Determinants of Canadian Politicians' Popularity on Wikipedia.*

Ethan Sansom

6 April 2022

Abstract

Voters internationally are increasingly using the internet as a primary tool to access information on politicians and political parties. Consequently, politcal organizers and researchers alike are using online platforms, such as Twitter, Facebook, and more recenlty Wikipedia, to analyze the sentiments of voters. This paper investigates the correlation between Canadian politicians' characteristics - including party affiliation, political position, age, and gender - and politicians' Wikipedia page views in the months prior to the 2021 Federal Election. Wikipedia page views are found to be greater for politicians in more popular parties, and for Party Leaders, Ministers, and Prime Ministers in particular.

Keywords: Canada elections, election forecasting, Wikipedia, linear regression

Contents

1 Introduction

Online search engines are increasingly the avenue by which voters internationally access information on political candidates in their countries. A reported 68% of internet-using voters in Germany, France, Italy, Spain, Poland, the United Kingdom, and the United States, for instance, rated online searches as an important factor in their voting choices (Pradel 2021). Such internet usage provides political organizers, pundits, and researchers a wealth of relatively easy to access on the voting preferences and attitudes of the electorate. An increasingly large number of researchers have used name mentions of political parties or individual politicians

*Code and data are available at: https://github.com/EthanSansom/wikipoliticians

on social media platforms such as Twitter and Facebook to measure the popularity of politicians and political parties (Smith and Gustafson 2017; Yasseri and Bright 2013). Such measures, however, are limited in their predictive abilities of voter behavior. Name mentions alone, for example, do not indicate a positive or negative attitude towards a politician or party (Yasseri and Bright 2013). Further, investigating social media platforms alone limits the sample of voters investigated to those who both participate in social media and use social media platforms as a venue for communicating political views (Yasseri and Bright 2013). Pressures to adhere to the political views of social circles, and potential negative outcomes at home or in the workplace of voicing divisive political views on a public platform, can additionally mask or misrepresent the intentions of voters on social media platform (Yasseri and Bright 2013). Authors Salem and Stephany (2021) argue that Wikipedia, the online encyclopedia, offers a better platform for accurately modelling voters' intentions. Specifically, the number of views of a given politician's or political party's Wikipedia page over a given time period can provide a strong measure of the number individuals searching for information about the politician or party (Yasseri and Bright 2013; Salem and Stephany 2021). Such information-seeking behavior has been shown to be reflective of the true public voting interest for certain candidates (Salem and Stephany 2021). Wikipedia page viewership, and other Wikipedia page metrics then, may offer a less ambiguous and more predictive measure of voting intent than that of other internet platforms.

To explore the potential of assessing voter behaviors and perceptions using Wikipedia in the context of Canadian elections, this paper investigates the Wikipedia presence and viewership of Canadian politicians from the five major Canadian federal political parties, the Liberal Party, Conservative Party, New Democratic Party (NDP), Green Party, and the Bloc Québécois. Using monthly data on Wikipedia page views from January 1st 2021 to January 1st 2022, this paper performs a multiple linear regression to investigate the correlation between characteristics of Canadian politicians and their Wikipedia page viewership. In particular, variables gender, age, political position, and party affiliation, are used to model the number of Wikipedia page views received by Canadian politicians. Additionally, the page views of Canadian federal party leaders, namely Justin Trudeau (Liberal), Erin O'Toole (Conservative), Jagmeet Singh (NDP), Annamie Paul (Green), and Yves-François Blanchet (Bloc Québécois), in the months preceding the 2021 Canadian federal election are compared with the 2021 election results. The Wikipedia page views of party leaders in the months preceding the election are shown to accurately predict the order of party vote share, from most to least votes. In the full sample of Canadian politicians on Wikipedia, it is found that political party affiliation is correlated with Wikipedia viewership. The Wikipedia pages of Green Party and Bloc Québécois politicians are shown to have substantially fewer page views on average than those of politicians in larger parties. Political position, specifically the positions of Prime Minister, Minister, and Party Leader, are shown to be very strongly correlated with page viewership as well.

