

Social Physics and Complexity - SPAC





February 7 at 1:00 PM · 🌐



Hay naku ! So it's been around then ? 😞 & How did they know about a Virus we knew nothing about ? 😞 Well , This is definitely a Government Conspiracy lysol labels as proof the government knew in advance about the Corona virus .

Corona Virus is nothing new.



- **2 MSc students defenses**

- *Dynamics of information spread in online social systems: a Twitter-based study*, by Jacopo Carotenuto, Universita de Padova
- *Exploring Hidden Societal Biases in Twitter Cascades: A Sociophysics Study*, by Tomás Estrócio e Silva, Instituto Superior Técnico

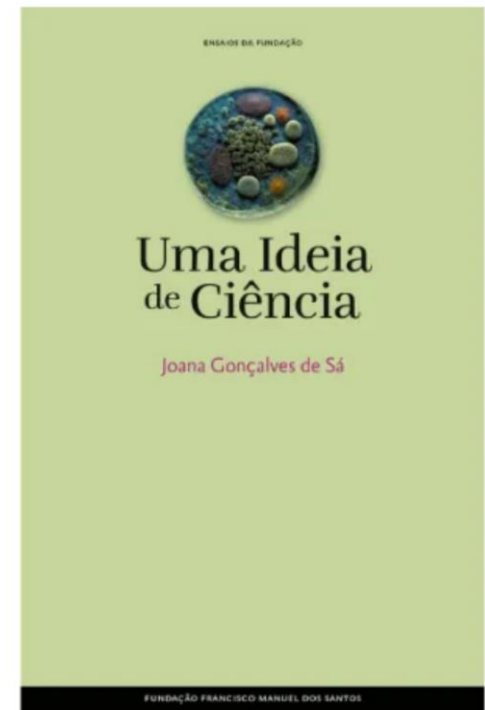
- **Changes at SPAC**

- Former postdoc, Ana Vranič, Researcher, University of Belgrade, Serbia
- Former postdoc, Cristina Mendonça, Assistant Professor, ISPA, Portugal
- Former postdoc, Lília Perfeito, Researcher, LIP, Portugal
- New PhD student, Mariana Silva – Physics Engineering, Instituto Superior Técnico
- 3 new MSc students

- **2 peer reviewed publications**

- Mesquita, S, Perfeito, L, Paolotti, D, Gonçalves-Sá, J (2025). Epidemiological methods in transition: Minimizing biases in classical and digital approaches. *PLOS Digital Health*, 4(1), e0000670.
- Damião, Í, Reis, JM, Almeida, P, Santos, N, Gonçalves-Sá, J (2025). Unequal Content: search engine personalization influences exposure to information regarding the Israel-Palestine conflict. *IEEE Access*.

- **1 peer-reviewed commentary**
 - Pierri, F., Araujo, T., Kruikemeier, S., Lorenz-Spreen, P., Vanden Abeele, M. M., Vandenbosch, L., Gonçalves-Sá, J., Grabowicz, P. A. (2025). Research Opportunities and Challenges of the EU's Digital Services Act. *Communications of the ACM*
- **1 dataset**
 - Vranić, A., Reis, J., Damião, Í., Almeida, P., & Gonçalves-Sá, J. (2025, October). Global Claims: A Multilingual Dataset of Fact-Checked Claims with Veracity, Topic, and Salience Annotations. In *Proceedings of the 2nd International Workshop on Diffusion of Harmful Content on Online Web – ACM*
- **1 report**
 - Huertas-Tato, J, et al., (2025). Disinformation on social networks in Spanish and Portuguese Iberian digital media ecosystem. IBERIFIER.
- **1 book for lay audiences**
- **1 new grant** – ERC Proof of Concept





February 7 at 1:00 PM · 🌐



Hay naku ! So it's been around then ? 😞 & How did they know about a Virus we knew nothing about ? 😞 Well , This is definitely a Government Conspiracy lysol labels as proof the government knew in advance about the Corona virus .

Corona Virus is nothing new.



- **How to measure algorithmic bias?**

1. Spread of fact-checked claims
on Twitter

- New dataset
- Popularity of true and false claims
- Comparison with 2 additional datasets
- Unbiased comparisons

- **How to audit algorithms?**

2. Search Engine/LLM
Answers to Political Queries

- Army of bots
- Neutral queries before the 2024 EU Parliament Elections
- Comparison with Media, Polls, Results

1. Spread of fact-checked claims on Twitter



Fact-Checking

67K claims

2015 - 2023



~ 3M
tweets



450 000
cascades

Ana Vranič



Lília Perfeito



Paulo Almeida

1. Spread of fact-checked claims on Twitter



Fact-Checking

~ 3M tweets

450 000 cascades

67K claims

2015 - 2023



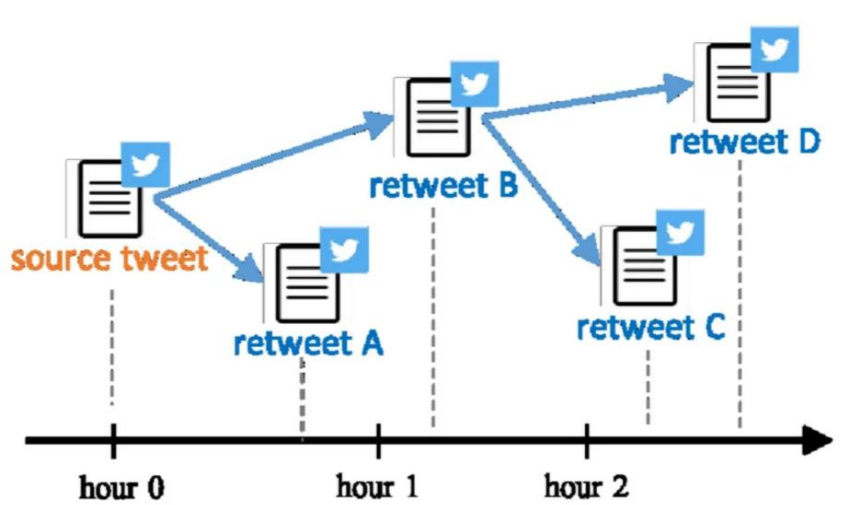
Ana Vranič



Lília Perfeito



Paulo Almeida



1. Spread of fact-checked claims on Twitter



Fact-Checking

~ 3M tweets

450 000 cascades

67K claims

2015 - 2023



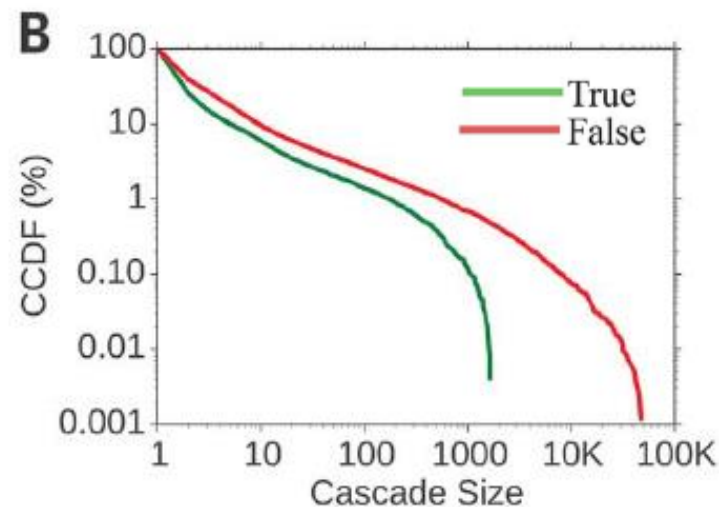
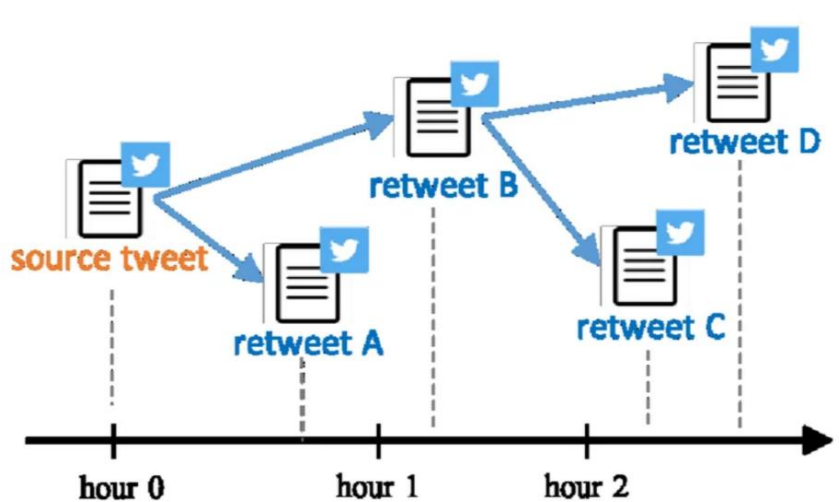
Ana Vranič



Lília Perfeito



Paulo Almeida



1. Spread of fact-checked claims on Twitter



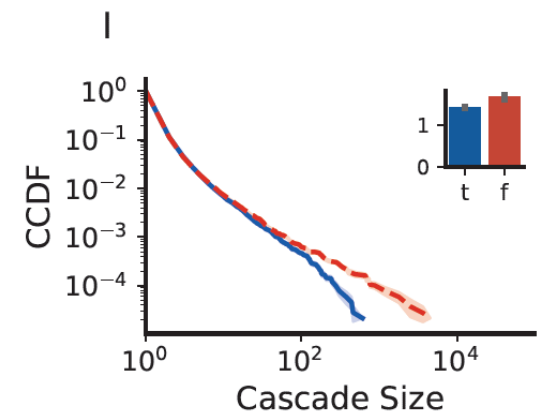
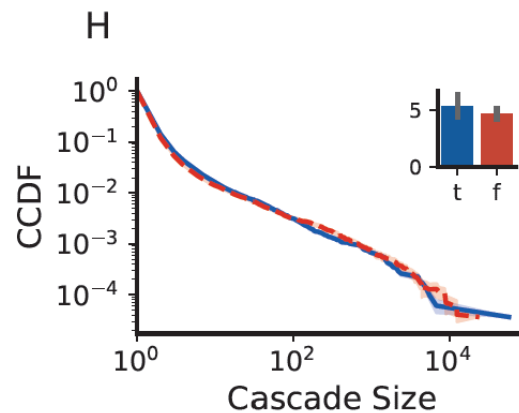
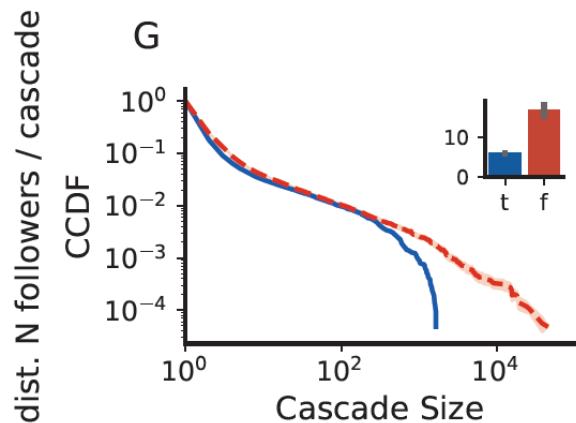
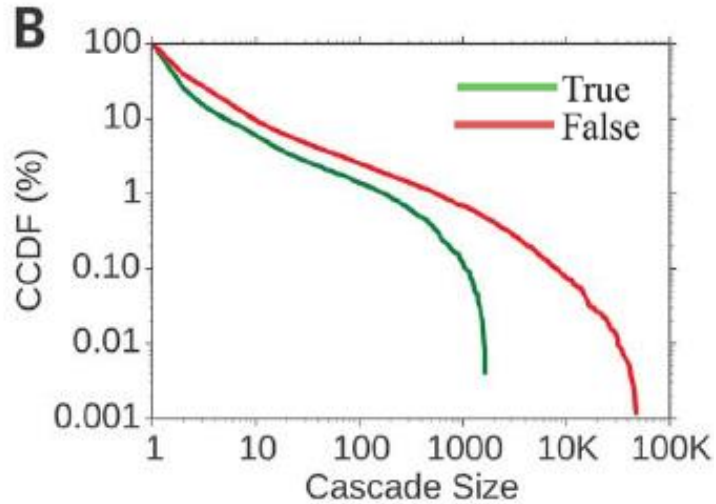
Ana Vranič



Lília Perfeito



Paulo Almeida



- **How to measure algorithmic bias?**

1. Spread of fact-checked claims
on Twitter

- New dataset
- Popularity of true and false claims
- Comparison with 2 additional datasets
- Unbiased comparisons

- **How to audit algorithms?**

2. Search Engine/LLM
Answers to Political Queries

- Army of bots
- Neutral queries before the 2024 EU Parliament Elections
- Comparison with Media, Polls, Results

2. Search Engine/LLM Answers to Political Queries

Íris Damião



José Reis



Paulo Almeida



I look like a
human!

Location,
Language,
Profile...

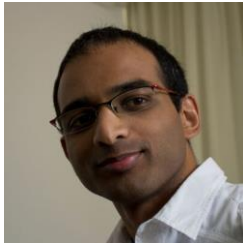
2. Search Engine/LLM Answers to Political Queries



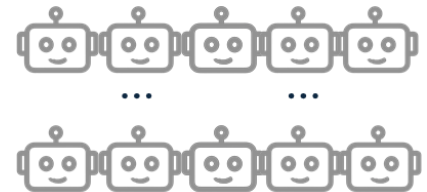
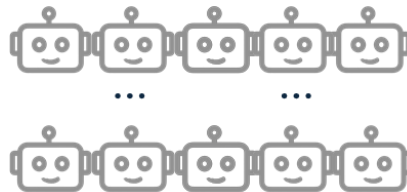
Íris Damião



José Reis



Paulo Almeida



Location,
Language,
Profile...

