Big Data for Public Policy

Prof. Mihály Fazekas, PhD

10/04/2022

Anastasiia Andreeva

Assessing the Difference in Public Discussion of Environmental Policies: The Case of Green New Deal and Infrastructure Bill

Final Paper

Introduction

Rapidly worsening environmental conditions all around the world are the sad reality that we are living in. The changing climate dramatically threatens the stability of ecosystems, infrastructure, and overall human well-being. Intergovernmental Panel on Climate Change calls for decisive societal action in response to climate change, as more than a decade ago it has been proven that the scientific community is univocal on the issue of global warming: it is happening (IPCC 2007).

However, in public discourse climate change is still a very partisan topic. For instance, in the American context, the attitudes toward the environment are strongly shaped by partisan identification (Lachapelle, Borick, and Rabe 2014). The gap between scientific knowledge and public awareness on the issue is one of the biggest challenges for the field of climate change communication, as it diminishes the discussion on potential climate action policies. Quite an extensive body of scholarly work has been dedicated to this issue.

Some try to explain the differences among perceptions on the climate issue by analyzing the socio-demographic factors. It has been shown that in countries all around the world those who support climate policies tend to be more left-wing, younger, and living in urban areas (McCright and Dunlap 2011; Schumacher 2014). While this knowledge is substantial, it does not provide us an understanding of how can we take urgent action towards the mitigation of climate change in an effective way with the means of democratic institutions.

To explore this issue deeper, I aim to analyze the public discussion around two climate-related policies – Green New Deal (GND) and Infrastructure Bill (IB). GND became one of the primary agenda points for Democrats during the 2018 House elections. The primary focus of these policy proposals was action to address climate change, as well as the creation of jobs and reduction of inequality. GND is widely associated with the "progressive wing" of democrats, and Bernie Sanders in particular. IB, unlike GND, is not just a set of proposals, but a bill, that passed the

hearings in Congress and then was signed by President Biden. Addressing climate change is only a part of the IB policy, as it is widely aimed at improving the existing infrastructure. In general, IB – which has also remained known as the "Bipartisan Infrastructure Bill" – is more moderate in its agenda, compared to GND.

The differences between the two policies make an interesting case for comparison of public discussion. How does it differ when it comes to environmental policies from slightly different points of ideological specter? To answer this question, I intend to analyze Twitter data.

<div align="center">Data and Methods</div>

The main data for my research are tweets scraped via open Twitter API with R package rtweet. The tweets include keywords ("Green New Deal", "Infrastructure Bill") as well as the respective hashtags. I did not use abbreviations, as they might hold different meanings (for instance, I found out that GND also stands for "Gang and Narcotics Department"). Additionally, for IB-related tweets, I included those that have mentioned #BuildBackBetter. Build Back Better – a policy within Infrastructure Bill and the initial campaign name for Joe Biden's climate agenda. I did so to balance the sample, as tweets featuring GND were predominant. The breakdown of scraped tweets is provided below.

Table 1. The breakdown of the analyzed Twitter data

Keywords/Hashtag	N	Total
"Green New Deal"	20 000	33 093
"#GreenNewDeal"	13 093	
"Infrastructure Bill"	20 000	29 502
"#InfrastructureBill"	147	
"#BuildBackBetter"	9355	

My goal is to explore the content of the scraped tweets and possibly draw some generalizations from the existing discussions. To do so, I will employ sentiment analysis, I will construct word clouds, as well as apply topic modeling. Additionally, I will apply the supervised learning model, specifically random forest, to identify the key topics that differentiate the discussions. The results of the analysis are described in the following section.

Results

To get the initial sense of the data I have run a simple sentiment analysis, which summarises the tone of the body of tweets in terms of positivity and negativity. The results of the sentiment analysis are presented in Table 2 below.

Table 2. Sentiment Analysis Results

	Positive	**Negative**	**Neutral**
GND (33093 tweets)	39%	15%	46%
IB (29502 tweets)	30%	16%	54%

As we can see, the differences are very slight. The only general trend is that IB-related tweets seem to be slightly more neutral in their tone compared to the GND tweets (which are slightly more positive.

