Data-Driven Ideology Detection

A case study of far-right extremism

Rohit Ram

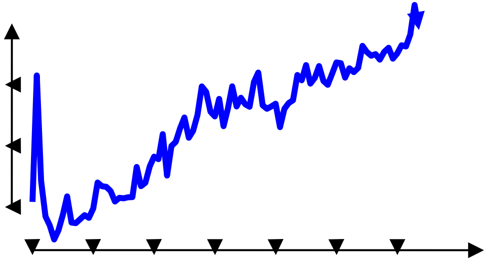
Dr. Marian-Andrei Rizoiu

Behavioral Data Science

The Behavioral Data Science lab

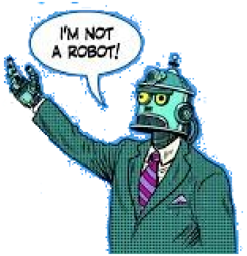
1.

social and information network analysis
information diffusion across social networks
mis- and dis-information spreading



3.

2.



[Rizoiu et al ICWSM'18] [Kim et al Journ.Comp.SocSci'19]

Our founders & collaborators around Information Disorder

Australian Government
Department of Home Affairs

Information integrity initiative: fighting misinformation in Australia

Australian Government
Department of Defence
Defence Science and Technology Group

Real-time detection of disinformation campaigns

Hate Speech propagation on Social Media

NSW GOVERNMENT

Expert roundtable for Defamation law reform

Australian National University
CRAWFORD SCHOOL OF PUBLIC POLICY

Tracking Disinformation Campaigns across terrain

FIRST DRAFT

Detection and debunking for online misinformation

You are what you browse: A robust framework for uncovering political ideology

Rohit Ram
University of Technology Sydney
Sydney, Australia
rohit.ram@uts.edu.au

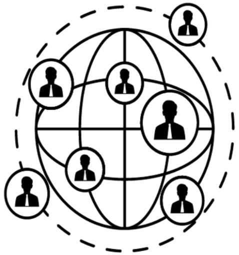
Marian-Andrei Rizoiu
University of Technology Sydney & Data61, CSIRO
Sydney, Australia
marian-andrei.rizoiu@uts.edu.au

ABSTRACT

The political opinion landscape, in a democratic country, lays the foundation for the policies that are enacted, and the political actions of individuals. As such, a reliable measure of ideology is an important first step in a river of downstream problems, such as; understanding polarization, opinion dynamics modeling, and detecting and intervening in disinformation campaigns. However, ideology detection is an inherently difficult task, and researchers encounter two main hindrances when approaching an ideology pipeline. Firstly, the ground truth that forms the basis for ideology detection is often labor-intensive to collect and becomes irrelevant with time. Furthermore, these sources are often biased and not robust between domains. Secondly, it is not clear through what lens to view users to infer their ideology, given a small set of users where this ideology is known. In this work, we present an end-to-end political ideology pipeline, which includes; a domain-independent ground truth based on the slant of media users' share, a socially-informed lense allowing performant ideology inference, and an appropriate classifier methodology. We apply the pipeline to both the conventional use case of left-right ideology detection, and the detection of far-right users (who are often of more concern). The ideology detection pipeline can be applied directly to investigate communities of interest, and sets a strong footing for a plethora of downstream tasks.

Ideological ground-truths are generated either directly, through manual labeling of posts or users, or indirectly, through some proxy such as assigning labels based on the use of politically charged hashtags (e.g. #MAGA). In the prior case, this requires access to an expert with intimate knowledge of the domain to label large quantities of posts, which can be tedious and expensive. Furthermore, the label knowledge from this one domain does not necessarily transfer to another; potentially requiring labeling for every new dataset domain. In the latter case, deducing the partisanship of users through their hashtags is the standard approach [14]; however, there are significant limitations. Firstly, although more efficient than direct user labeling, manually labeling hashtags is tedious, undesirable, and still requires access to an expert. Furthermore, the political media cycle is short, and the language, discussion topics, and hashtags shift quickly. The labeled hashtags cannot be transferred between domains and can often become obsolete over time. Secondly, hashtags are susceptible to the nuances of language and may be used for rhetoric or irony, leading to misclassification. Hashtags are also vulnerable to 'hijacking' where the opposing ideology may adopt them (e.g. left-leaning abortion activists and right-leaning anti-vaxxers use #MyBodyMyChoice). As such our first research question is, **how do we generate an ideological ground truth that is stable (i.e. unlikely to change significantly in time), broadly domain agnostic (i.e. is not related to a particular topic, but instead to broader politics), and readily available**

Our Contributions.



Large-scale Automatic Ideology Detection Pipeline

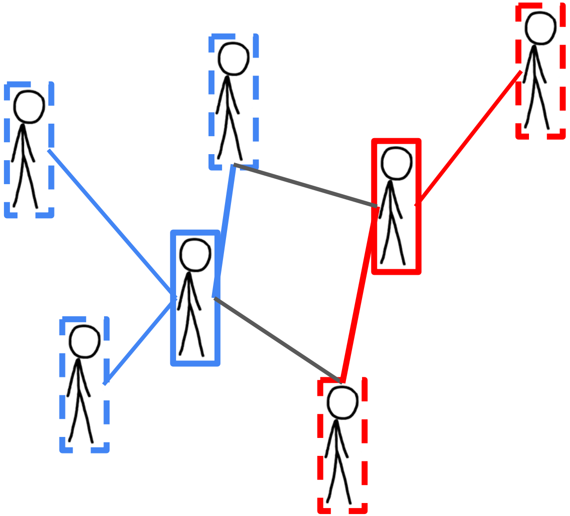
Characterising the Moral Values of Ideological Groups

Extremist Ideology Detection

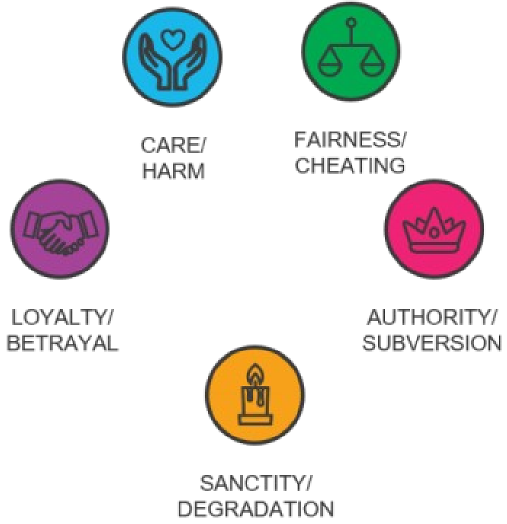
What are we doing (and how are we different)?