2. Search Engine/LLM Answers to Political Queries



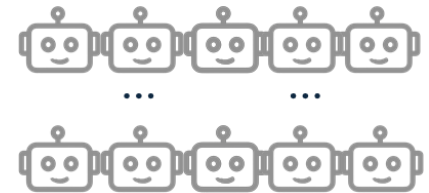
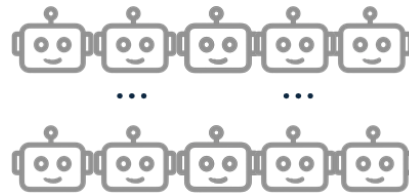
Íris Damião



José Reis



Paulo Almeida



I look like a human!

Location,
Language,
Profile...



2. Search Engine/LLM Answers to Political Queries

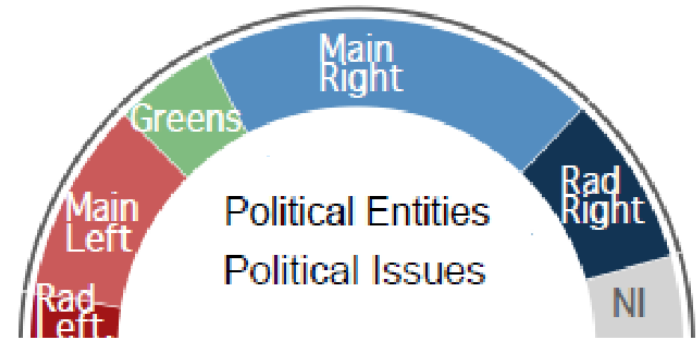
Íris Damião



José Reis

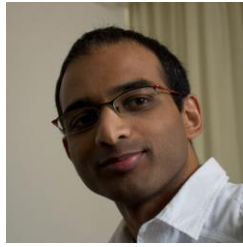


Paulo Almeida



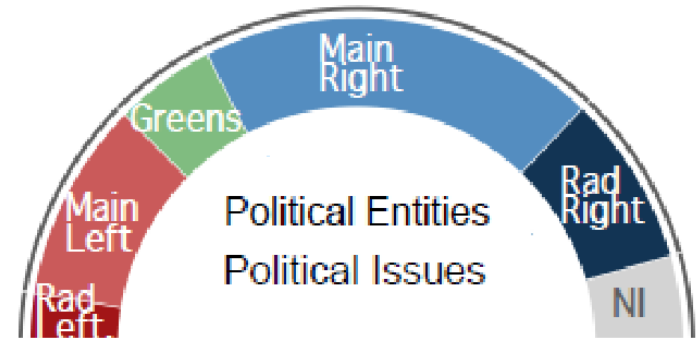
2. Search Engine/LLM Answers to Political Queries

Íris Damião



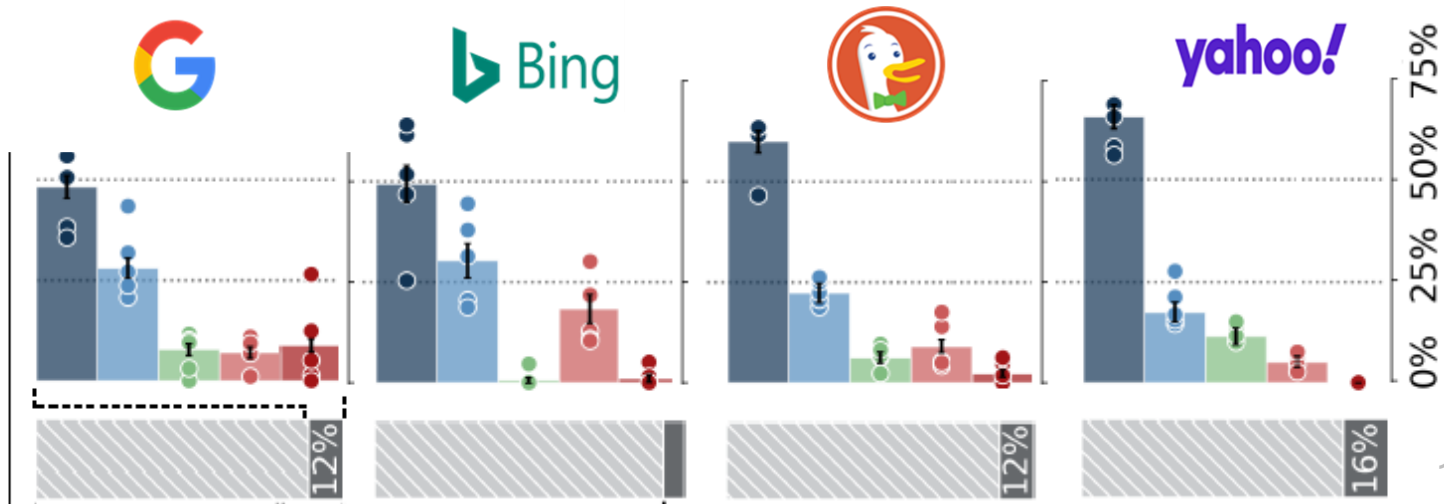
José Reis

Paulo Almeida



EU Parliament Election

Political Entities



1. Spread of fact-checked
claims on Twitter

- Differences between true and false news are not universal and depend strongly on
 - The time period,
 - The content
 - Dataset
- Imbalances can bias conclusions about whether false or true content spreads more
- Applying sampling and balancing methods often reduces or even reverses initially found effects

1. Spread of fact-checked
claims on Twitter

- Differences between true and false news are not universal and depend strongly on
 - The time period,
 - The content
 - Dataset
- Imbalances can bias conclusions about whether false or true content spreads more
- Applying sampling and balancing methods often reduces or even reverses initially found effects

2. Search Engine/LLM Answers to
Political Queries

- Traditional media give unbalanced attention to radical right entities
- This cannot be explained by electoral outcomes or predictions
- This is “carried” and amplified by search engines and LLMs
- Citizens making broad electoral queries are disproportionately directed to radical right content

DIGITAL SERVICES ACT

**Vetted
researchers
(40.4)**

Reserved for researchers
fulfilling specific
conditions, gives access
to non-public data to study
systemic risks in the EU

*Delegated Act published
in July 2025*

**‘Qualified’ Researchers
(40.12)**

For researchers and
CSOs, gives access to
public data with tools
such as APIs, and protects
those who use automated
access means (e.g.
scraping) to study
systemic risks in the EU

No DSC intervention

Strengths

Solid multidisciplinary team; Demonstrated capacity to attract high quality researchers, competitive funding and publications in high impact journals.

Weaknesses

Limited history of collaboration with researchers at LIP or with LIP's main research partners; Breadth of interests, that limit the creation of focused, critical mass.

Opportunities

Ongoing research receiving wide (both popular and scientific) interest; Ample space to establish SPAC as a leader in a very novel research field; High interest in future collaborations both inside and outside of LIP.

Threats

Very competitive research areas, particularly misinformation and digital epidemiology (and bias detection); Main funding source ending in 2026

Thank you!





Fact-Checking

67K claims

2015 - 2023

<https://dl.acm.org/doi/abs/10.1145/3746275.3762201>

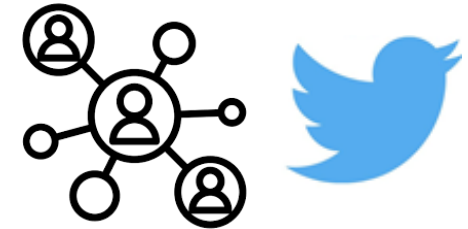
<https://zenodo.org/records/16942245>

- Expert
- mTurk workers
- ChatGPT



Structured datasets

	politics
	health
	environment
	science
	society
	other



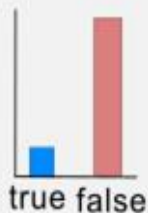
Online spread

450 000 cascades
~ 3M tweets



dataset	Vosoughi News			Global Claims			Truth Seeker		
timespan	2008-2016			2015-2022			2008-2022		
Topics	US Poliiics			Diverse Topics			Politics and Health		
veracity	TRUE	FALSE	TOTAL	TRUE	FALSE	TOTAL	TRUE	FALSE	TOTAL
N claims	490	1 699	2 189	136	2 648	2 784	578	477	1 055
N cascades	24 409	82 605	107 014	29 622	421 510	451 132	66 081	62 377	128 458
N tweets	202 348	3 525 344	3 727 692	261 915	2 600 162	2 862 077	434 316	510 876	945 192
N users	20 266	63 589	83 855	28 120	298 854	326 974	139	45	184

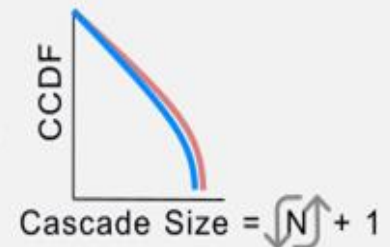
B Unbalanced dataset



sampling
claims/
cascades



unbiased
comparison



How sensitive are the observed differences to sample size and other sampling effects?

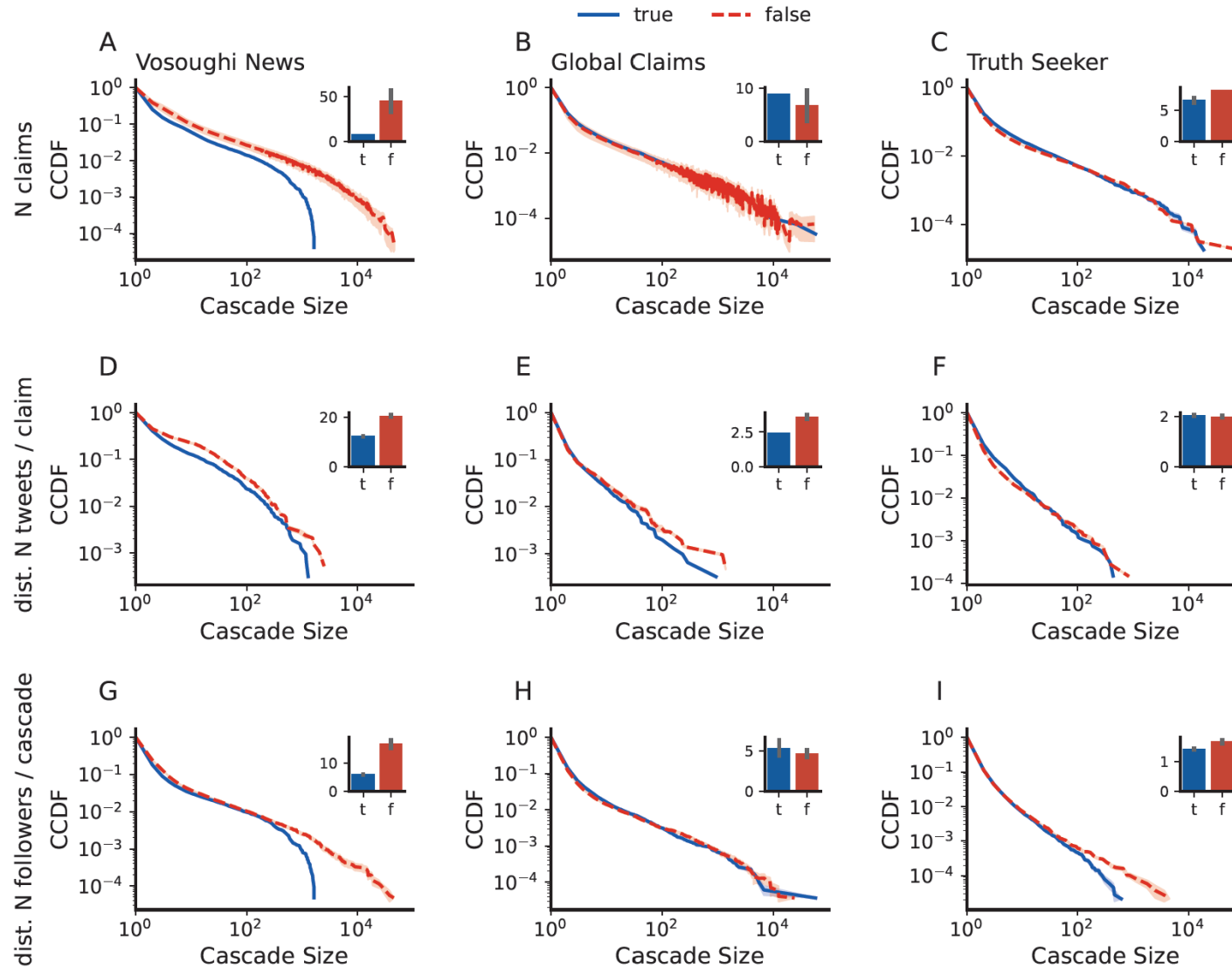
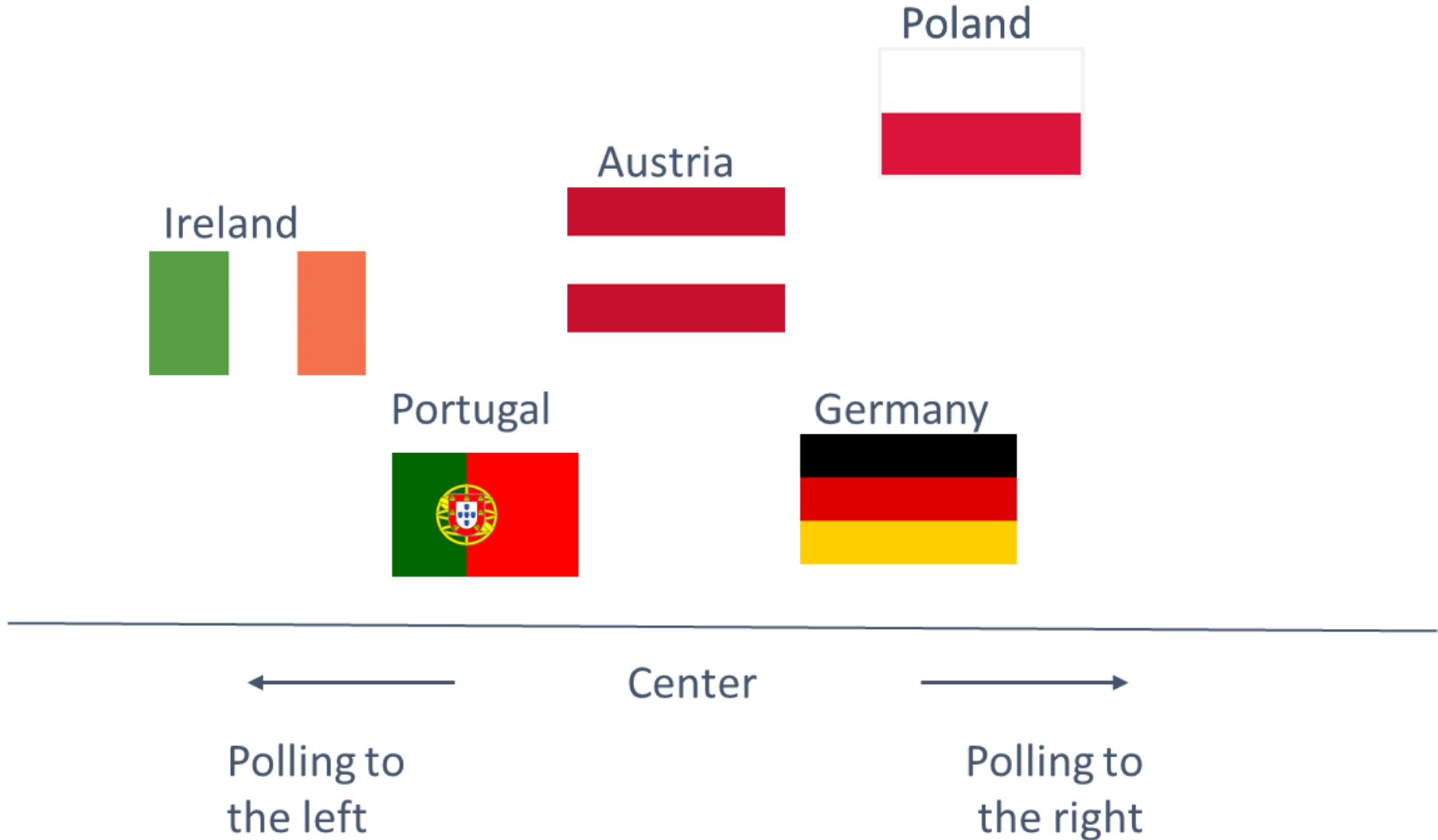