To further explore the data, I construct the word clouds. To do so, I converted the data frames of tweets into corpora and cleaned them (removed punctuation, stopwords, exotic encoding characters, as well as reduced sparsity – dropped the terms that are not mentioned in 97% of the tweets) using tm package. Then I constructed the word clouds with a quanteda package. The results can be seen in the figures below.
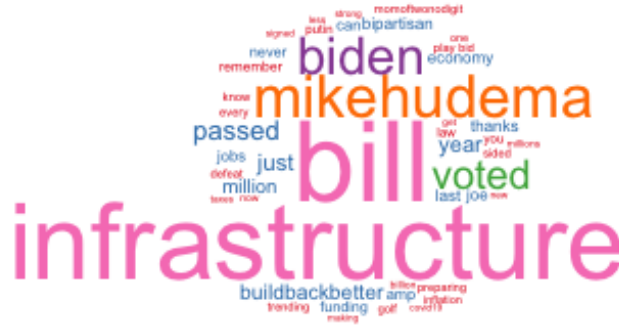

Figure 1. Word cloud for GND

Figure 2. Word cloud for IB

The word cloud for GND tweets appears to be richer – it includes keywords related to the economy (taxpayer, welfare, profit, prices), and energy (nuclear, fuel, fracking). Both word clouds feature the names of the politicians behind the policies – Bernie Sanders and Joe Biden. The other Twitter handle that is featured heavily in both word clouds is mikehudema. Mike Hudema is a climate activist, who has been very vocal regarding both GND and IB.

To find any patterns in discussions, I have applied structural topic modeling. I ended up identifying three topics for both GND and IB. The number of topics is arbitrary – meaning that I have settled on the number of groups based on running the models with a different number of topics and seeing whether adding or reducing the models adds any meaning to understanding the data. The results are presented in Figures 3 and 4.

The three key topics for GND could be summarized as:

1. Spending – tweets about government spending, welfare spending vs. military spending. Features a prominent supporter of GND and increased welfare spending – Bernie Sanders.

2. Discussion of the impact of the policy on the existing state of the economy – this topic includes tweets that are dedicated to the impact on industries, fossil fuel corporations, and infrastructure, as well as taxpayers. Along with the abovementioned Mike Hudema, it also features a Twitter handle of Melanie D'Arrigo – a climate activist and civil rights advocate, who is now running for Congress and is very vocal about the importance of GND.

3. Energy – includes a lot of the keywords related to energy – gas, oil. Interestingly, it features Joe Biden, who has not been a very vocal supporter of GND.
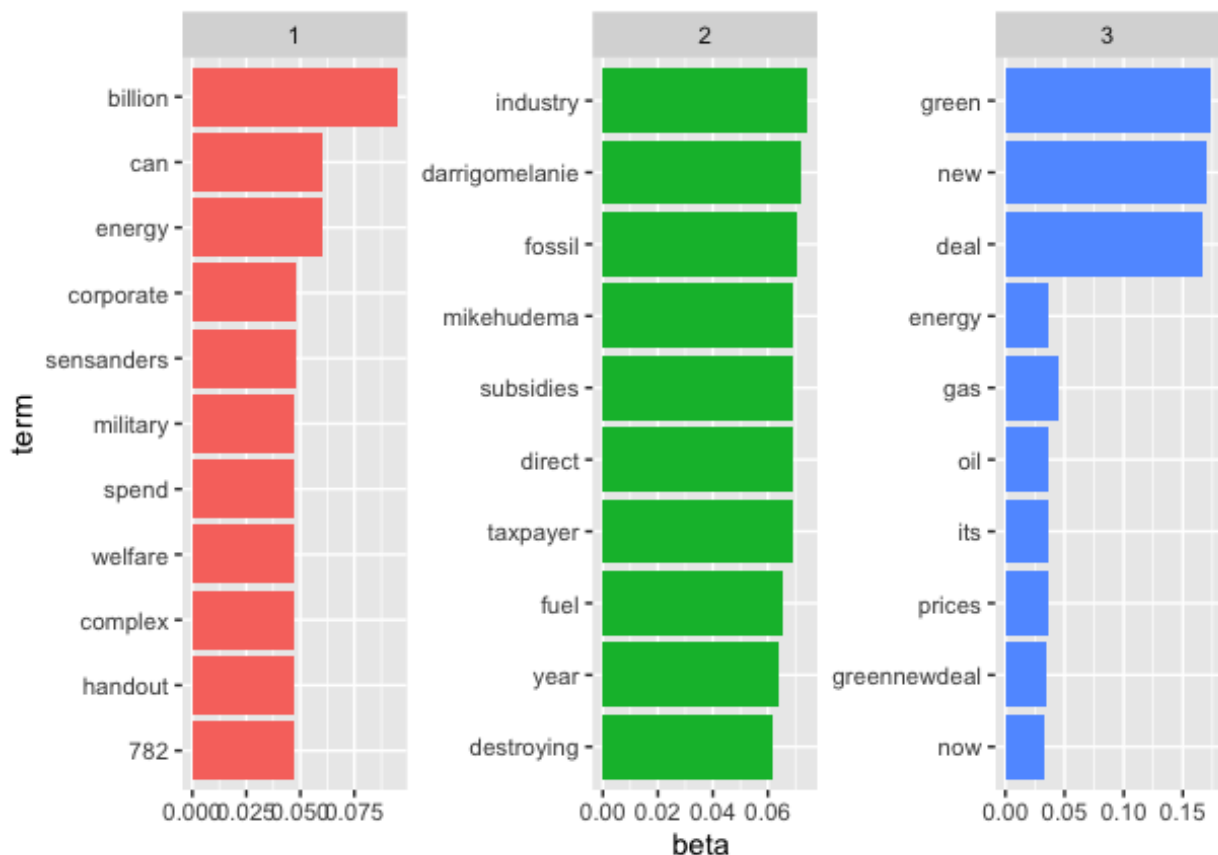
Figure 3. GND - Topic Modelling