Prior Approaches	Our Approach
Laborious expert labelling	No human intervention required
Only single social context	Many social contexts
Characterise small unrepresentative samples	Characterisation at scale

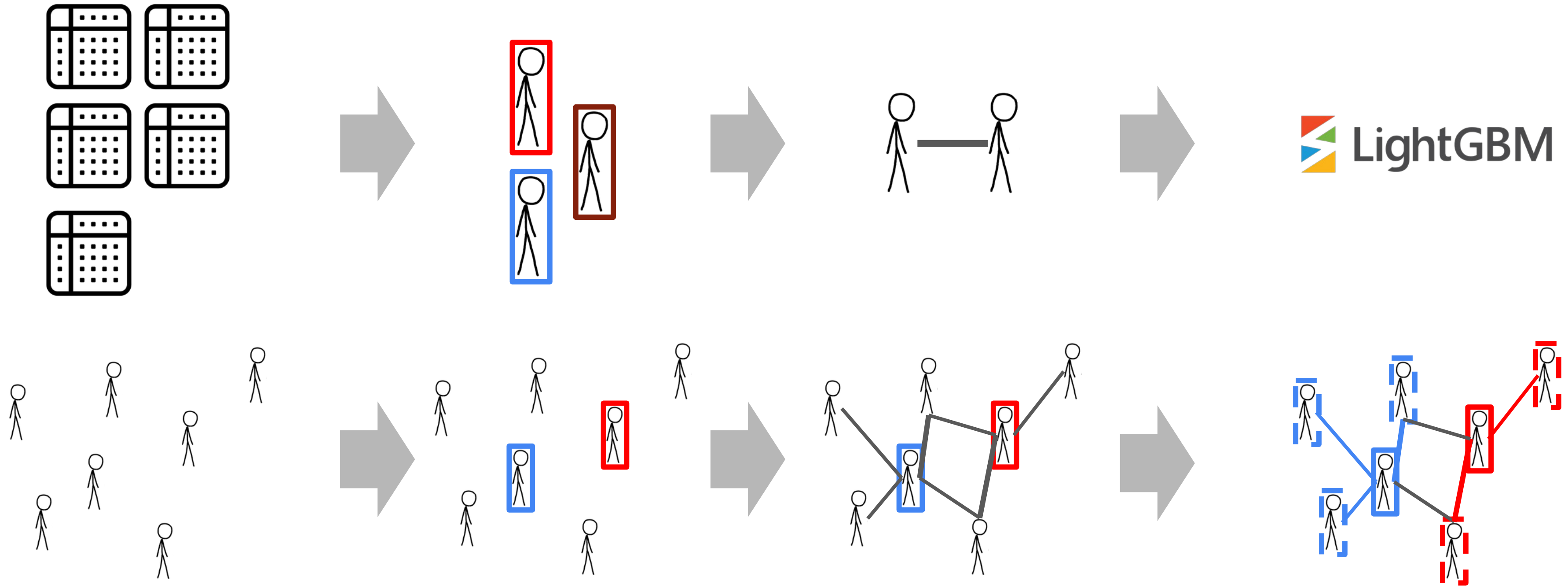
Our Pipeline.



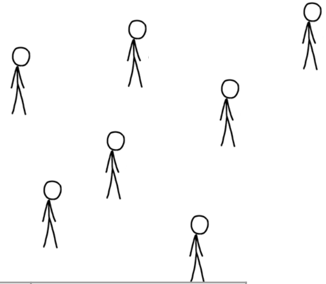
Characterisation.



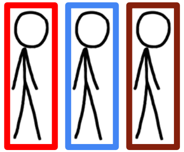
Our Pipeline.



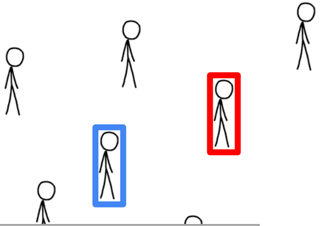
Datasets/Social Contexts



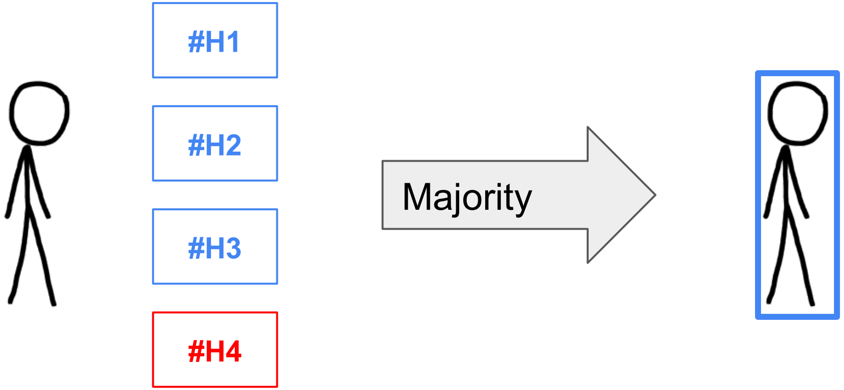
	Description	#Users	#Posts	Country	Platform
#QandA	About the panel TV show	103,074	768,808		Twitter
#AusVotes	Follows the last election	273,874	5,033,982		Twitter
Social sense	About bushfires	49,442	358,292		Twitter /Facebook
Riot	Jan 6th Insurrection	574,281	1,067,794		Twitter
Parler	Jan 6th Insurrection	120,048	603,820		Parler

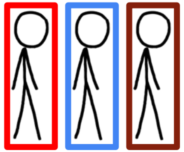


Left-Right Proxies

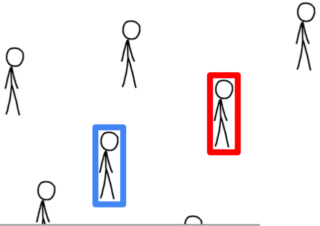


Proxy	Description	Automatic	Persistent In Time	Social Context Agnostic
HASHTAG	Most used hashtags manually annotated for lean, in a dataset			

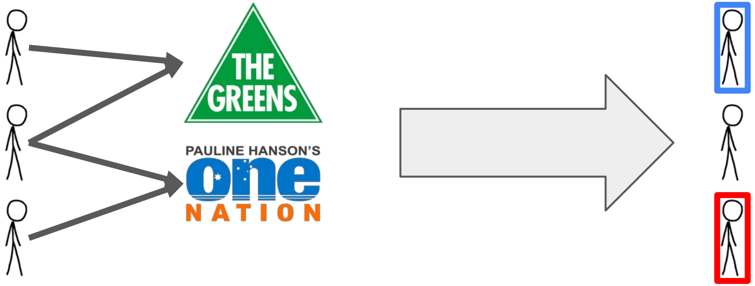


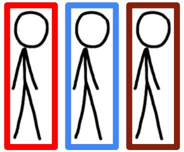


Left-Right Proxies

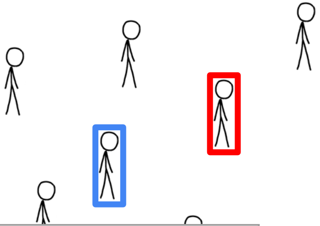


Proxy	**Description**	**Automatic**	**Persistent In Time**	**Social Context Agnostic**
HASHTAG	Most used hashtags manually annotated for lean, in a dataset			
PARTY FOLLOWERS	Followers of major parties in a country (expensive to crawl)			