Fig. 4. Effect of sampling strategies on cascade-size distributions for the Fake News (left column), Global Claims (middle column), and Truth Seeker datasets (right column), for true (blue lines) and false (red dotted lines) cascades. Each row corresponds to a different sampling method: (top) random sampling of claims, (middle) matching true and false claims by the number of claims, and (bottom) matching by user follower distributions. Panels show the average complementary cumulative distribution functions (CCDFs) of cascade sizes across 10000 samples for true and false claims under each sampling method. Insets display the corresponding mean cascade sizes.



Thank you!

<https://socialcomplexity.eu>

spac@lip.pt





S. Vosoughi et al., 2018

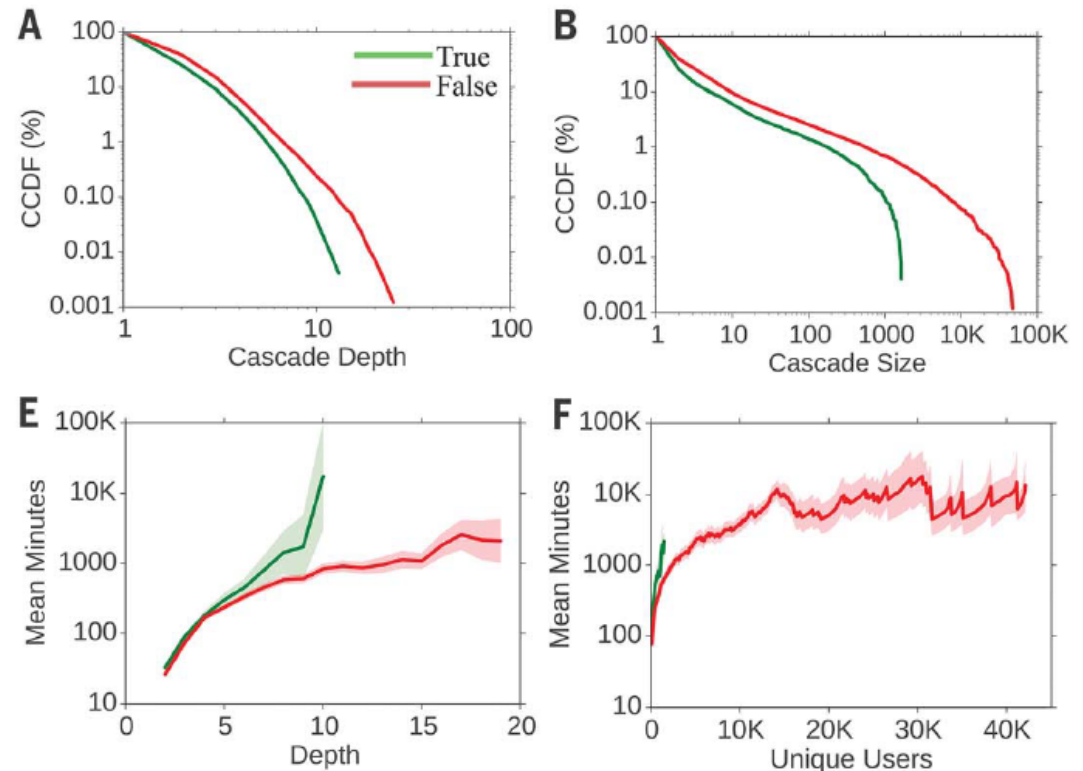


Fig. 2. Complementary cumulative distribution functions (CCDFs) of true and false rumor cascades. (A) Depth. (B) Size. (C) Maximum breadth. (D) Structural virality. (E and F) The number of minutes it takes for true and false rumor cascades to reach any (E) depth and (F) number of unique Twitter users. (G) The number of unique Twitter

“Falsehood spread farther, faster, deeper, and more broadly than the truth”

“Falsehood spread farther, faster, deeper, and more broadly than the truth”



Confirmed by others

- **On Weibo in China and Twitter in Japan** fake news propagate differently than true Zhao (2020)
- **On Twitter** Larger false cascades, driven by emotional response Pröllochs (2021)
- **Tweets in German** Unreliable news are more retweeted (Garcia 2025)
- **False news are retweeted more on different platforms** (Mosleh 2025)

English



Showing otherwise

- **Matching on cascade size distribution differences in cascade properties disappear** (Juul 2021)
- **On Twitter** - Limited reach - “Overall, disinformation outlets obtain less engagement than others” Cinelli (2020)
- **On Facebook** - vaccine hesitancy, “Small exposure of fact-checked posts, and non-flagged data can be more harmful” Allen (2024)
- **On Twitter** - “From 2016 to 2020 volume of fake news decreased, but echo chambers increased” Flamino (2023)

~Political News

2006 - 2016



Fact-Checking

67K claims

2015 - 2023

<https://dl.acm.org/doi/abs/10.1145/3746275.3762201>

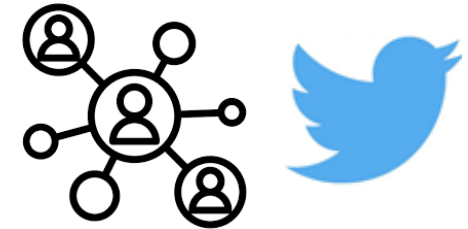
<https://zenodo.org/records/16942245>

- Expert
- mTurk workers
- ChatGPT



Structured datasets

	politics
	health
	environment
	science
	society
	other

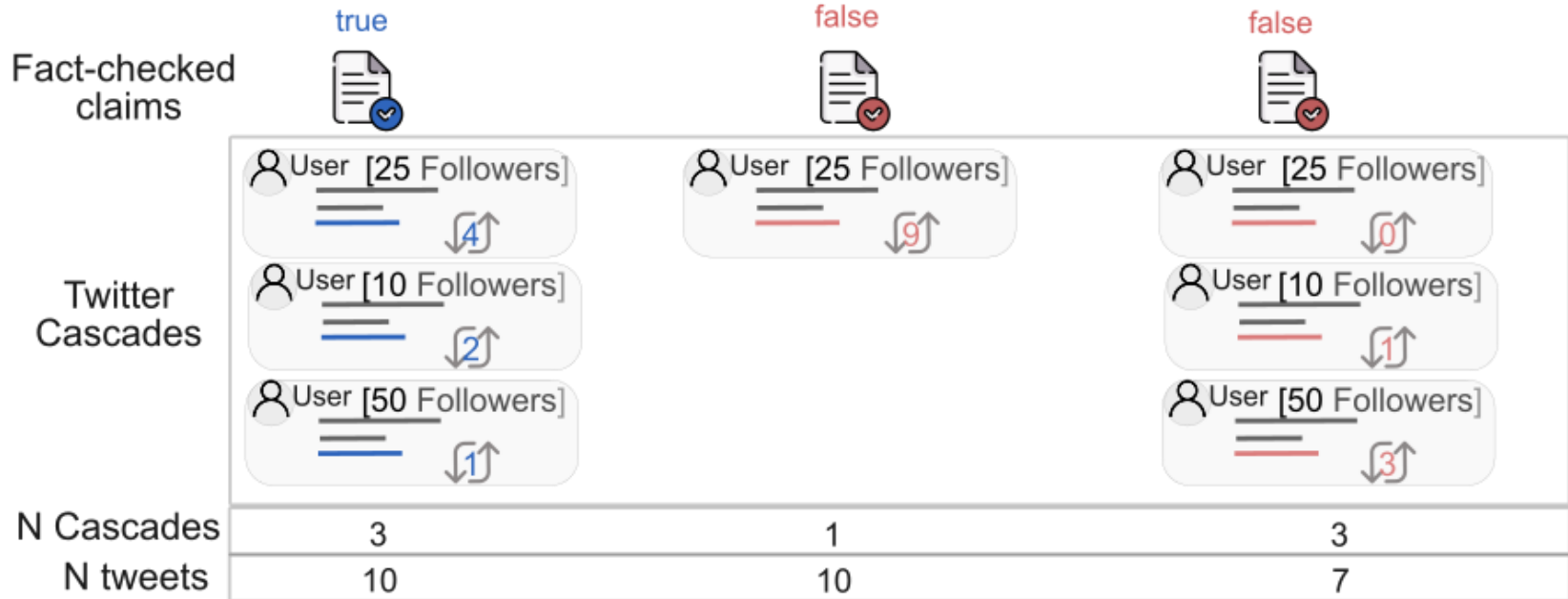


Online spread

450 000 cascades
~ 3M tweets

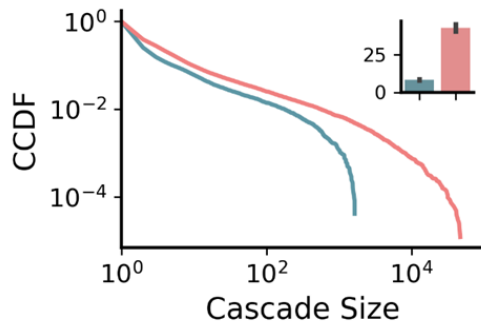


A

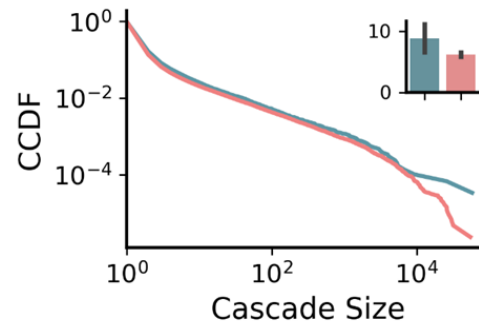


dataset	Vosoughi News			Global Claims			Truth Seeker		
timespan	2008-2016			2015-2022			2008-2022		
Topics	US Poliics			Diverse Topics			Politics and Health		
veracity	TRUE	FALSE	TOTAL	TRUE	FALSE	TOTAL	TRUE	FALSE	TOTAL
N claims	490	1 699	2 189	136	2 648	2 784	578	477	1 055
N cascades	24 409	82 605	107 014	29 622	421 510	451 132	66 081	62 377	128 458

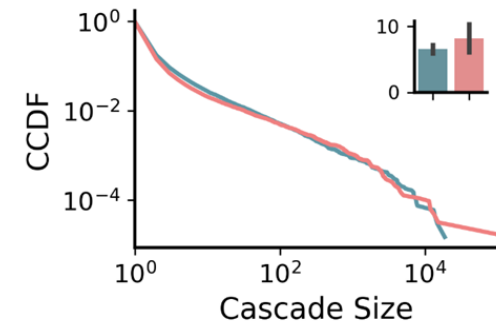
Vosoughi News



Global Claims



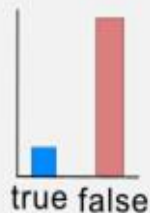
Truth Seeker



- Vosoughi News (2008-2016)
 - False spread more than true news
- Global Claims (2014-2022)
 - True news spread more, distributions mostly differ in tail
- Truth Seeker (2008-2022)
 - False news have heavier tail and appear produce larger cascade sizes on average

dataset	Vosoughi News			Global Claims			Truth Seeker		
timespan	2008-2016			2015-2022			2008-2022		
Topics	US Poliiics			Diverse Topics			Politics and Health		
veracity	TRUE	FALSE	TOTAL	TRUE	FALSE	TOTAL	TRUE	FALSE	TOTAL
N claims	490	1 699	2 189	136	2 648	2 784	578	477	1 055
N cascades	24 409	82 605	107 014	29 622	421 510	451 132	66 081	62 377	128 458
N tweets	202 348	3 525 344	3 727 692	261 915	2 600 162	2 862 077	434 316	510 876	945 192
N users	20 266	63 589	83 855	28 120	298 854	326 974	139	45	184

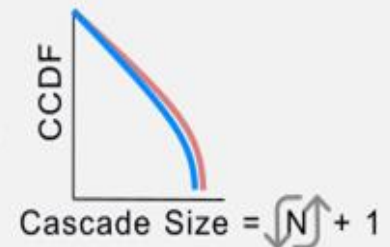
B Unbalanced dataset



sampling
claims/
cascades



unbiased
comparison



How sensitive are the observed differences to sample size and other sampling effects?

