883 French Members of Parliament (MPs)
- 400 media domains (lemonde.fr, lefigaro.fr, ...)

And...

- Activity metrics (number of tweets and retweets)
- Popularity metrics (number of times retweeted)
- Partisanship: most retweeted party

Dimensions of interest



Dimensions of interest





Example (users)

pseudo_id	Irecon_23	galtan_23	eu_position_23	antielite_salience_23
user_0	3.331	0.674	9.814	2.266
user_1	5.383	6.459	3.246	7.364
user_2	5.591	5.647	4.673	6.438
user_3	6.023	8.610	1.732	9.007
user_4	4.672	4.069	6.031	5.178
user_979084	4.030	5.654	1.814	8.606
user_979085	6.334	8.621	1.344	8.392
user_979086	2.917	4.895	1.821	7.959
user_979087	5.199	6.614	2.199	8.474
user_979088	6.228	6.165	5.256	5.627

Example (MPs)

pseudo_id	party	lrecon_23	galtan_23	eu_position_23	antielite_salience_23
user_2003	PCF	3.217	4.162	3.472	6.867
user_2340	UDI	6.918	5.431	7.599	3.356
user_2626	LFI	3.371	3.656	3.782	7.617
user_2652	LFI	2.637	2.683	4.427	6.896
user_3599	RN	5.918	8.932	0.540	9.301
user_975106	NaN	5.601	4.080	6.892	4.994
user_976402	LC	6.476	5.405	7.067	3.569
user_976502	Independent	3.256	1.513	8.304	3.362
user_976811	LR	7.478	7.093	5.574	5.071
user_977837	PS	2.742	2.606	5.154	5.958

Example (media)

domain	media_category	user_count	tweet_count	antielite_salience_23_mean	eu_position_23_mean	Irecon_23_mean
lemonde.fr	Centre	21139	179198	5.745	5.319	4.860
lefigaro.fr	Hyper-centre	16205	153754	6.195	4.558	5.647
francetvinfo.fr	Hyper-centre	16195	136294	6.032	4.939	4.968
radiofrance.fr	NaN	11836	62895	5.644	5.463	4.765
liberation.fr	Centre	11509	71206	6.131	4.901	4.574
jeune-nation.com	Identitarian	101	583	8.647	1.415	5.768
yetiblog.org	Right Wing	101	346	7.973	2.362	5.312
cointribune.com	NaN	100	611	6.610	4.041	5.729
attali.com	NaN	100	182	6.099	4.701	5.677
lemondemoderne.media	NaN	100	137	7.656	3.092	4.267

Example (activity)

	pseudo_id	n_tweets	n_retweets	n_retweeted	n_retweets_MPs	partisanship	partisanship_reliability
0	user_5	1.0	36.0	234.0	2.0	LFI	1.000
1	user_6	0.0	45.0	0.0	0.0	NaN	NaN
2	user_8	54.0	8889.0	1381.0	231.0	LFI	0.649
3	user_12	1.0	4.0	3.0	0.0	NaN	NaN
4	user_13	2315.0	424.0	53373.0	1.0	LFI	1.000
455273	user_979078	0.0	4.0	0.0	0.0	NaN	NaN
455274	user_979080	95.0	9.0	159.0	0.0	NaN	NaN
455275	user_979081	1.0	318.0	18.0	2.0	LR	1.000
455276	user_979085	0.0	1.0	0.0	0.0	NaN	NaN
455277	user_979088	0.0	4.0	0.0	0.0	NaN	NaN

455278 rows × 7 columns

Data Collection

January 2023.

886 MPs on Twitter (/925 total).

Collect their followers.

Remove bots and followers without enough political interest.

Data Collection

January 2023.

886 MPs on Twitter (/925 total).

Collect their followers.

Remove bots and followers without enough political interest.



Latent space embedding



Assumption: underlying homophily mechanism

Prob $(A_{ij} = 1 | \alpha_i, \beta_j, \gamma, \phi_i, \phi_j) = \text{logit}^{-1}(\alpha_i + \beta_j - \gamma \| \phi_i - \phi_j \|^2),$ (1) activity Latent coordinates

Pablo Barbera. Birds of the same feather tweet together: Bayesian ideal point estimation using Twitter data. Political Analysis, 23(1):76–91, 2015. ISSN 1047-1987, 1476-4989. Doi: 10.1093/pan/mpu011.