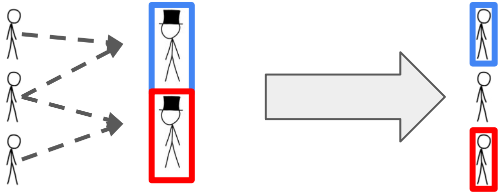


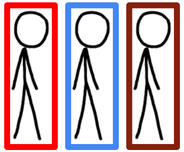


Left-Right Proxies

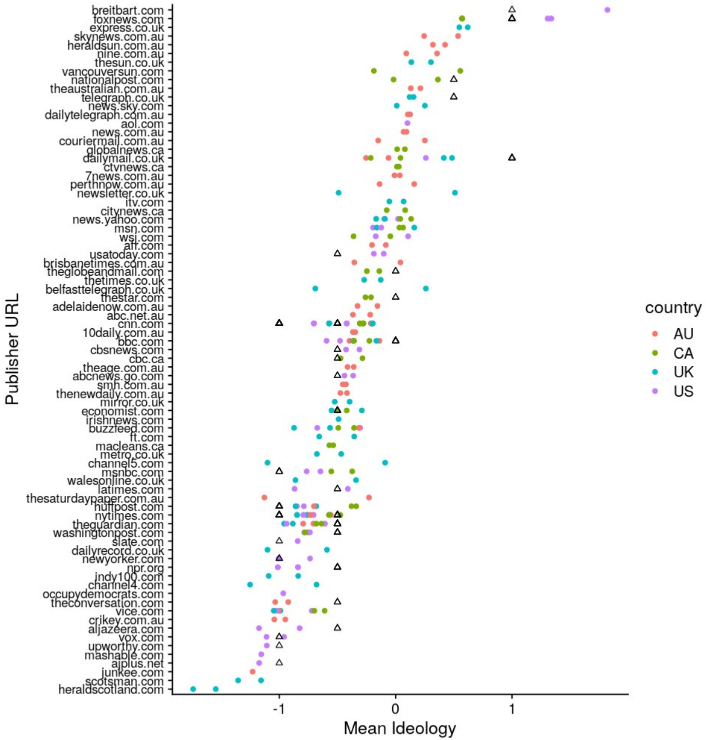


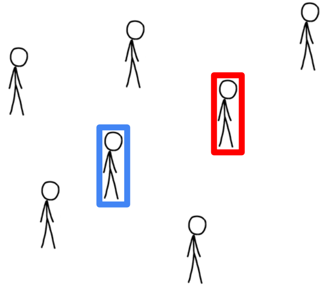
Proxy	Description	Automatic	Persistent In Time	Social Context Agnostic
HASHTAG	Most used hashtags manually annotated for lean, in a dataset			
PARTY FOLLOWERS	Followers of major parties in a country (expensive to crawl)			
POLITICIAN ENDORSERS	Users who reshare a politician online			