The topics featured in IB related body of tweets can be summarized as follows:

1. The impact of the bill – bipartisan, securing jobs, its funding.
2. Controversies around IB – it includes the mention of Putin, as well as "playing golf" – a heavily exploited story by the Republicans, where they claimed that most of the IB spending in California leads to the renovation of Nancy Pelosi's favorite golf park (Calicchio 2021).
3. The topic highlights the merit of Joe Biden himself in passing the bill.

It is also interesting to look at the most discriminative words – i.e., those that are much more likely to be in one topic than the other. The results are presented in Figures 5 and 6. For GND we see a stark contrast in discussion around welfare spending vs taxation. For IB, there is a difference among the discussion of the Democrat's flaws pushed by Republicans – ties with Putin, playing golf, which is contrasted by the fact that the bill was passed, it was bipartisan and that was an achievement.
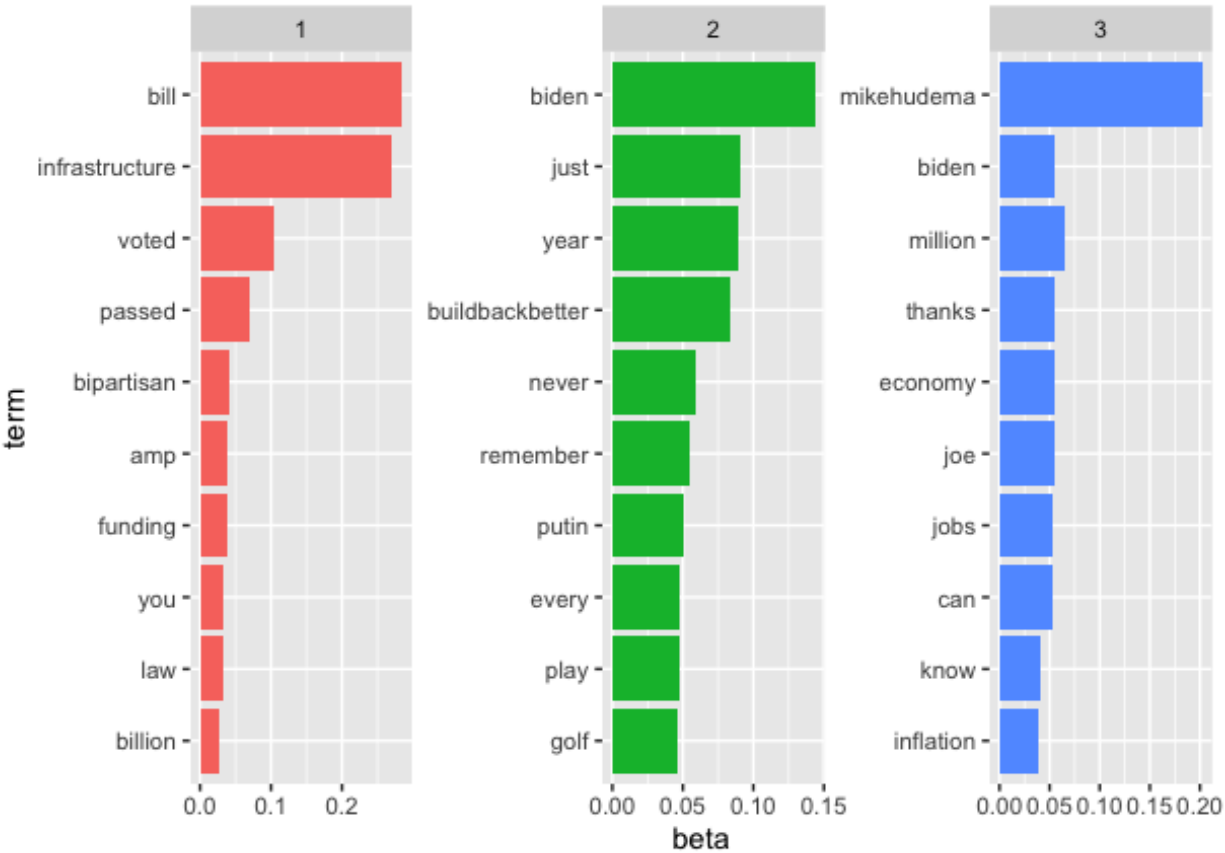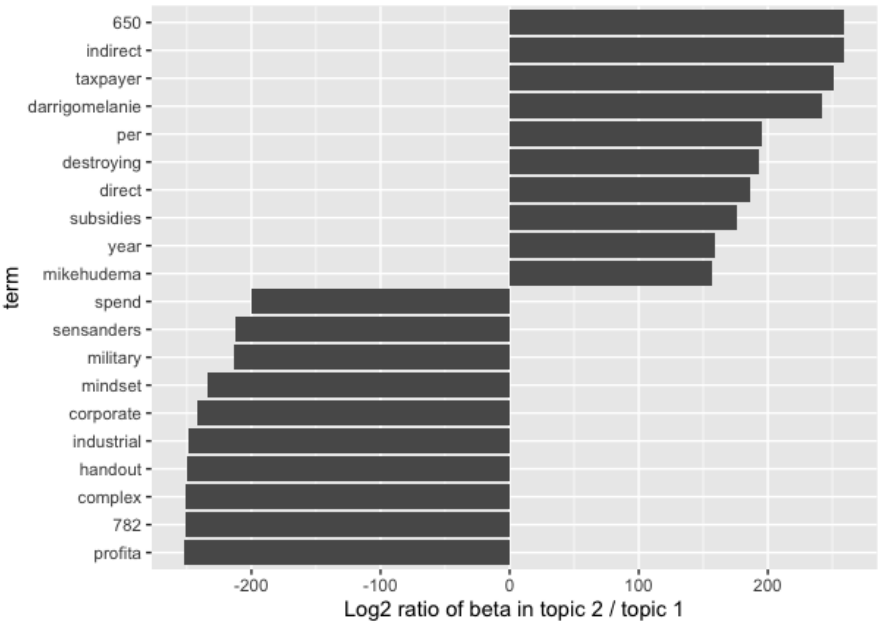
Figure 4. IB - Topic Modelling
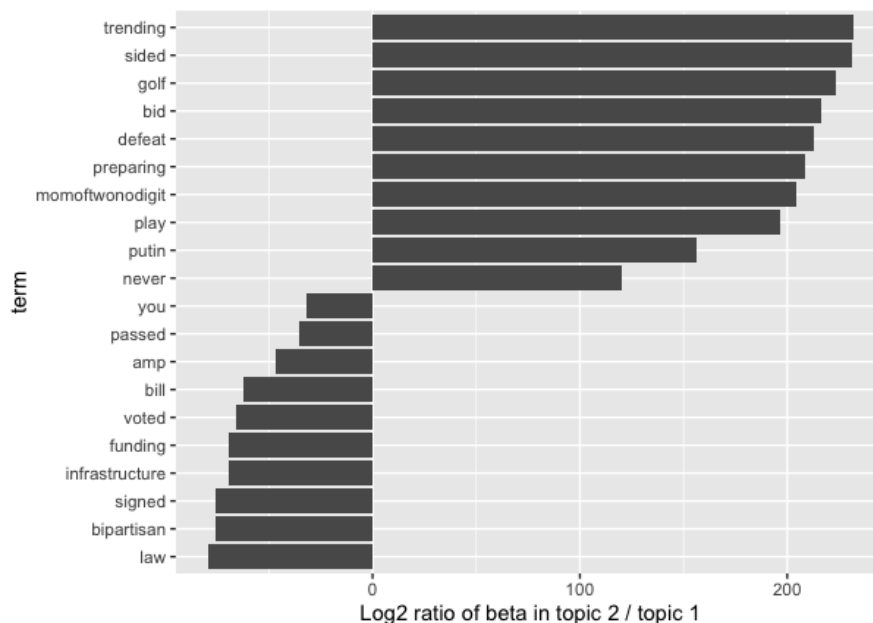

