

Bias and Stereotypes towards Politicians: A Multilingual Study with Reddit Data

Indira Sen

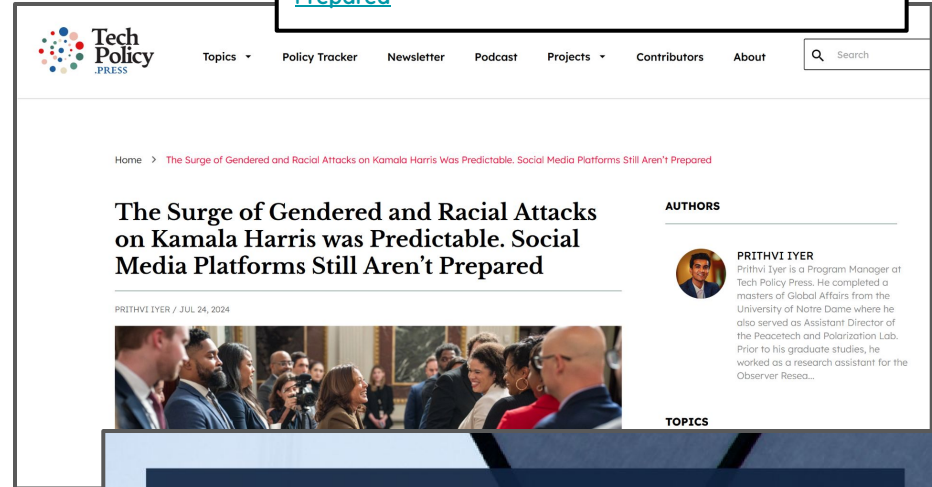
indira.sen@uni-mannheim.de

Why study this and why Reddit?

- A politician's background, such as their gender and ethnicity, can affect how people perceive them
- Many survey-based studies on this
- But, social media platforms are important fora for discussing politics
 - And include a great deal of abuse directed towards certain politicians
- Reddit has many affinity based communities, e.g., r/the_Donald*, r/politics, r/JoeBiden

*banned in 2020 for violating Reddit rules on harassment and targeting.

[The Surge of Gendered and Racial Attacks on Kamala Harris was Predictable. Social Media Platforms Still Aren't Prepared](#)



The screenshot shows the Tech Policy Press website. The header includes the logo, navigation links (Topics, Policy Tracker, Newsletter, Podcast, Projects, Contributors, About), and a search bar. The main content area features the article title "The Surge of Gendered and Racial Attacks on Kamala Harris was Predictable. Social Media Platforms Still Aren't Prepared" by Prithvi Iyer, dated July 24, 2024. An author bio for Prithvi Iyer is visible on the right, along with a "TOPICS" section. A photograph of a group of people is partially visible at the bottom of the article.

Twitter's Response to Abuse and Bigotry
Directed at Vice President Kamala Harris

Report

05-26-2022

[Twitter's Response to Abuse and Bigotry Directed at Vice President Kamala Harris](#)

Datasets

- Data from 605 subreddits, spanning 12 years (2008-2019)
- Subset of [Pushshift Reddit Dataset](#)
- Data: <https://doi.org/10.5281/zenodo.5851729>
- Data wrangling scripts: <https://github.com/valentinhofmann/politosp here>

Proceedings of the Sixteenth International AAAI Conference on Web and Social Media (ICWSM 2022)

The Reddit Politosphere: A Large-Scale Text and Network Resource of Online Political Discourse

Valentin Hofmann,^{1,3} Hinrich Schütze,³ Janet B. Pierrehumbert^{2,1}

¹Faculty of Linguistics, University of Oxford

²Department of Engineering Science, University of Oxford

³Center for Information and Language Processing, LMU Munich
valentin.hofmann@ling-phil.ox.ac.uk

Abstract

We introduce the Reddit Politosphere, a large-scale resource of online political discourse covering more than 600 political discussion groups over a period of 12 years. It is to the best of our knowledge the largest and ideologically most comprehensive dataset of its type now available. One key feature of the Reddit Politosphere is that it consists of both text and network data, allowing for methodologically-diverse analyses. We describe in detail how we create the Reddit Politosphere, present descriptive statistics, and sketch potential directions for future research based on the resource.

Introduction

The computational analysis of political discourse has attracted growing interest over the last few years (Garimella et al. 2018; Johnson and Goldwasser

to diverse methodological approaches. To facilitate research, we annotate the data in several ways and release scripts for easy data access. We publish the Reddit Politosphere on Zenodo at <https://doi.org/10.5281/zenodo.5851729> and the accompanying collection of scripts in a GitHub repository at <https://github.com/valentinhofmann/politosphere>.

The remainder of this paper is structured as follows. We first provide an overview of related work, focusing particularly on research on political discourse, online political discourse, and the complementary use of text and network data in the computational social sciences. We then present in detail our method of creating the Reddit Politosphere, followed by exploratory analyses that showcase potential avenues for future research. Finally, we provide a short conclusion and discuss ethical aspects of the resource.

Hofmann et al., [The Reddit Politosphere: A Large-Scale Text and Network Resource of Online Political Discourse](#)

Methods

- Task 1: Link reddit posts and comments with mentioned politicians using [Wikidata](#) => allows us to also get background info about the politicians
- Task 2: Disaggregate the linked data based on demographics of the mentioned politicians
 - Gender
 - Political party
 - Age
 - ...

Donald Trump (Q22686) Item Identifier


Label

Description
president of the United States from 2017 to 2021


Donald John Trump | Donald J. Trump | Trump | The Donald | POTUS 45 | Donald J Trump | President Donald Trump | President Trump | President Donald J. Trump | President Donald John Trump | DJT | Mr Trump | President Donald J Trump | John Barron | John Miller | David Dennison | Individual 1 | @realDonaldTrump Aliases

[In more languages](#)

Statements

instance of	human <small>value</small>
	1 reference
logo image	 <small>value</small>
	Trump Text Logo.svg 512 x 67; 1 KB
	0 references

Properties

image	 <small>value</small>
	Donald Trump official portrait.jpg

Task 3: Analyze differences in how politicians of varying backgrounds are discussed in different subreddits and languages [open-ended!]

- Topics
- Sentiment / Emotion
- Toxicity and hate speech
- Gendered language and sexism
- Network Analysis: which politicians are important?
- ...