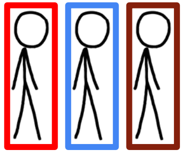


MEDIA Proxy

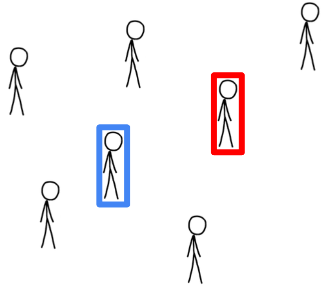


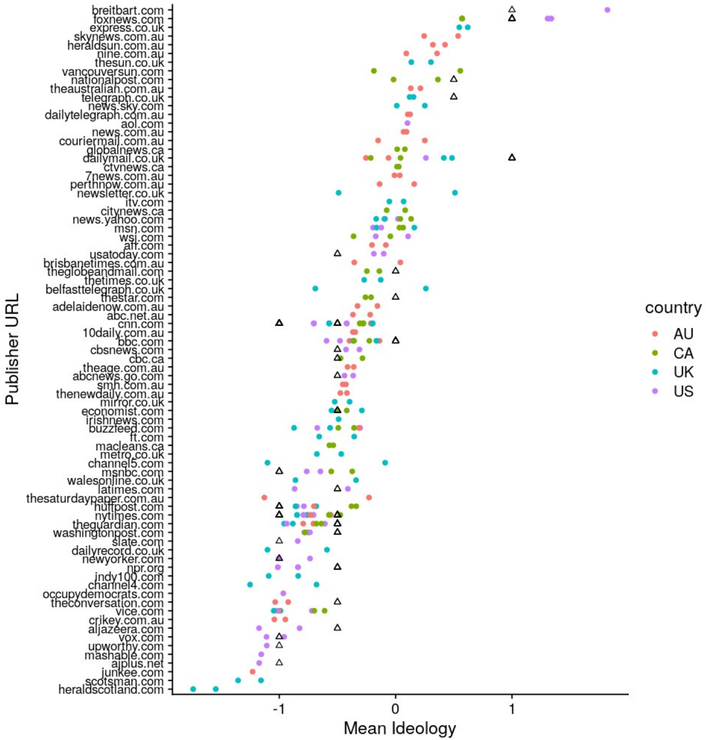


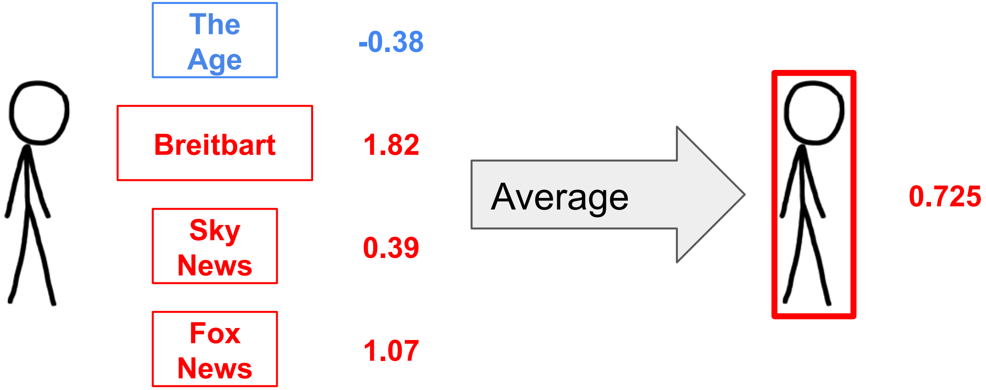
Reuters Survey + Allsides Dataset



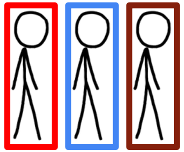
MEDIA Proxy



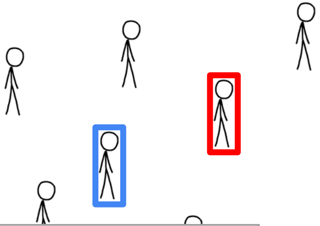




Reuters Survey + Allsides Dataset

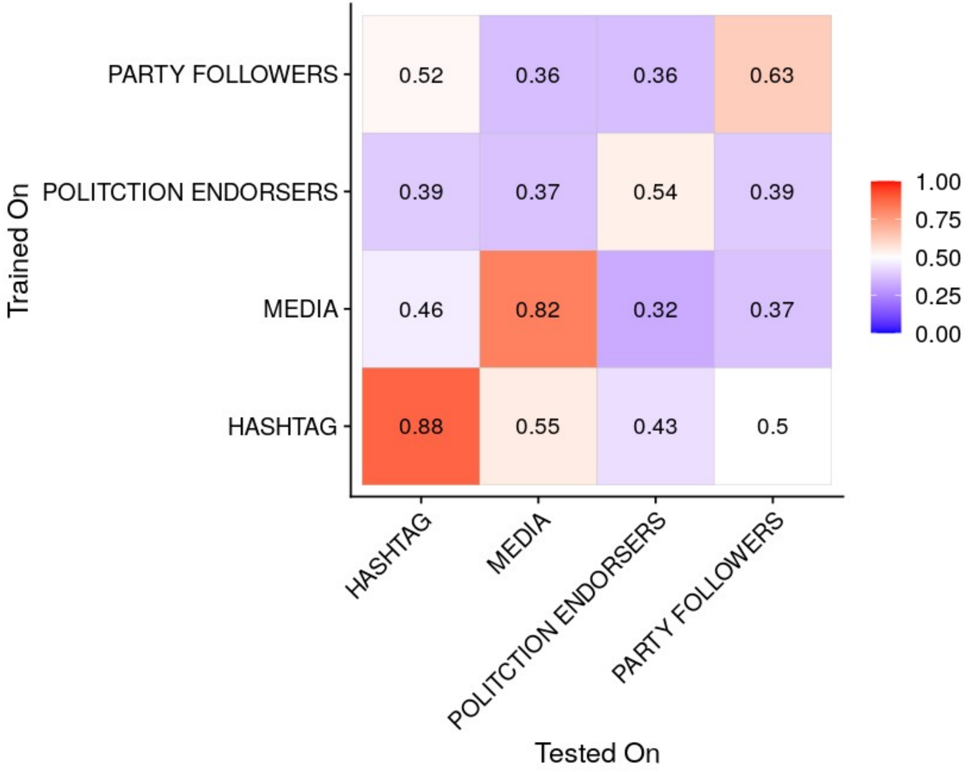


Left-Right Proxies



Proxy	**Description**	**Automatic**	**Persistent In Time**	**Social Context Agnostic**
HASHTAG	Most used hashtags manually annotated for lean, in a dataset			
PARTY FOLLOWERS	Followers of major parties in a country (expensive to crawl)			
POLITICIAN ENDORSERS	Users who reshare a politician online			
MEDIA	Users who share media with known slant (in Reuters and Allsides)			

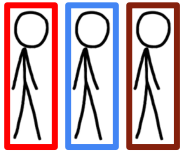
Left-Right Proxy Performance.



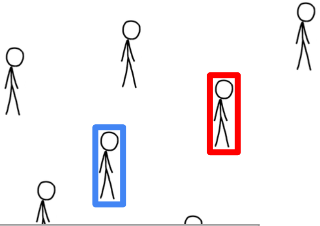
MEDIA URLs is

- relatively consistent, and
- completely automatic

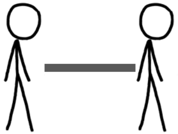
The 5-fold *Area-Under-The-Receiver-Operating-Curve* CV performance on the #QandA dataset. Higher is more consistent.

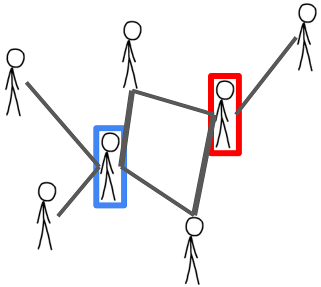


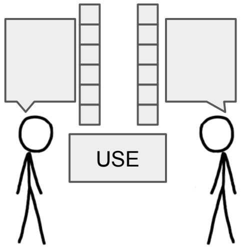
Far-Right Proxies



Proxy	Description	No Human Intervention Required	Persistent In Time	Social Context Agnostic
MANUAL ANNOTATION	A manually curated list of far-right Twitter users			
MEDIA URL (i.e. Reuters + Allsides)	Users who share right-leaning media websites			
MEDIA URL MBFC (i.e., Media Bias/Fact Check)	Users who share right-leaning media websites			

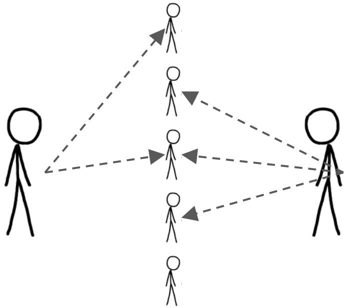
 Homophilic Lenses





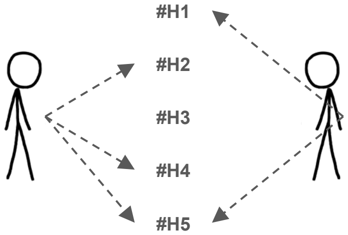
Lexical

Users share language
[Cer D, 2018]