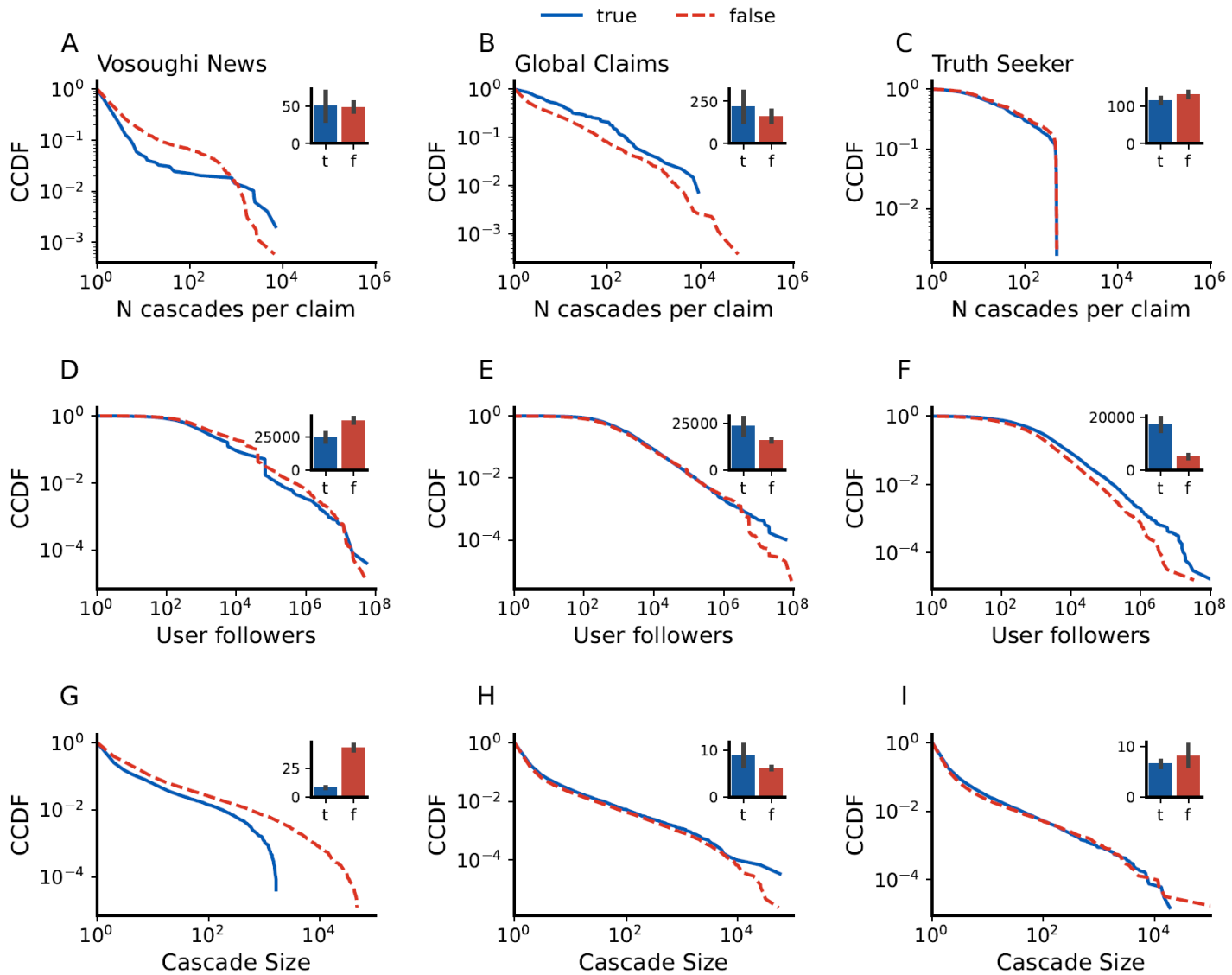
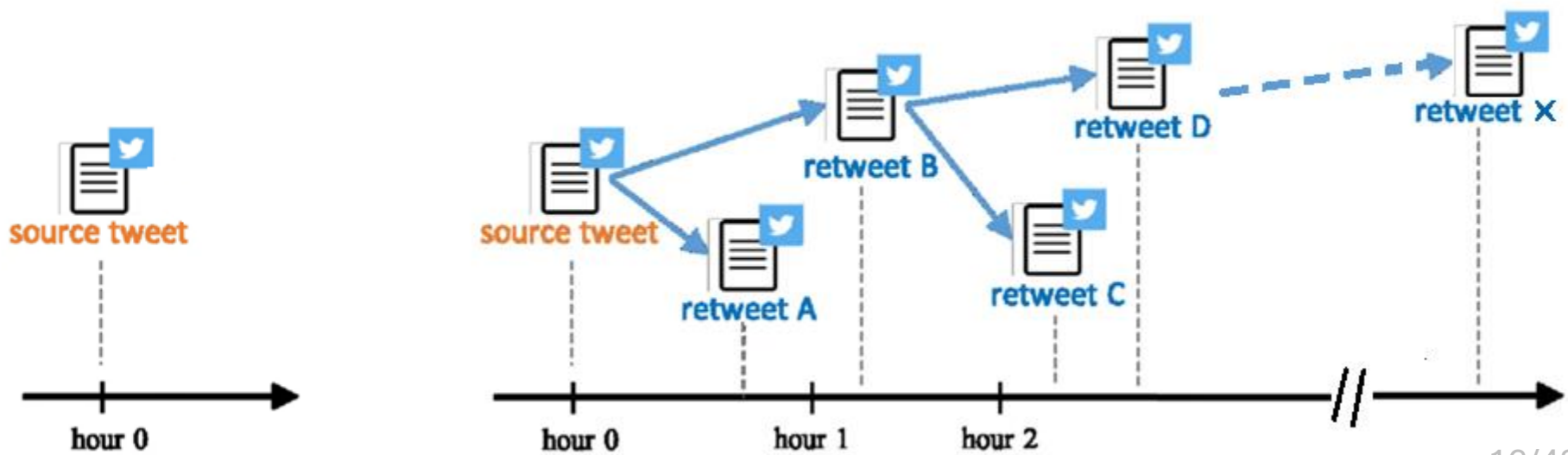
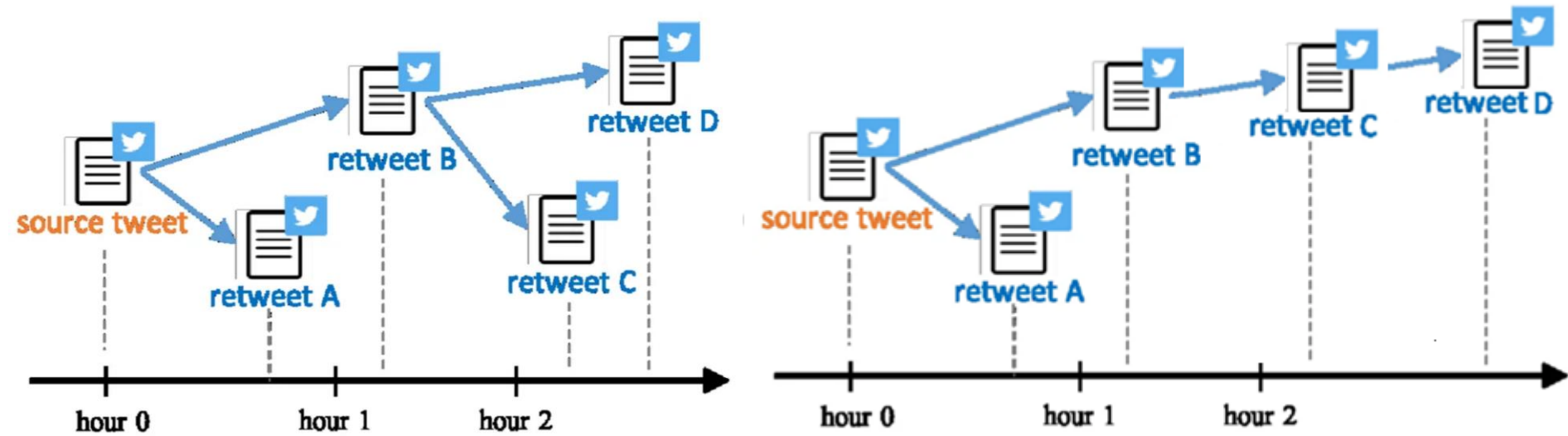
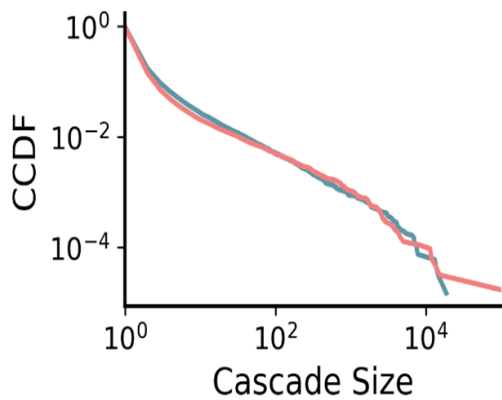
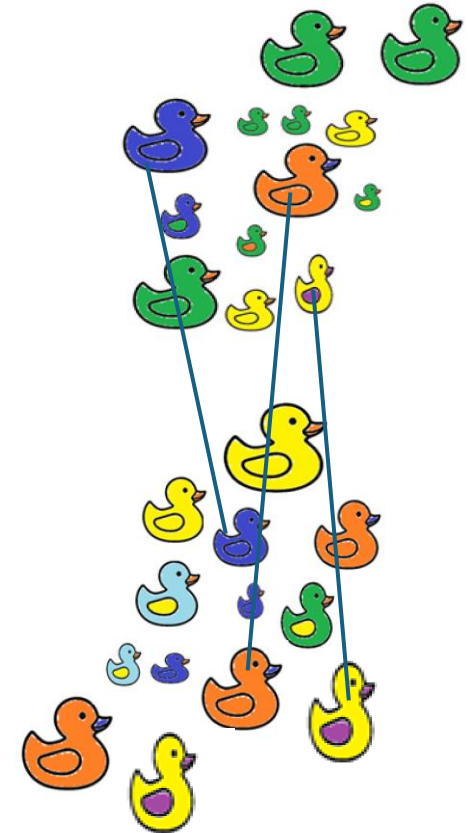
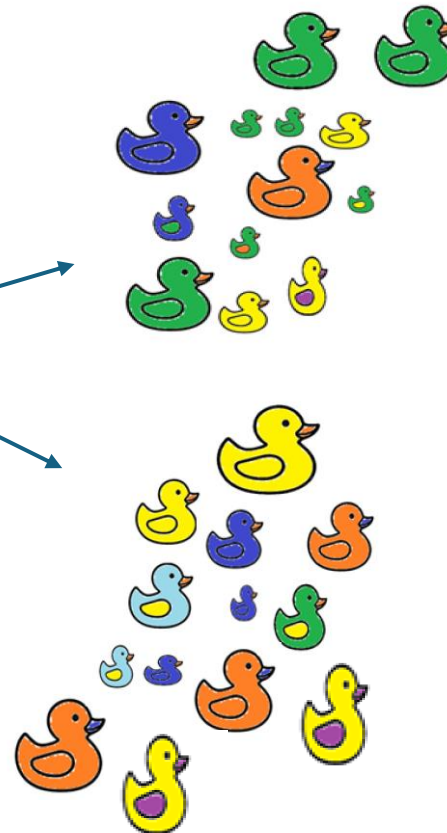
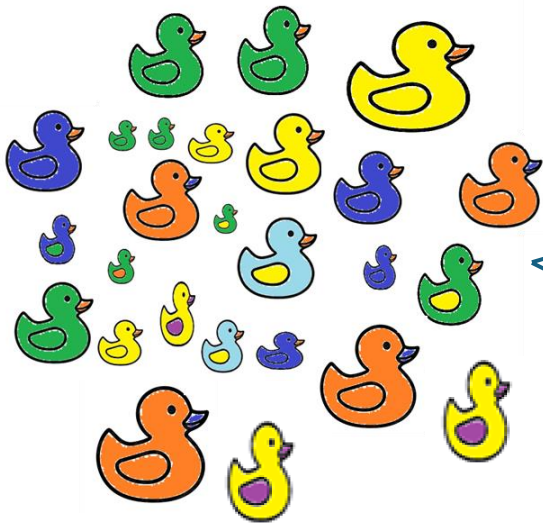


Fig. 3. Distributions of cascade characteristics for the Fake News (left column), Global Claims (middle column), and Truth Seeker datasets (right column), for true (blue lines) and false (red dotted lines) cascades. Complementary cumulative distribution functions (CCDFs) of the number of cascades per claim panels A, B and C). CCDFs of follower counts for users who initiated each cascade (panels D, E and F). CCDFs of cascade sizes, measured as the total number of tweets within each cascade (panels G, H and I). Insets in each panel show the corresponding mean values for true and false cascades.