In the following sections of this paper, a brief description of the data collection and cleaning process is provided, followed by an investigation of the data itself. The Model 3 section describes the multiple linear regression framework used in this analysis, including a discussion of the model's assumptions and subsequent limitations. In the Results 4 section, the implications of the model estimates and the results of the data investigation are described. Finally, in the Discussion 5 section, the results are analysed in the context of the 2021 federal election results and the political views of the Canadian electorate more broadly. The paper concludes with a discussion if its limitations, including the challenges to the external validity of the model and the ability of Wikipedia page viewership to serve as a valid proxy for voter interest. For this analysis, the R statistical software (R Core Team 2021) is used to import and clean Wikipedia page data, perform regression model estimation and data visualization, and compile the paper itself. Packages `tidyverse` (Wickham 2021b) and `dplyr` (Wickham, François, et al. 2021) are used to manipulate data, `ggplot2` (Wickham, Chang, et al. 2021), `gridExtra` (Auguie 2017), and `kableExtra` (Zhu 2021), are used for creating tables and plots, and `ggrepel` (Slowikowski 2021) is used to customize plot labels. To access web APIs, the package `httr` (Wickham 2020) is used, and for scraping HTML data the package `rvest` (Wickham 2021a) is used. The package `here` (Müller 2020) is used for to simplify reading and writing data within the file directory. Lastly, packages `bookdown` (Xie 2021a), `knitr` (Xie 2021b), and `tinytex` (Xie 2021c) are utilized to format and compile the final paper.

2 Data

3 Model

4 Results

5 Discussion

6 Appendix

References

Auguie, Baptiste. 2017. *gridExtra: Miscellaneous Functions for "Grid" Graphics*. https://CRAN.R-project.org/package=gridExtra.

Müller, Kirill. 2020. *Here: A Simpler Way to Find Your Files*. https://CRAN.R-project.org/package=here.

Pradel, Franziska. 2021. "Biased Representation of Politicians in Google and Wikipedia Search? The Joint Effect of Party Identity, Gender Identity and Elections." *Political Communication* 38 (4): 447–78. https://doi.org/10.1080/10584609.2020.1793846.

R Core Team. 2021. *R: A Language and Environment for Statistical Computing*. Vienna, Austria: R Foundation for Statistical Computing. https://www.R-project.org/.

Salem, Hamza, and Fabian Stephany. 2021. "Wikipedia: A Challenger's Best Friend? Utilizing Information-Seeking Behaviour Patterns to Predict US Congressional Elections." *Information, Communication & Society* 0 (0): 1–27. https://doi.org/10.1080/1369118X.2021.1942953.

Slowikowski, Kamil. 2021. *Ggrepel: Automatically Position Non-Overlapping Text Labels with 'Ggplot2'*. https://CRAN.R-project.org/package=ggrepel.

Smith, Benjamin K., and Abel Gustafson. 2017. "Using Wikipedia to Predict Election Outcomes: Online Behavior as a Predictor of Voting." *Public Opinion Quarterly* 81 (3): 714–35. https://doi.org/10.1093/poq/nfx007.

Wickham, Hadley. 2020. *Httr: Tools for Working with URLs and HTTP*. https://CRAN.R-project.org/package=httr.

———. 2021a. *Rvest: Easily Harvest (Scrape) Web Pages*. https://CRAN.R-project.org/package=rvest.

———. 2021b. *Tidyverse: Easily Install and Load the Tidyverse*. https://CRAN.R-project.org/package=tidyverse.

Wickham, Hadley, Winston Chang, Lionel Henry, Thomas Lin Pedersen, Kohske Takahashi, Claus Wilke, Kara Woo, Hiroaki Yutani, and Dewey Dunnington. 2021. *Ggplot2: Create Elegant Data Visualisations Using the Grammar of Graphics*. https://CRAN.R-project.org/package=ggplot2.

Wickham, Hadley, Romain François, Lionel Henry, and Kirill Müller. 2021. *Dplyr: A Grammar of Data Manipulation*. https://CRAN.R-project.org/package=dplyr.

Xie, Yihui. 2021a. *Bookdown: Authoring Books and Technical Documents with r Markdown*. https://CRAN.R-project.org/package=bookdown.

———. 2021b. *Knitr: A General-Purpose Package for Dynamic Report Generation in r*. https://yihui.org/knitr/.

———. 2021c. *Tinytex: Helper Functions to Install and Maintain TeX Live, and Compile LaTeX Documents*. https://github.com/yihui/tinytex.

Yasseri, Taha, and Jonathan Bright. 2013. "Can Electoral Popularity Be Predicted Using Socially Generated Big Data?" *It - Information Technology* 56 (December). https://doi.org/10.1515/itit-2014-1046.

Zhu, Hao. 2021. *kableExtra: Construct Complex Table with Kable and Pipe Syntax*. https://CRAN.R-project.org/package=kableExtra.