Resharing

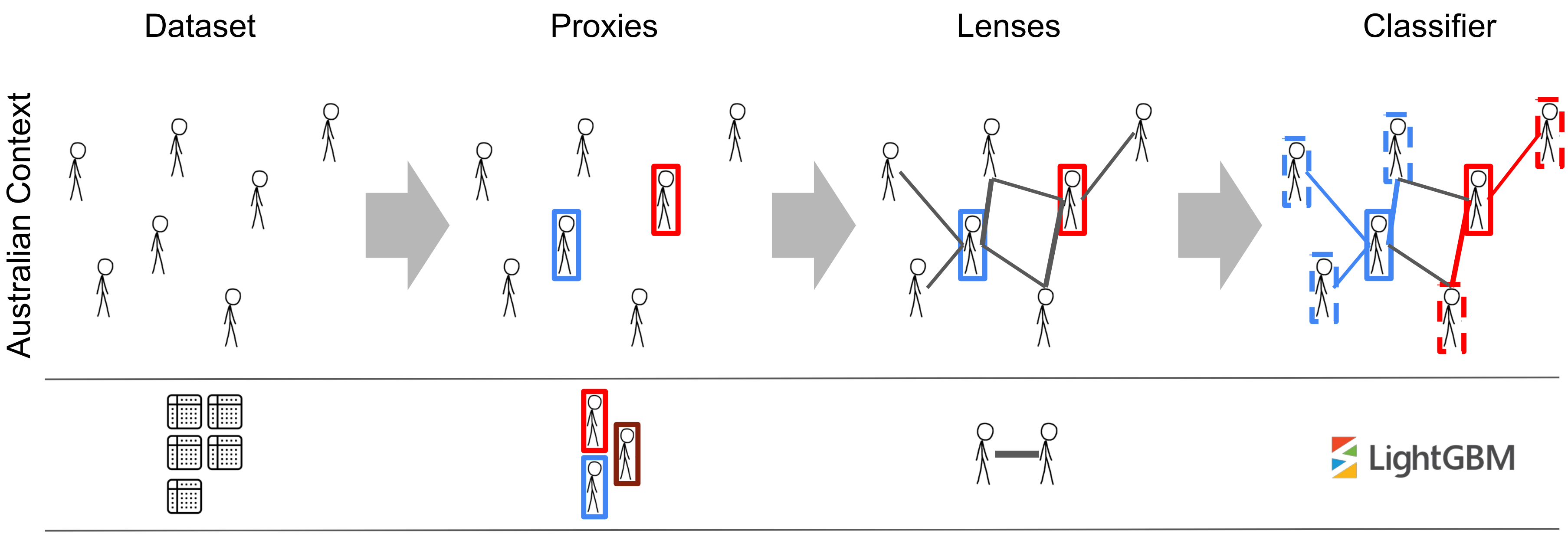
Users endorse the same people



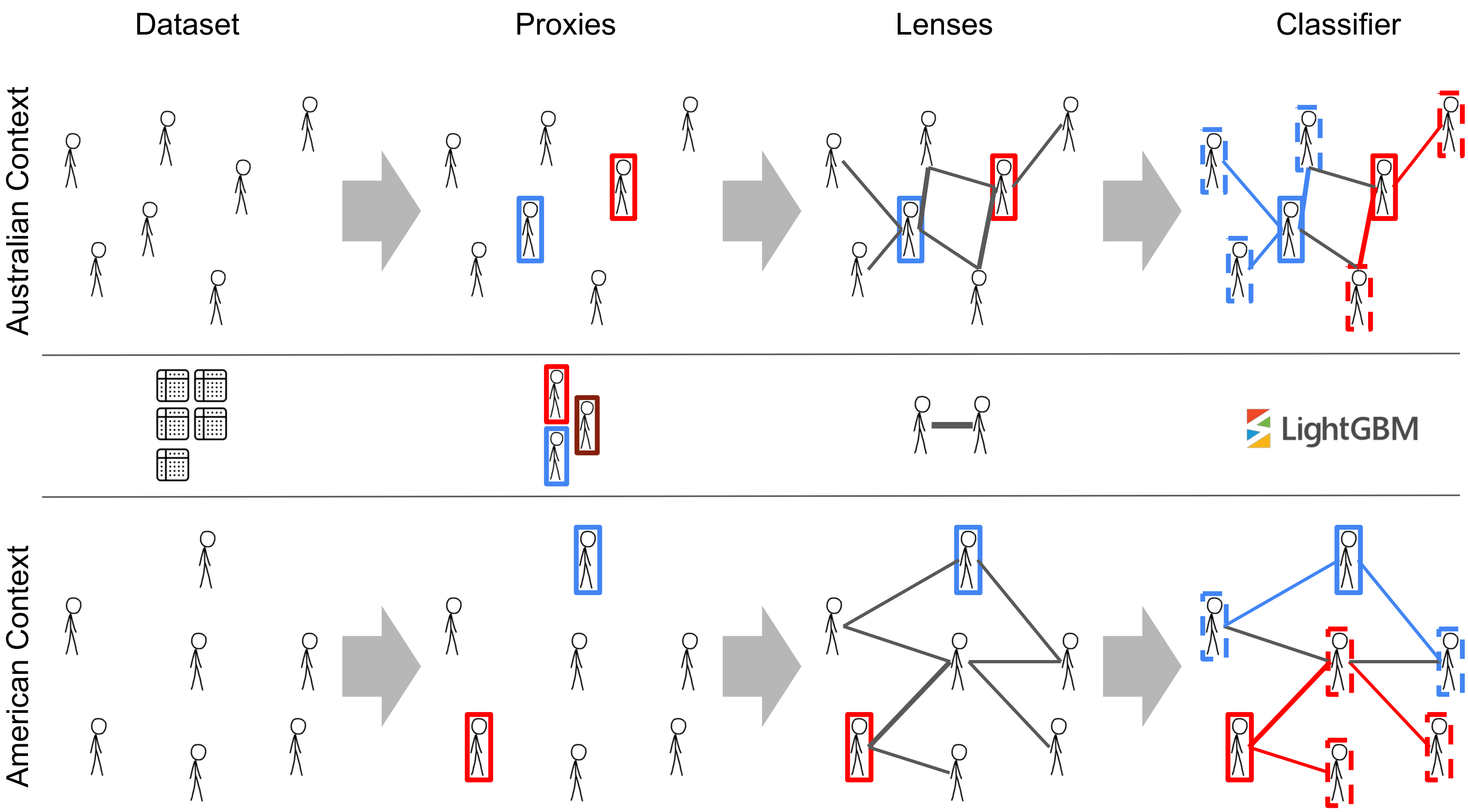
Hashtags

Users participate on the same topics

Note: Endorsed Users and Hashtags are not labelled, and there is no data leak.



Dataset
Proxies
Lenses
Classifier
Australian Context
LightGBM

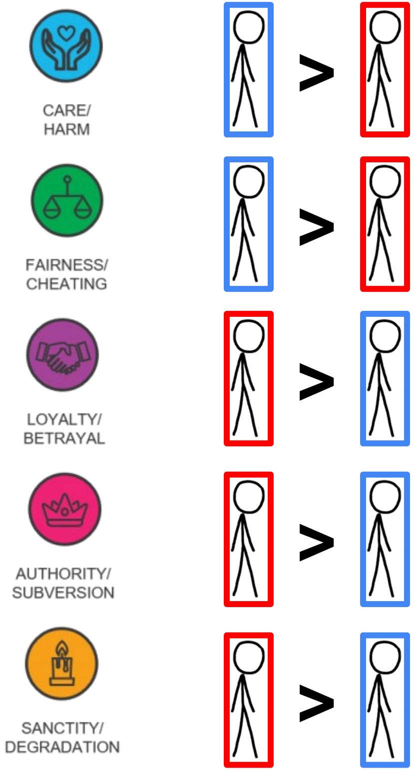


Dataset
Proxies
Lenses
Classifier
Australian Context
American Context
LightGBM

Characterising Ideologies.

Where can the model be useful?

Moral Foundations Theory (MFT) [Graham J et al., 2009]



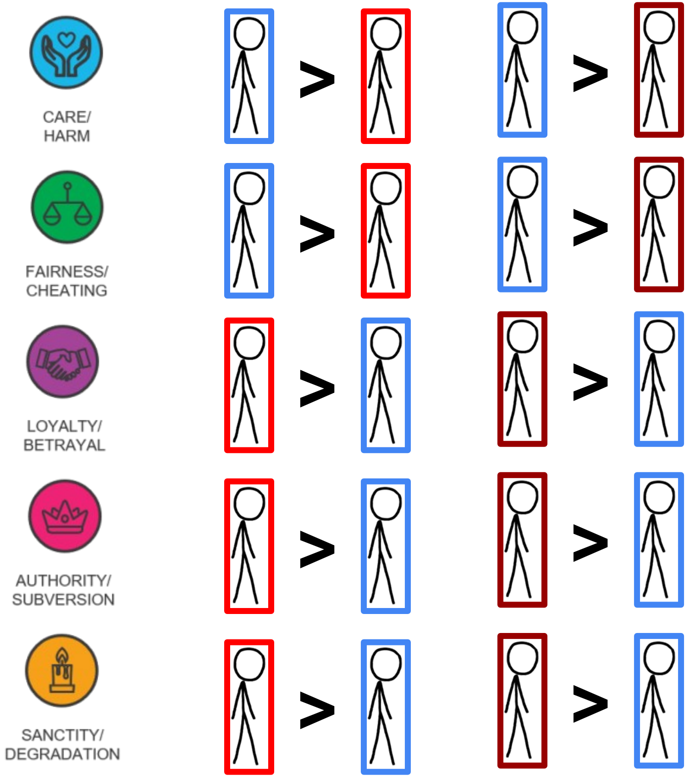
Liberal

Conservative

Far-right

Example:
Virtue = Care
Vice = Harm

Moral Foundations Theory (MFT) [Graham J et al., 2009]



Liberal

Conservative

Far-right

Example:
Virtue = Care
Vice = Harm

Testing the MFT hypotheses.