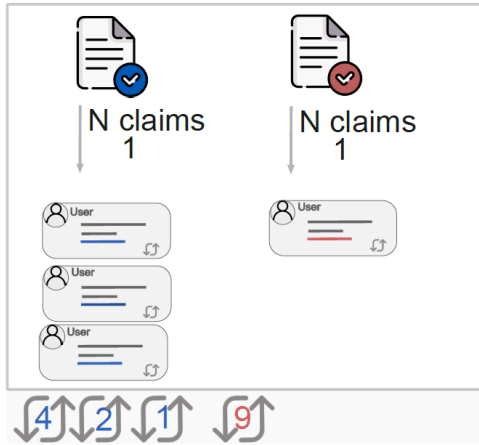
Random sampling

Sample Matching

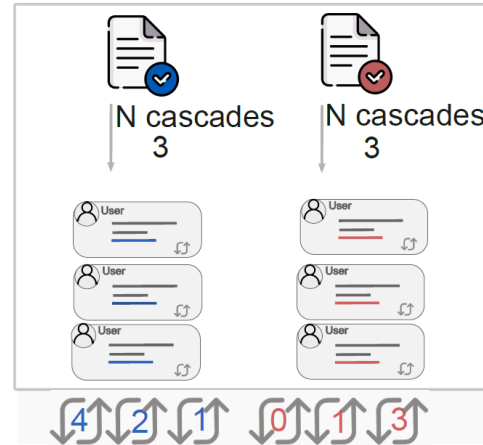


Matching on Claim properties

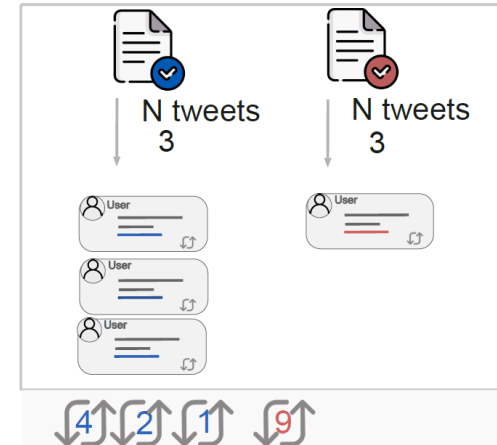
a) same N of claims



b) same distributions of N cascades per claim

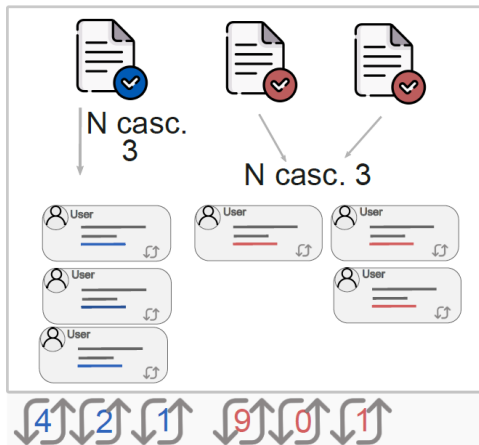


c) same distributions of N tweets per claim

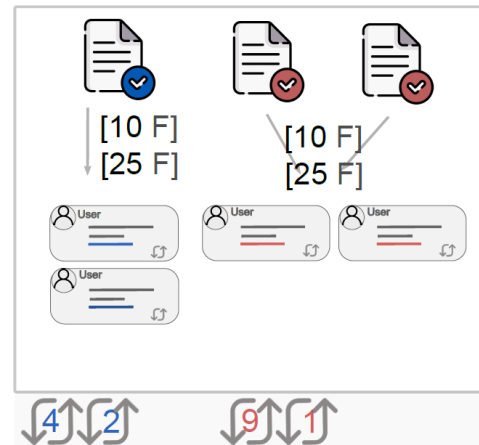


Matching on Cascade properties

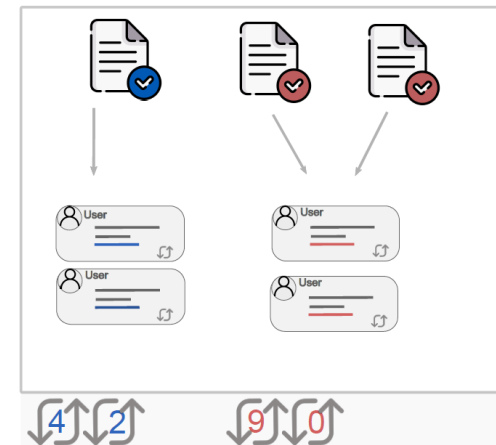
d) same N of cascades



e) same distributions of N followers per cascade



f) propensity scores matching on N casc., N tweets., N followers



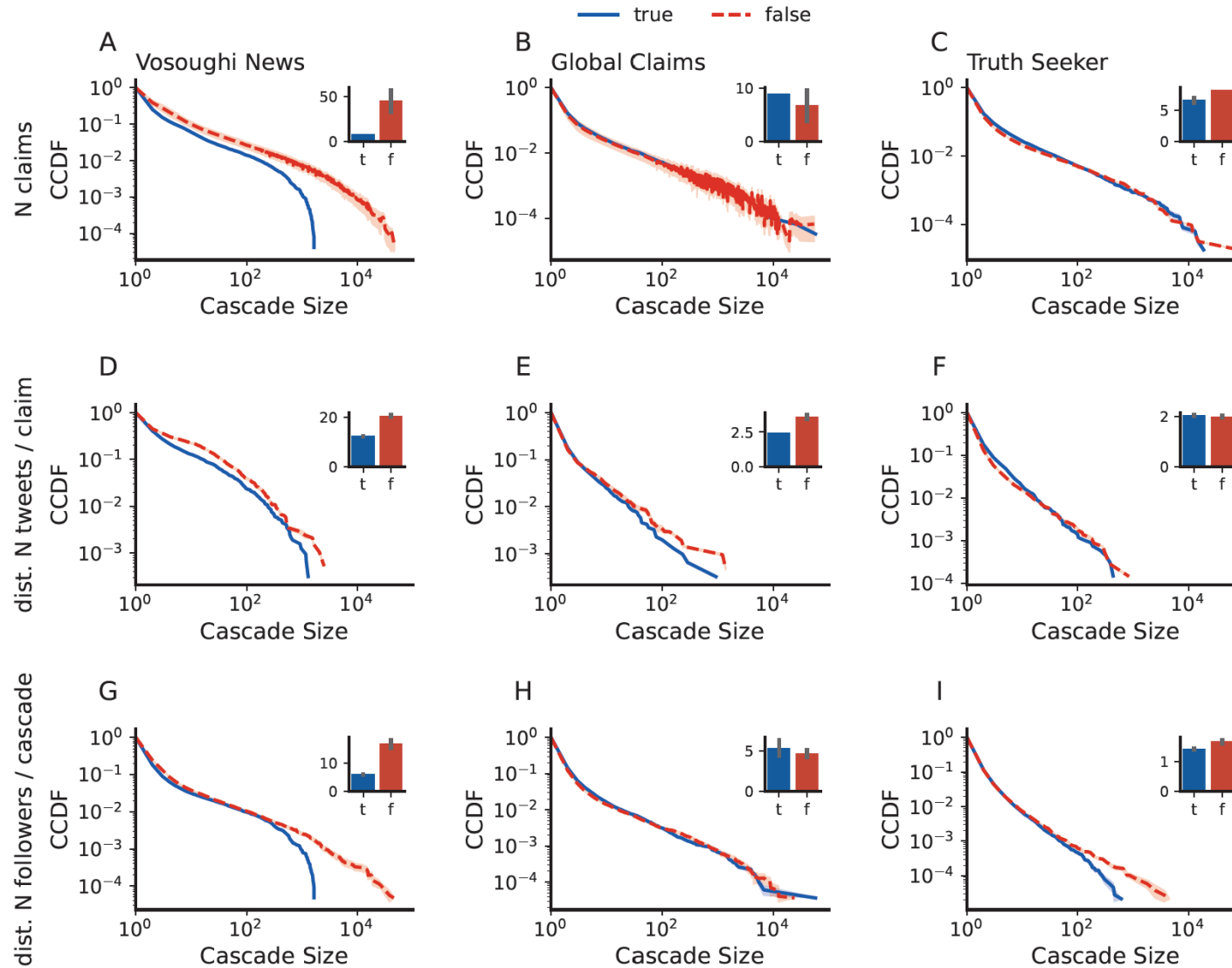
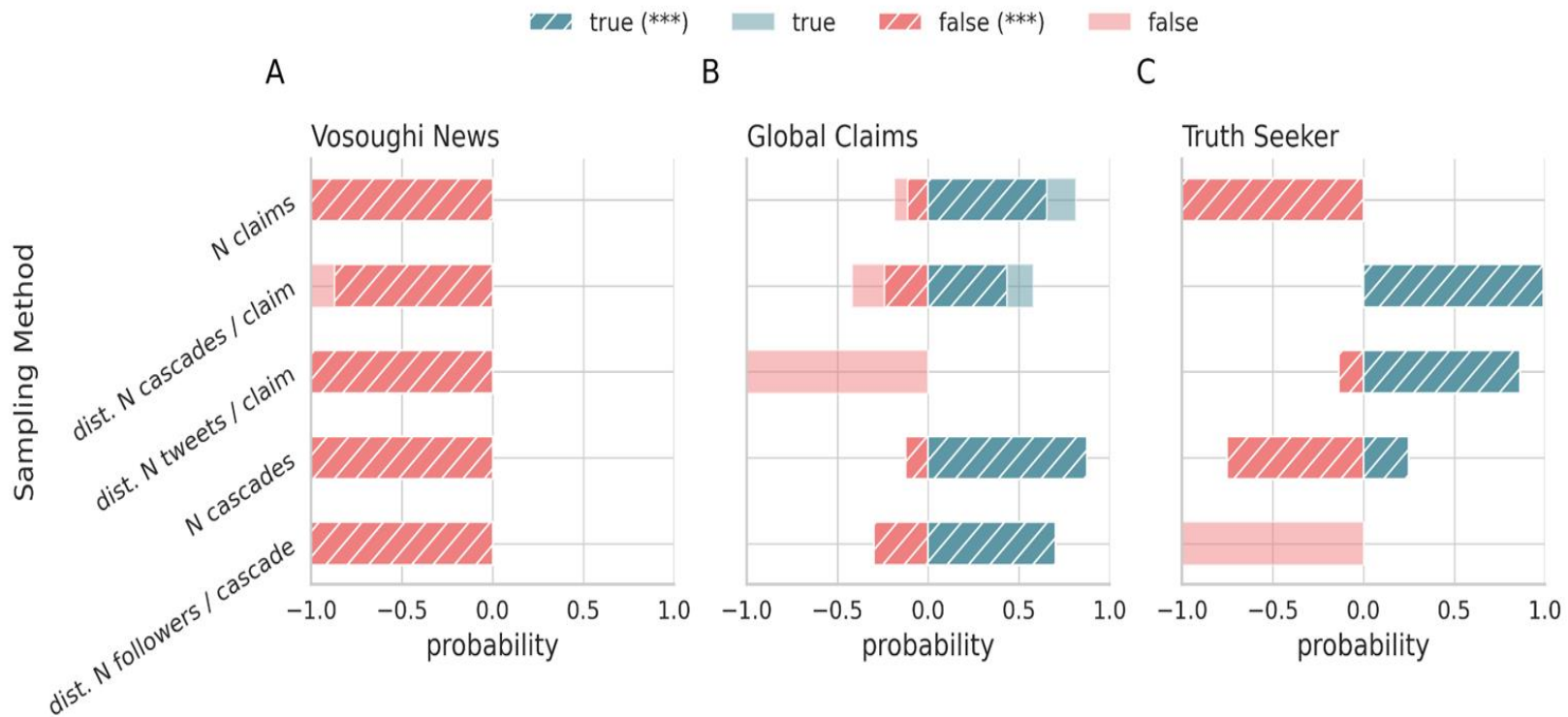
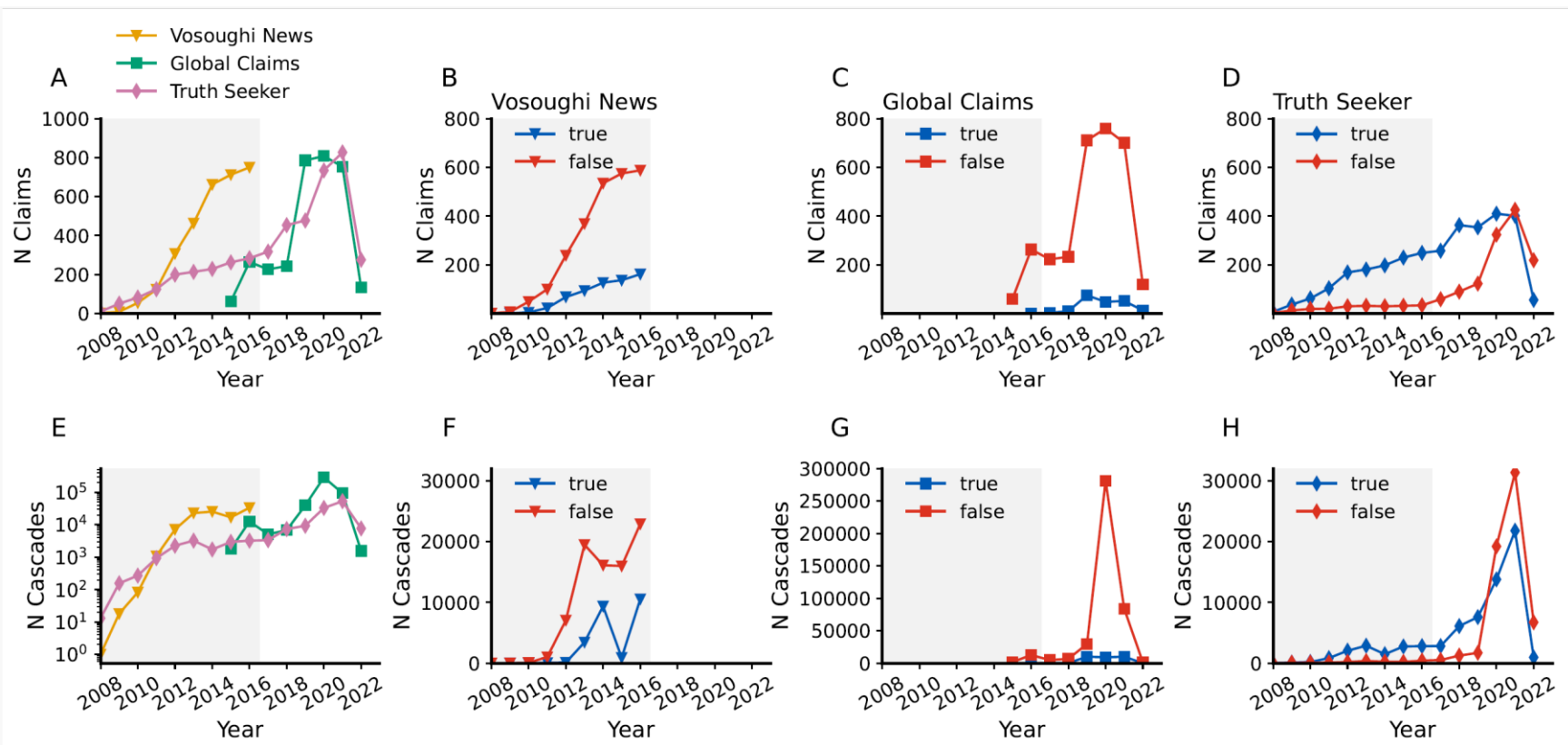