Assumption: underlying homophily mechanism

popularity $\operatorname{Prob}(A_{ij} = 1 | \alpha_i, \beta_j, \gamma, \phi_i, \phi_j) = \operatorname{logit}^{-1}(\alpha_i + \beta_j - \gamma \| \phi_i - \phi_j \|^2),$ activity Latent coordinates Approximation: correspondence analysis. Pablo Barbera. Birds of the same feather tweet together: Bayesian id data. Political Analysis, 23(1):76–91, 2015. ISSN 1047-1987, 1476-4

Ideological patterns already present...

... but we can do better !

Chapel Hill Expert Survey (CHES)

- Expert survey (incl. Sciences Po academics)
- Multiple countries
- Positions of political parties on several dimensions : left-right, EU, immigration, environment...
- Continuous values [0,10]
- We use the 2019 and 2023 waves.



Mapping onto CHES dimensions

- 1. Compute party positions in the latent space.
- 2. Infer linear mapping between party positions in latent space and in their political positions in CHES.
- 3. Apply the mapping to all users and MPs.



Activity and media domain positions

Large-scale database of tweets in French from January 2022 to July 2023.

Media: average position of users who shared an URL toward the media domain.

Domains selected from

• Crépel, M. et al. French media ecosystem map. SciencesPo data (2024).

Now let's see how it looks !

Position of users, MPs and media domains on selected axes (1)



Position of users, MPs and media domains on selected axes (2)



Validation

Compare political positions with opinions expressed by users in their Twitter bio.

- Manually (precise but time-consuming)
- With LLM (faster)

Evaluation via logistic regression.

Validation



Validation



Media domain positions (validation)

Media classification from

Cointet, J.-P. et al. Uncovering the structure of the French media ecosystem (2021).



Comparison with general population



Dimensionality of the political



More than 90% of opinion differences can be summarized with **two** ideological axes.

Dimensionality of the political Twittersphere

First axis: "Liberal-conservative"



Dimensionality of the political Twittersphere

First axis: "Liberal-conservative"

First PCA dimension (X users) Gender equality GAL-TAN Immigration Left-Right Environment Redistribution EU integration Anti-elitism Followers MPs Direct democracy 10 12 14 0 2 8 16 Λ 6 Contribution (%)

Second axis: "Populism"



Dimensionality of the political Twittersphere (bonus)

Third axis (<5% explained variance): "urban-rural divide"

Dimensionality and activity



The liberal-conservative axis gains in importance and the populism axis loses in importance as we focus on more and more active users.

Partisan sorting



Potential usages

- Dimensionality of politics
- Issue alignment
- Ideological polarization
- Partisan sorting
- Online media consumption
- Opinion dynamics
- Etc...

Thank you !

SciencesPo

User Guide Support English - Log In

data.sciencespo > Sciences Po, dépôt de données de recherche > médialab (MEDIALAB) >

Population of X/Twitter users and web domains embedded in a multidimensional political opinion space

Version 2.0

հ	1-4
пі	
ш	
ш	
ш	
ш	$ \longrightarrow $
II	

Antoine Vendeville; Jimena Royo-Letelier; Duncan Cassells; Jean-Philippe Cointet; Maxime Crépel; Tim Faverjon; Théophile Lenoir; Béatrice Mazoyer; Benjamin Ooghe-Tabanou; Armin Pournaki; Hiroki Yamashita; Pedro Ramaciotti, 2025, "Population of X/Twitter users and web domains embedded in a multidimensional political opinion space", <u>https://doi.org/10.21410/7E4/QPECFF</u>, data.sciencespo, V2, UNF:6:ppwt/IF80vKAIGr21iwpDw== [fileUNF]

Cite Dataset 🕶

Learn about Data Citation Standards.

Description 🚱

The undertaking of several studies of political phenomena in social media mandates the operationalization of the notion of political stance of users and contents involved. Relevant examples include the study of segregation and polarization online, the study of political diversity in content diets in social media, or AI explainability. While many research designs rely on operationalizations best suited for the US setting, few allow addressing more general design, in which users and content might take stances on multiple ideology and issue dimensions, going beyond traditional Liberal-Conservative or Left-Right scales. To advance the study of more general online ecosystems, we present a dataset of X/Twitter population of users in the French political Twittersphere and web domains embedded in a political space

Access D	Dataset 🗸
Contact Owner	Share
Make Data Count (N since 2024-01-01	/IDC) Metrics 📀
161 Views 🔞	

0 Citations 🕄