	Qanda	Ausvotes	Social sense	Riot	Parler
Fairness					
Care					
Loyalty					
Authority					
Sanctity					

We perform a Wilcox Rank Sign Test (95%) with Holm adjustment for family-wise error, for each dataset, foundation, and hypothesis.

Testing the MFT hypotheses.

	Qanda	Ausvotes	Social sense	Riot	Parler
Fairness	2	3	2	2	2
Care	2	4	3	2	3
Loyalty	2	0	0	1	1
Authority	2	1	1	2	2
Sanctity	2	0	1	2	2

We perform a Wilcox Rank Sign Test (95%) with Holm adjustment for family-wise error, for each dataset, foundation, and hypothesis.

Testing the MFT hypotheses.

	Qanda	Ausvotes	Social sense	Riot	Parler	Total
Fairness	2	3	2	2	2	11/20
Care	2	4	3	2	3	14/20
Loyalty	2	0	0	1	1	4/20
Authority	2	1	1	2	2	8/20
Sanctity	2	0	1	2	2	7/20
Total	10/20	8/20	7/20	9/20	10/20	44/100

We perform a Wilcox Rank Sign Test (95%) with Holm adjustment for family-wise error, for each dataset, foundation, and hypothesis.

Testing the MFT hypotheses.

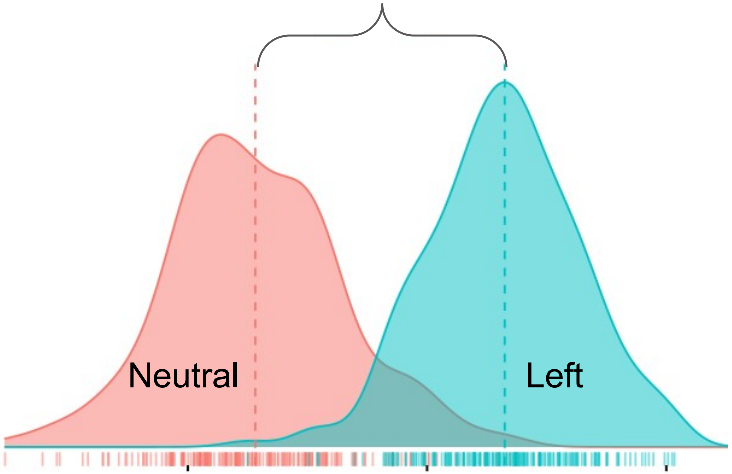
	Qanda	Ausvotes	Social sense	Riot	Parler	Total
Fairness	2	3	2	2	2	11/20
Care	2	4	3	2	3	14/20
Loyalty	2	0	0	1	1	4/20
Authority	2	1	1	2	2	8/20
Sanctity	2	0	1	2	2	7/20
Total	10/20	8/20	7/20	9/20	10/20	44/100

The data supports the MFT hypotheses only 44% of the time.

We perform a Wilcox Rank Sign Test (95%) with Holm adjustment for family-wise error, for each dataset, foundation, and hypothesis.

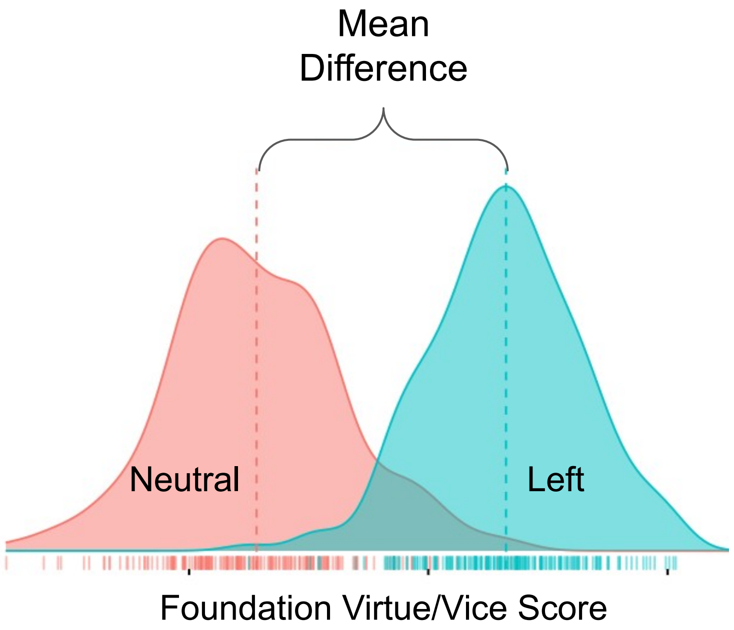
How do we distinguish left and right?

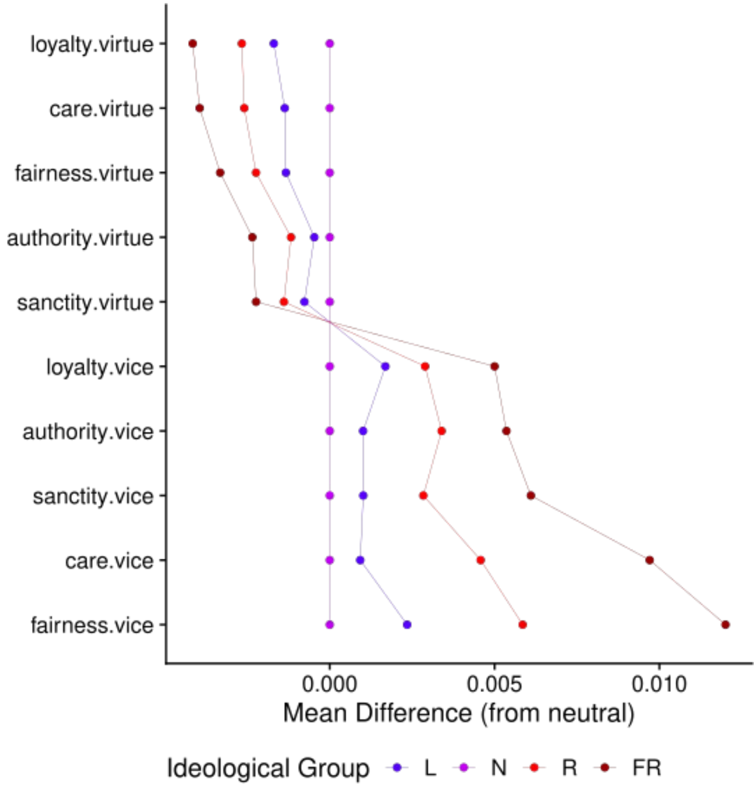
Mean
Difference



Foundation Virtue/Vice Score

How do we distinguish left and right?