Fig. 4. Effect of sampling strategies on cascade-size distributions for the Fake News (left column), Global Claims (middle column), and Truth Seeker datasets (right column), for true (blue lines) and false (red dotted lines) cascades. Each row corresponds to a different sampling method: (top) random sampling of claims, (middle) matching true and false claims by the number of claims, and (bottom) matching by user follower distributions. Panels show the average complementary cumulative distribution functions (CCDFs) of cascade sizes across 10000 samples for true and false claims under each sampling method. Insets display the corresponding mean cascade sizes.

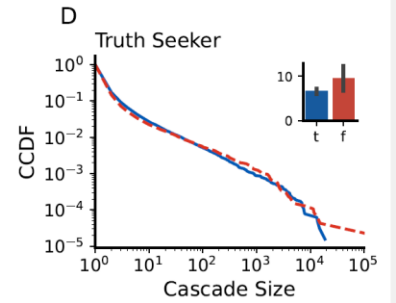
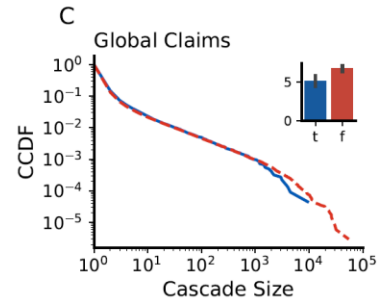
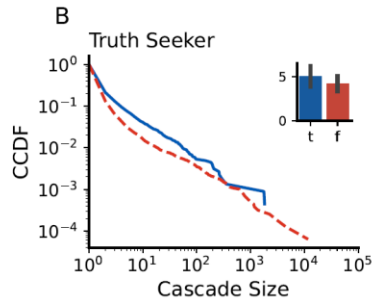
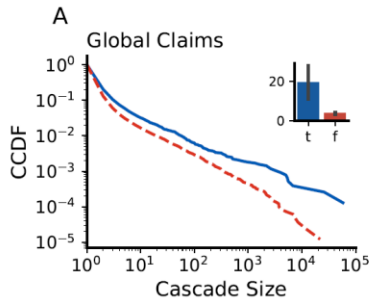
- Probability across 1000 bootstrap samples that true (blue bars) or false (red bars) have larger mean cascade size
- Dashed bars - statistically significant comparisons between true and false (KS test)



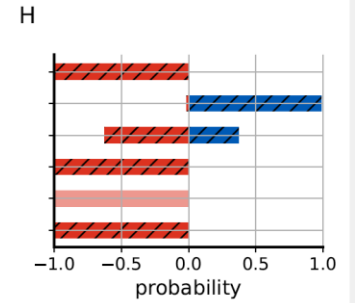
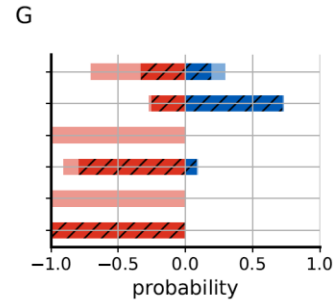
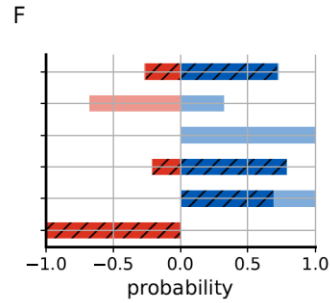
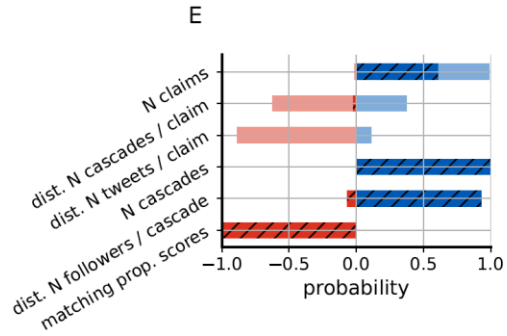


Covid claims — true — false

Non-covid claims

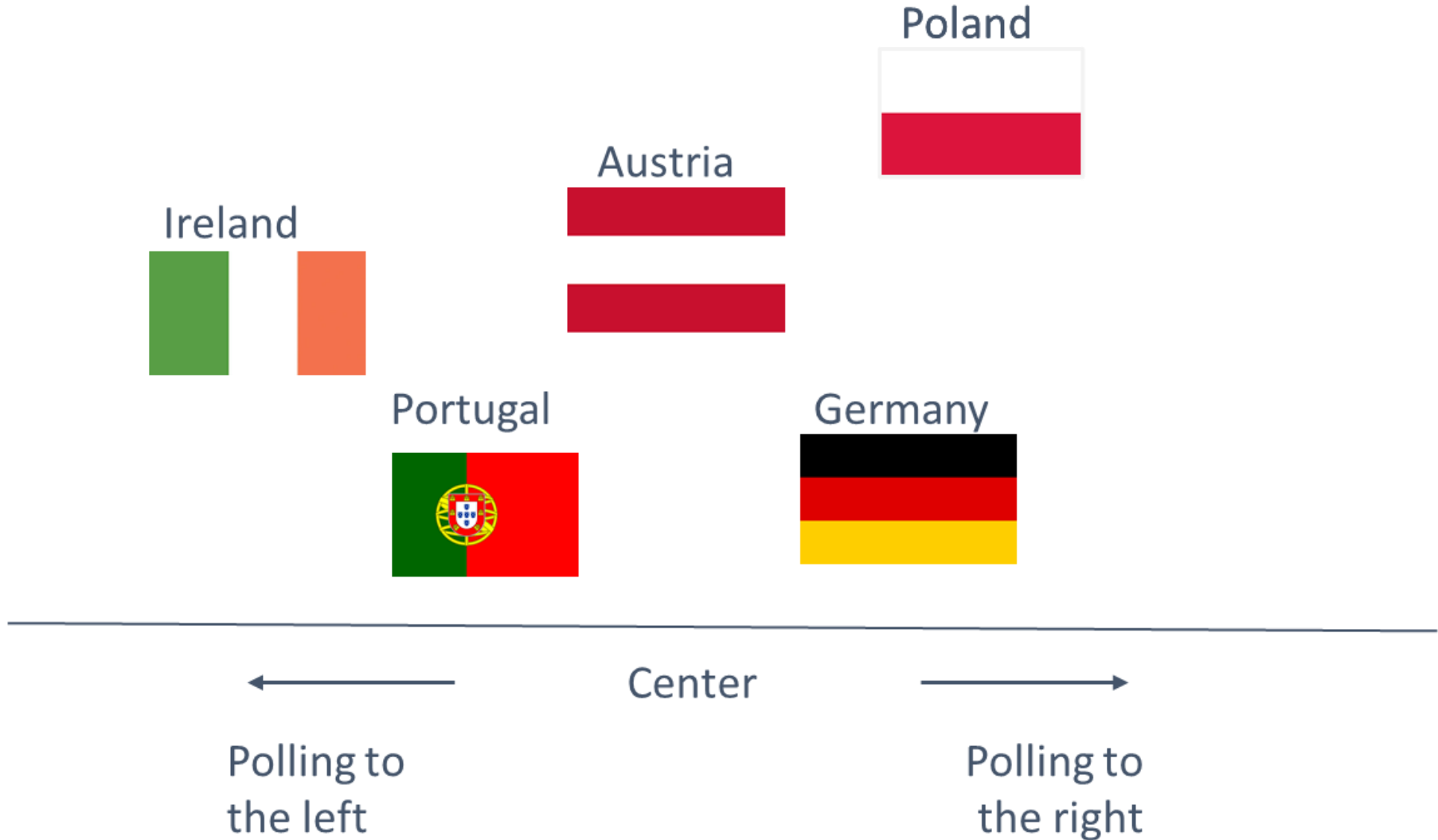


■ true (***) ■ true ■ false (***) ■ false

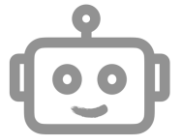




What is the best party
Who should I vote for
European Elections



Location,
Language,
Profile...

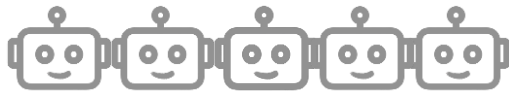


I look like a human!

OpenWPM webcrawler



AT



...

...

PT



Proxy system

(kindly provided by
bright data)



1. News Collection

Media Sources choice

+

Keywords selection



Media Cloud



SemRush
Top-20 Newspapers AT

1. YouTube.com ✗
2. orf.at ✓
3. krone.at ✓
- ⋮
20. focus.de ✓

EU - elections
(e.g., "European elections")

G1 ●

General elections
("Elections", "Polls")

G2 ●

EU - politics
(e.g., "MEP", "EU Parliament")

G3 ●

+

AND/OR




European Election-related News DB

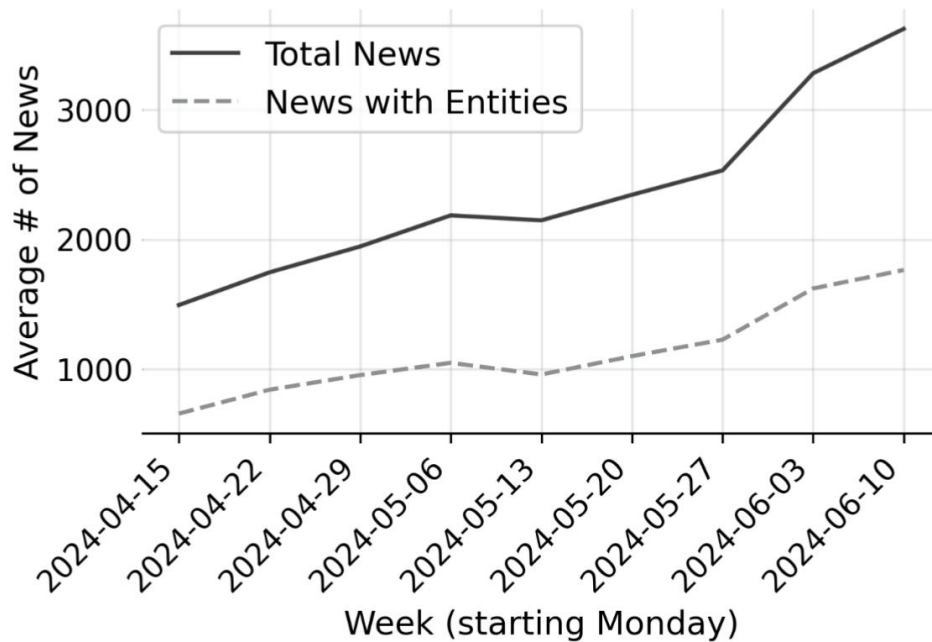
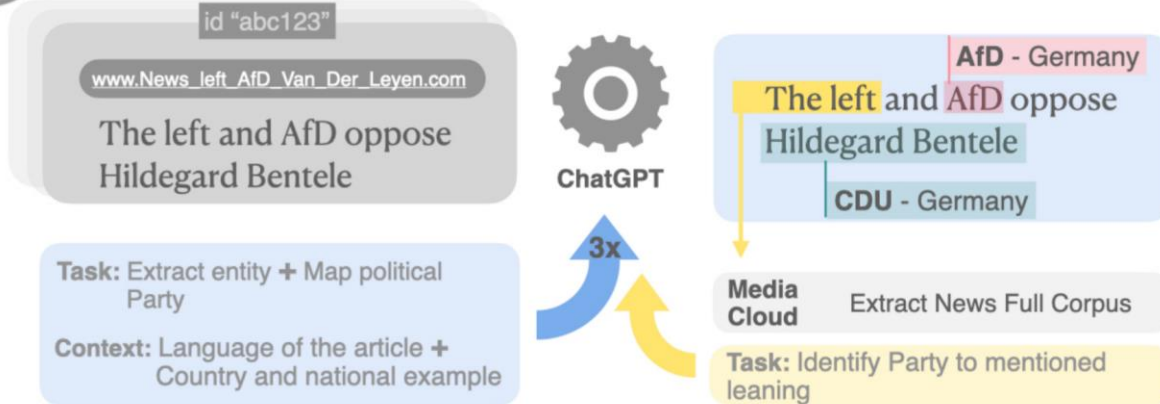
Translated via Google Translate, ChatGPT and native speakers

Country	Total News
Austria	3,480
Germany	3,861
Ireland	4,449
Poland	3,297
Portugal	6,441
Total	21,528

2. Entities Extraction



Entity Extraction for all types of mentions

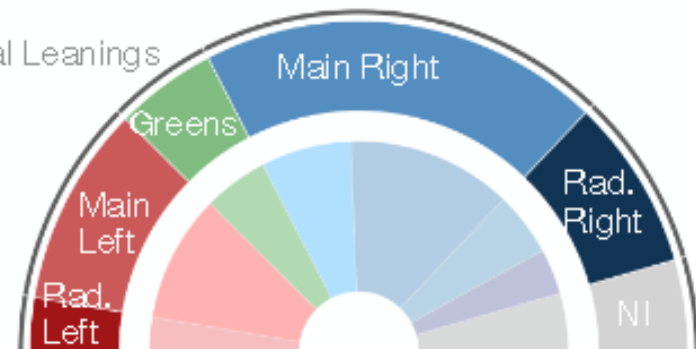
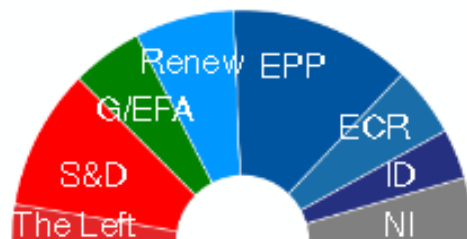


id	entity	mapping
"abc123"	Hildegard Bentele	CDU - Germany
"abc123"	AfD	AfD - Germany
"abc123"	The left	Die Linke - Germany

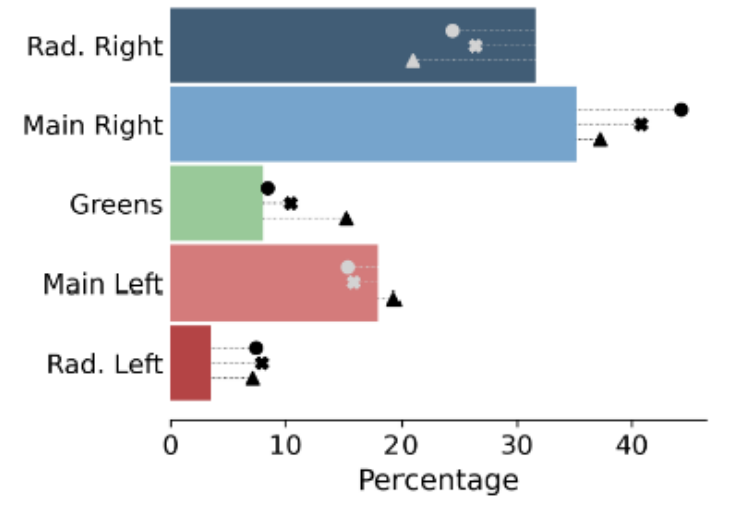
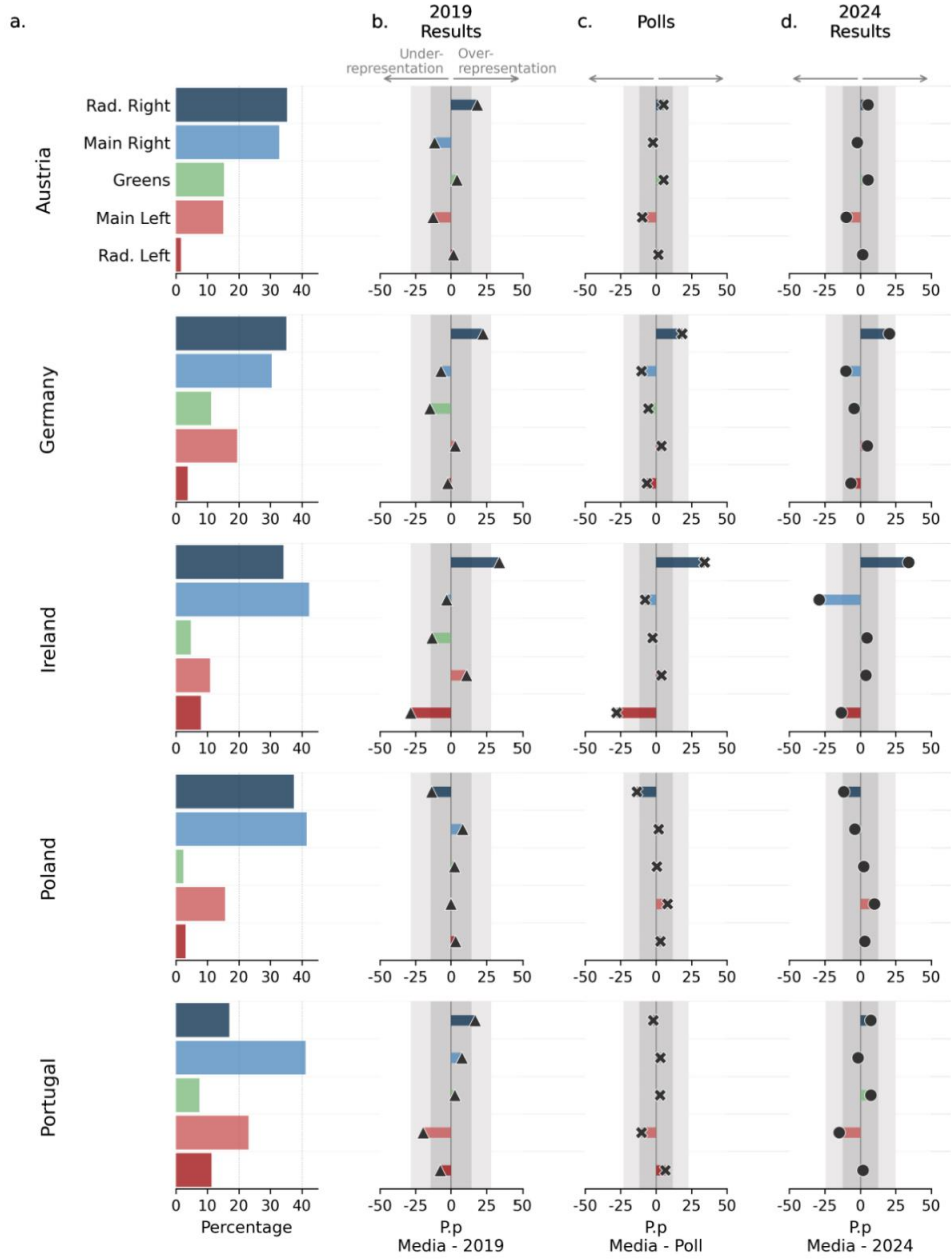


Mapping to
EU Groups

Mapping to
Main Political Leanings





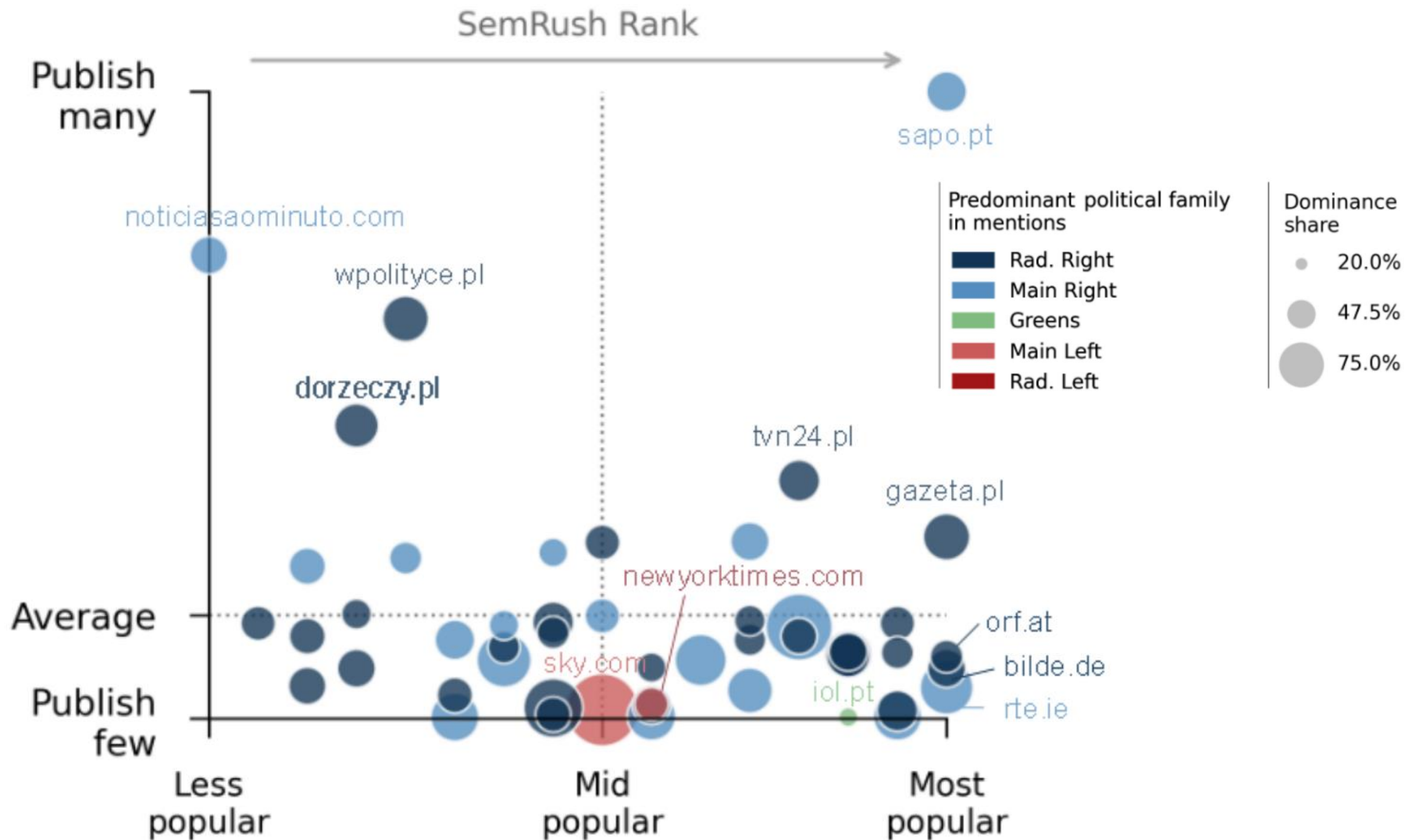


▲ 2019 Seats * Polls ● 2024 Seats

Rad. Left Main Right

Main Left Rad. Right

Greens

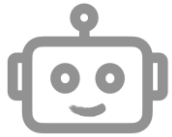


- Traditional media give unbalanced attention to radical right entities
- This cannot be explained by electoral outcomes or predictions



What is the best party
Who should I vote for
European Elections

Location,
Language,
Profile...

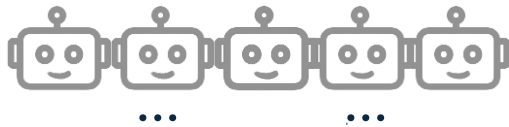


I look like a human!

OpenWPM webcrawler



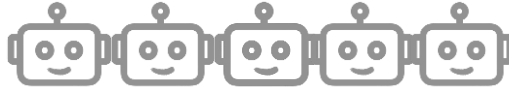
AT



...

...