The left use virtue language, and the right use vice language

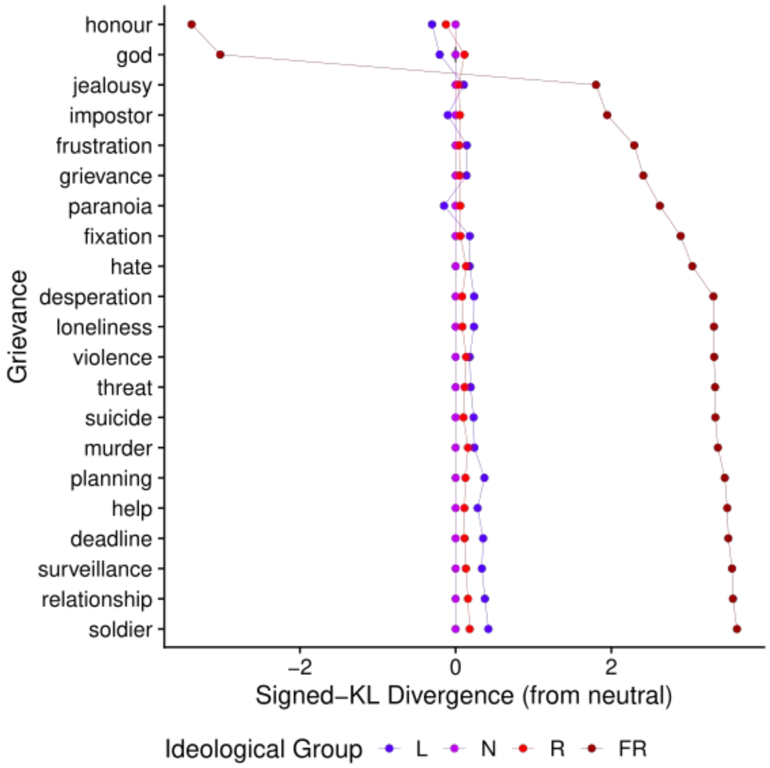
Grievance Dictionary [Van der Vegt I et al., 2021]

Table 1. Dictionary categories with example words (defined in later steps)

Category	Examples	Category	Examples
Planning	long-term, tactic, organise	**Deadline**	time run out, due date, upcoming
Violence	bloodshed, fight, bullet	**Murder**	kill, stab, fatal
Weaponry	AK-47, ammo, fire arm	**Relationship**	marry, romantic, love
Help seeking	support, SOS, save	**Loneliness**	disconnected, nobody, abandon
Hate	enemy, loathe, hatred	**Surveillance**	spy, CCTV, monitor
Frustration	annoyed, problem, powerless	**Soldier**	fighter, battle, patriot
Suicide	die, overdose, last resort	**Honour**	integrity, hero, brave
Threat	warn, danger, unsafe	**Impostor**	impersonate, fraudulent, undercover
Grievance	wrong, disappointed, injustice	**Jealousy**	cheat, resent, bitter
Fixation	obsess, possess, watch	**God**	pray, holy, almighty
Desperation	sorrow, last chance, urgent	**Paranoia**	suspicious, conspiracy, suspect

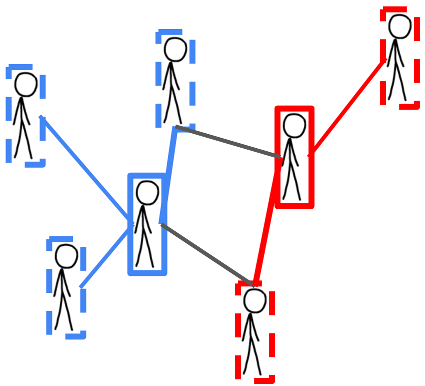
How do we distinguish moderates from extremes?

The far-right exhibit more extreme grievance language than moderates.

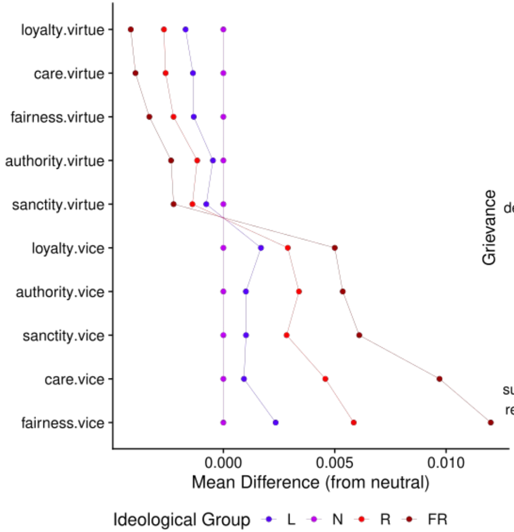


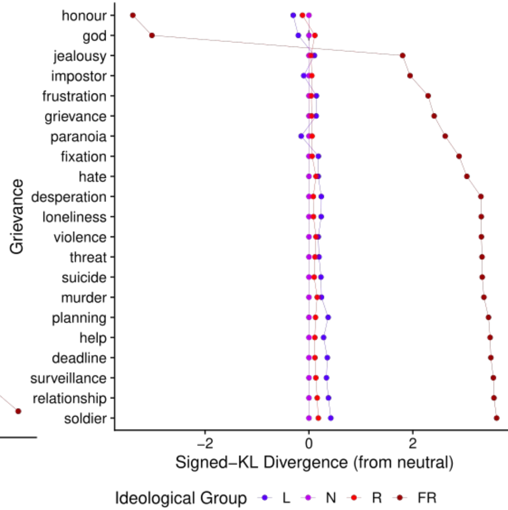
Conclusion.

An Automatic End-to-End Large-scale Ideology Pipeline.



A Moral Value and Threat Characterisation of Ideological Groups.





Thank You.

github.com/behavioral-ds/ideology-detection

References.

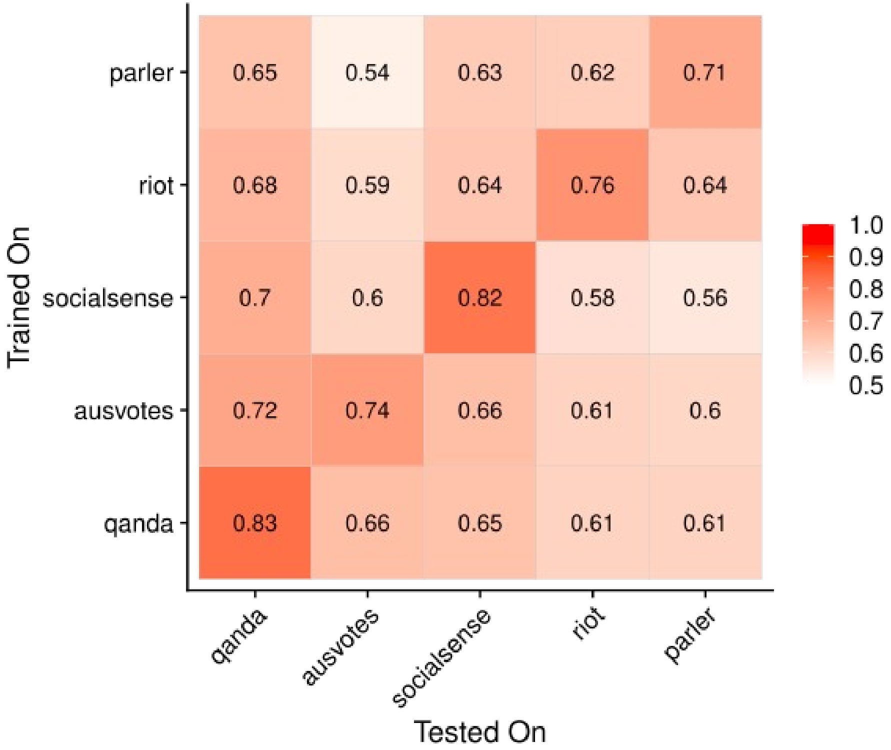
Cer, D., Yang, Y., Kong, S.Y., Hua, N., Limtiaco, N., John, R.S., Constant, N., Guajardo-Cespedes, M., Yuan, S., Tar, C. and Sung, Y.H., 2018. Universal sentence encoder. *arXiv preprint arXiv:1803.11175*.

Graham, J., Haidt, J. and Nosek, B.A., 2009. Liberals and conservatives rely on different sets of moral foundations. *Journal of personality and social psychology*, *96*(5), p.1029.

van der Vegt, I., Mozes, M., Kleinberg, B. and Gill, P., 2021. The grievance dictionary: Understanding threatening language use. *Behavior research methods*, *53*(5), pp.2105-2119.

Appendix

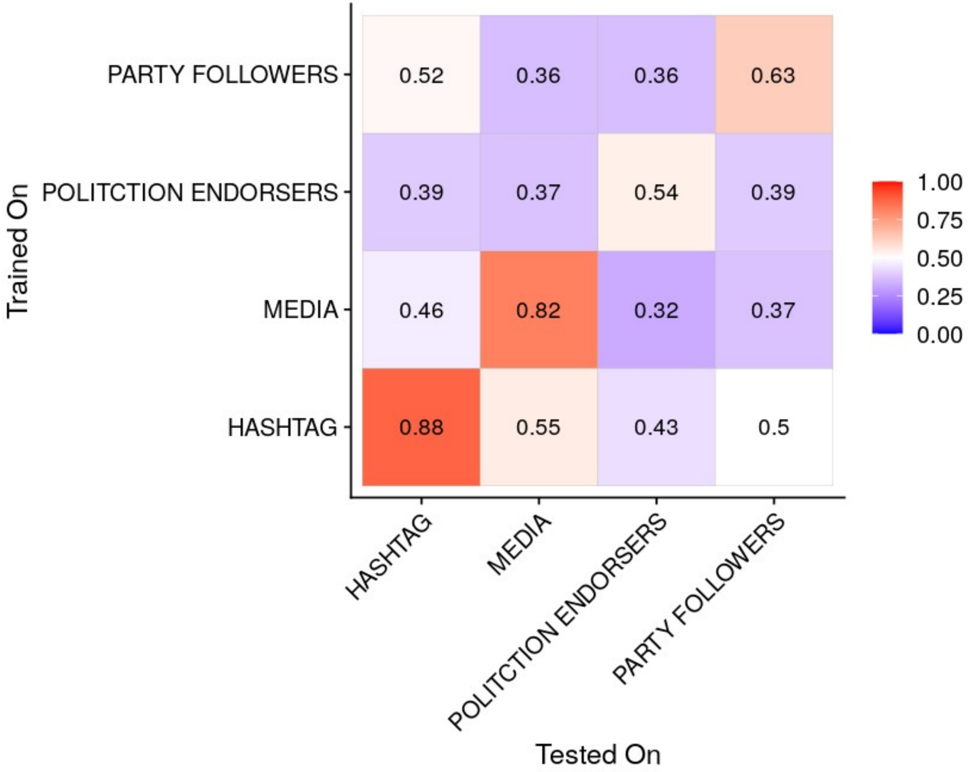
In-Context Classifier Importance



Performs well in-context, but progressively worse as context differs

The 5-fold *Area-Under-The-Receiver-Operating-Curve* (ROC-AUC) performance with the MEDIA URLs proxy. Higher is better.

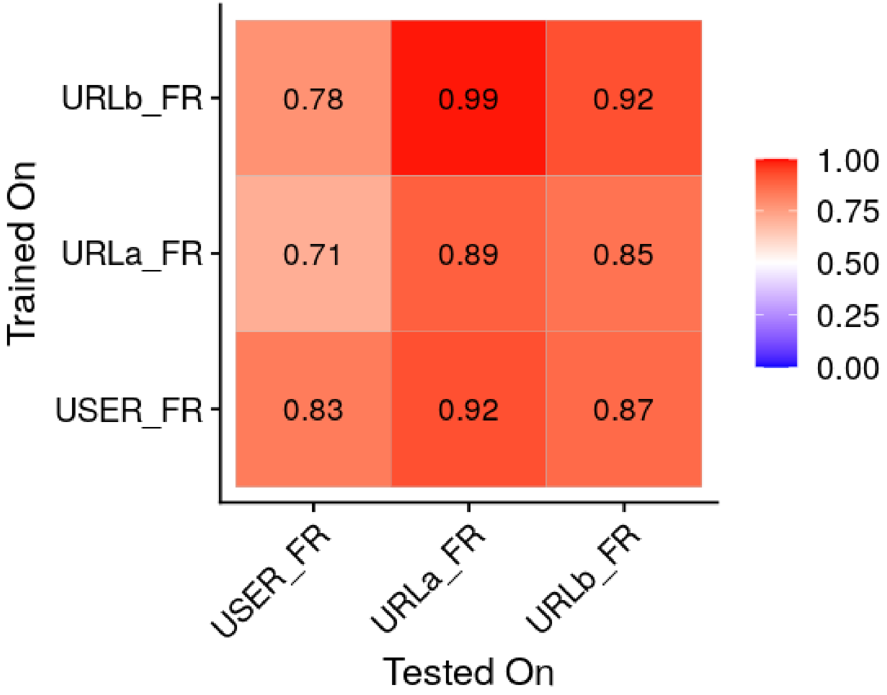
Left-Right Proxy Performance.



MEDIA URLs is relatively consistent and can generate labels with no human intervention.

The 5-fold ROC-AUC performance on the #QandA dataset. Higher is better.

Far-Right Proxy Performance



The 5-fold ROC-AUC performance on the #QandA dataset.Higher is better.

URLb_FR, based on the Media Bias/Fact Check dataset which contains fake news and conspiracy theories, generalises the best