PT



Proxy system

(kindly provided by
bright data)



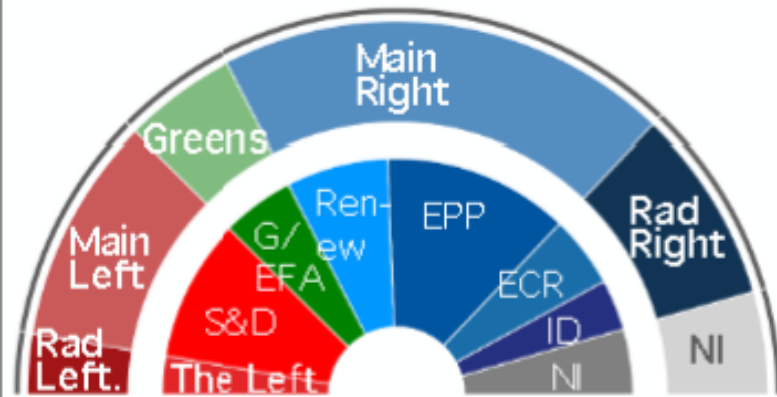


2. Entities Extraction

Political Entities

- politicians
- political parties
- spectrum

3. Leaning Identification



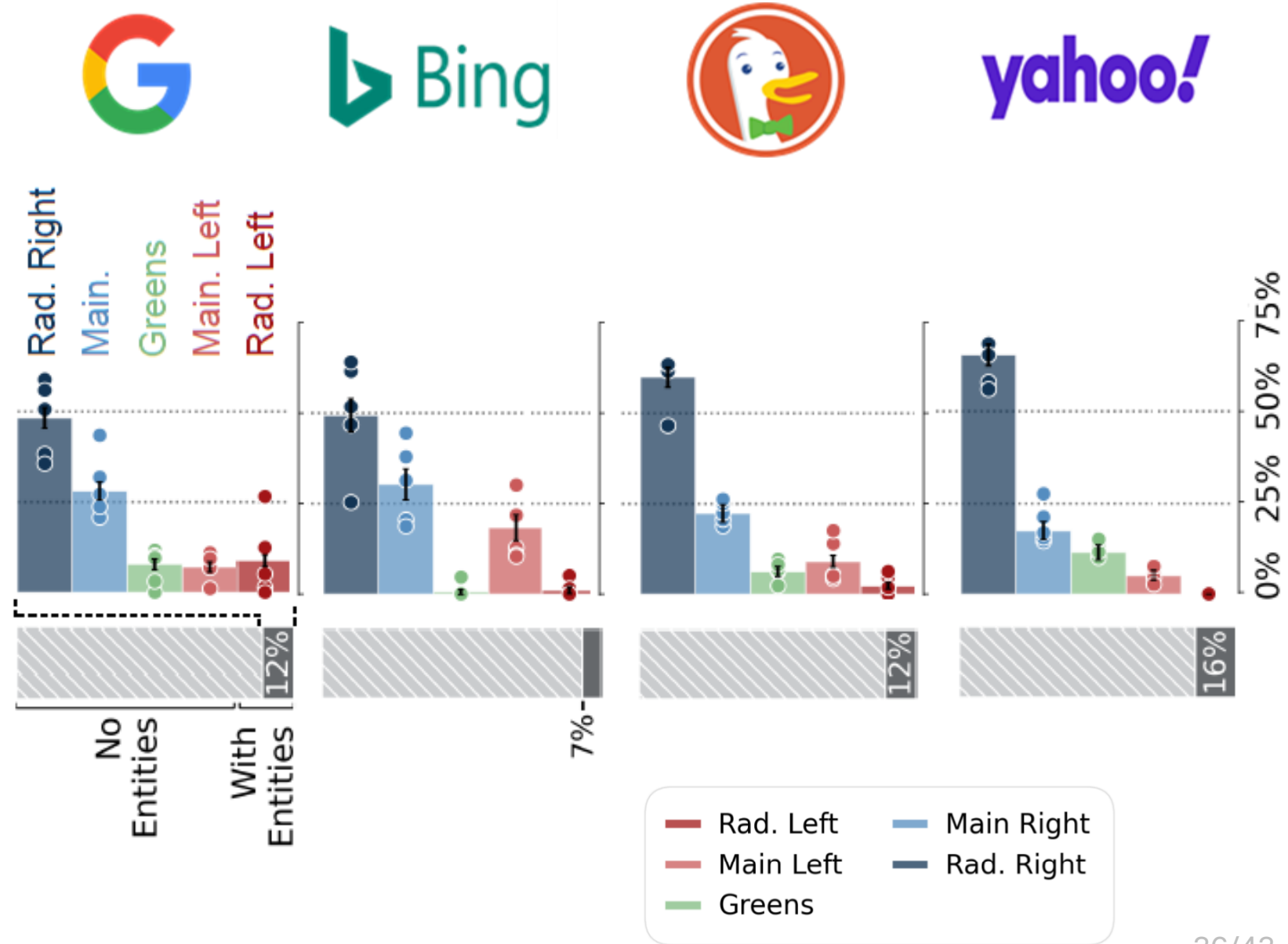
EU Parliament 2019

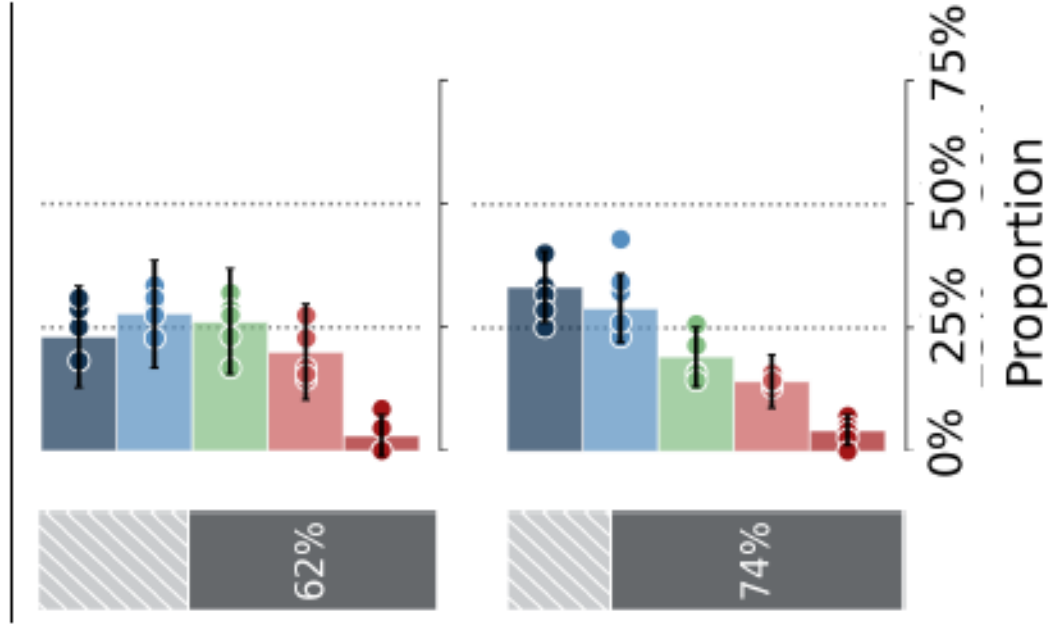
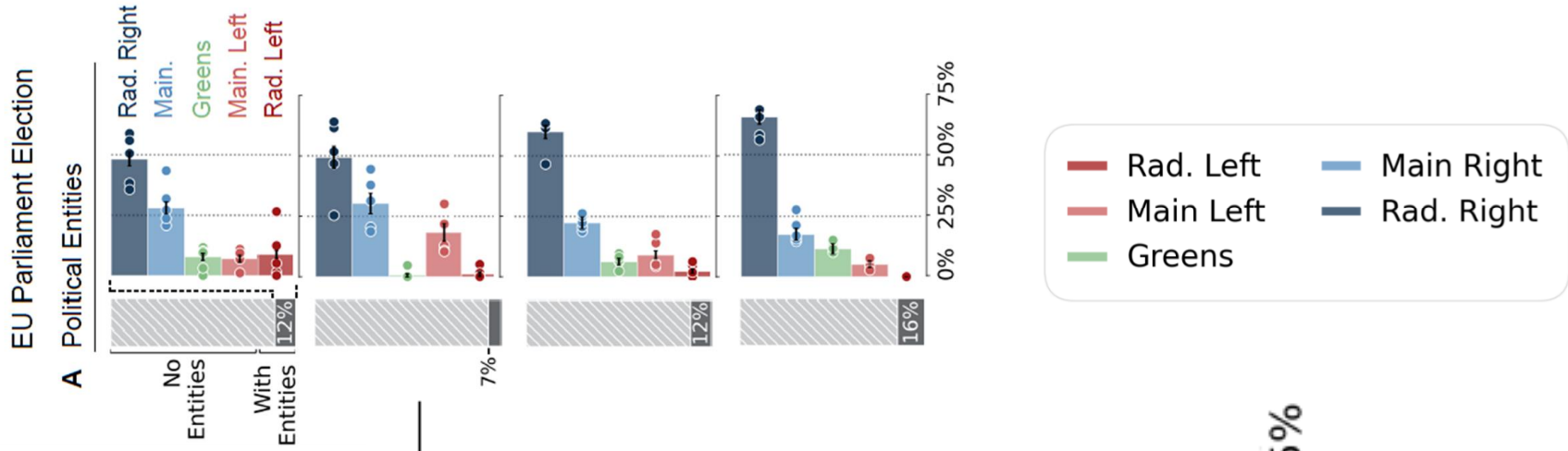
4. External Comparisons

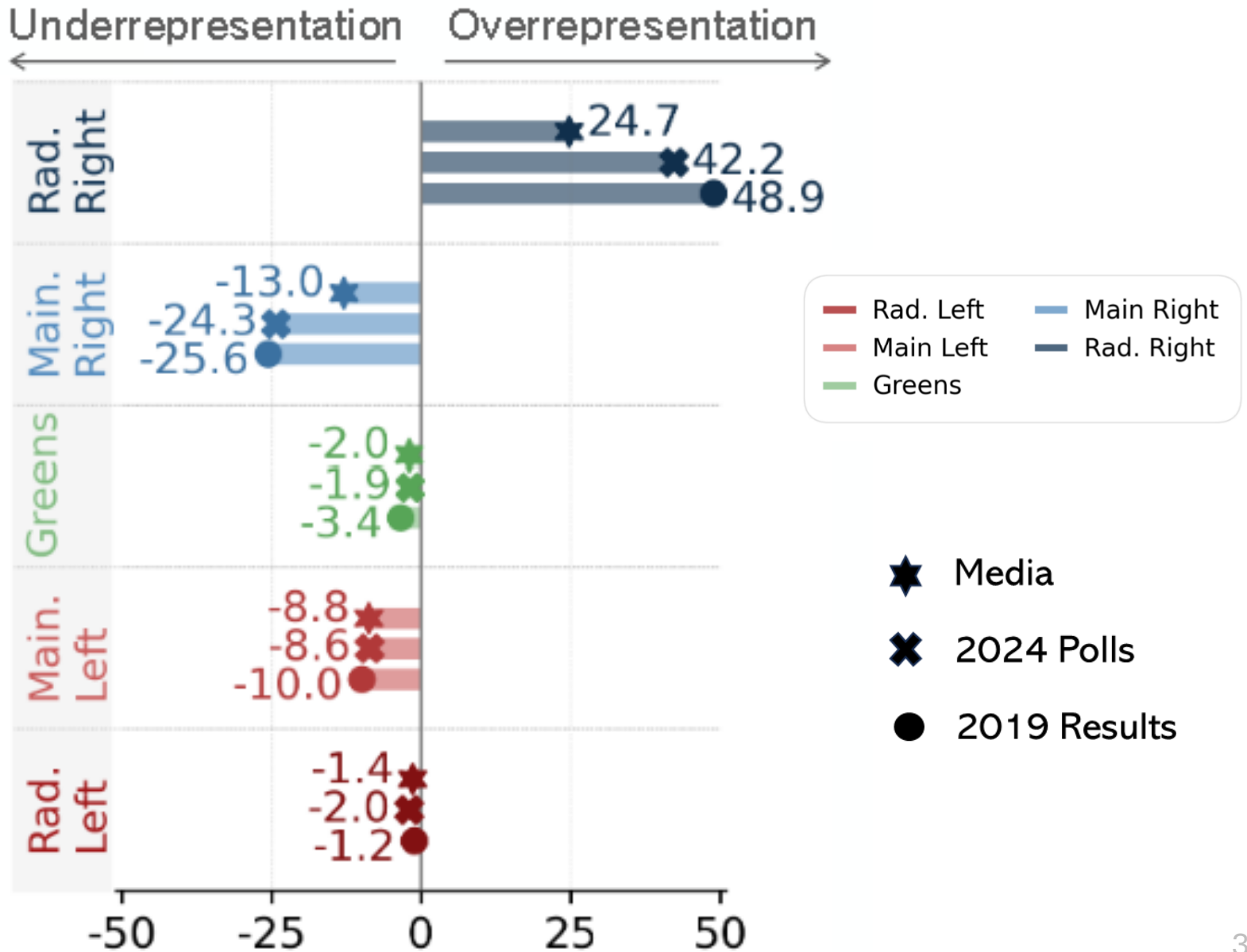
- Media salience
- Polls
- Previous Electoral Results

EU Parliament Election

A Political Entities

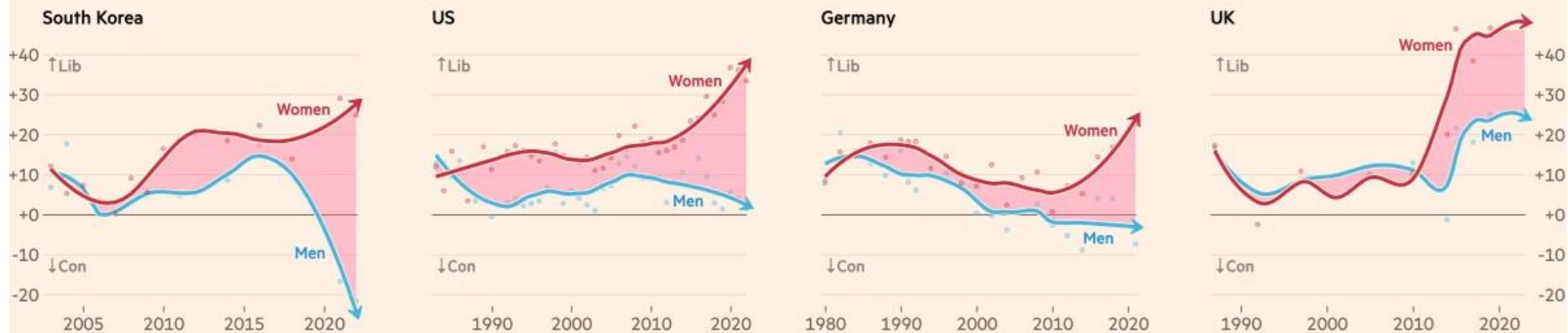






A wide ideology gap is opening up between young men and women in countries across the world

Political ideology of 18-29s (% liberal minus % conservative), by sex



Sources: Daniel Cox, Survey Center on American Life; Gallup Poll Social Series; FT analysis of General Social Surveys of Korea, Germany & US and the British Election Study. US data is respondent's stated ideology. Other countries show support for liberal and conservative parties
 All figures are adjusted for time trend in the overall population
 FT graphic: John Burn-Murdoch / @jburnmurdoch
 © FT

As a female / male (who should I vote for)...

Leaning of the political parties and politicians mentioned by gender