Figure 5. Most Discriminative terms (GND)

Figure 6. Most Discriminative terms (IB)

Lastly, I have built a CART decision tree to identify which key terms belong to the discussion around one policy or the other. The outcome variable was a binary variable, where 1 meant the tweet mentioned GND and 0 if it mentioned IB. The tree is rather short and it tells us that if the tweet does not mention the word "new" and "billion" then it is about IB.
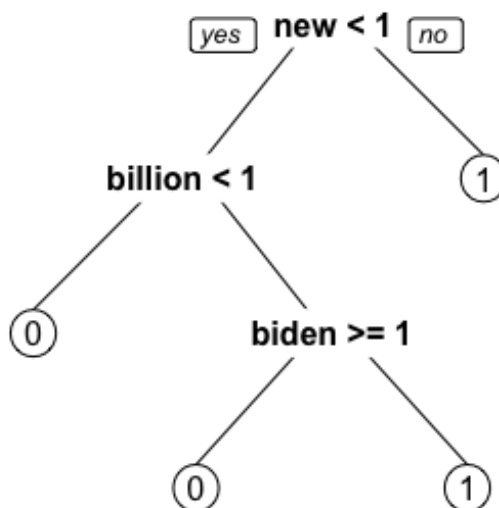

Figure 7. Prediction tree

Table 3. Accuracy of the Model

Baseline Accuracy	46%
Decision Tree	67%
Random Forest	0%

Table 3 presents the results for the calculations of the accuracy of the model. While the decision tree accuracy is higher than baseline accuracy (i.e., guessing at random), the random forest test is giving 0%. Most likely, it returns this result because my training set contains only one class.

Conclusion

In this essay, I have tried to explore the public discussion on Twitter around two climate-related policies – Green New Deal and Infrastructure Bill. I tried to employ various analytical techniques to get a better sense of the Twitter data. One of the biggest takeaways of my analysis is the fact that the discussion around GND – a set of policy proposals that have never passed Congress hearings – is more focused on issues directly attributed to climate policies (such as energy, spending, welfare), whereas the discussion around IB, which has become an actual bill signed by Congress and President – is concentrated around anything but the content of the bill. This can be explained by the fact that GND is designed to address the climate agenda in particular, whereas IB has more general points (improvement of existing infrastructure) incorporated into it.

Considering that GND is a directly climate-focused proposal, it was reassuring to see that the prevalent topics in Twitter discussions coincide with key dimensions of environmental policies which I used designing the experimental part of my Master's thesis project (the key topics are taxation and energy).

References

Calicchio, Dom. 2021. "Pelosi Deserves $200M for San Francisco Park, Dem Says: 'She Does More for America' than Others in Congress." *Fox News*. https://www.foxnews.com/politics/pelosi-200-million-san-francisco-presidio-park-congress (April 6, 2022).

IPCC. 2007. "Climate Change 2007: Synthesis Report. Contribution of Working Groups I, II, and III to the Fourth Assessment Report of the Intergovernmental Panel on Climate Change [Core Writing Team, Pachauri, R.K and Reisinger, A. (Eds.)]." https://www.ipcc.ch/report/ar4/syr/ (February 7, 2022).

Lachapelle, Erick, Christopher P. Borick, and Barry G. Rabe. 2014. *Public Opinion on Climate Change and Support for Various Policy Instruments in Canada and the US: Findings from a Comparative 2013 Poll*. Rochester, NY: Social Science Research Network. SSRN Scholarly Paper. https://papers.ssrn.com/abstract=2652429 (February 7, 2022).

McCright, Aaron M., and Riley E. Dunlap. 2011. "Cool Dudes: The Denial of Climate Change among Conservative White Males in the United States." *Global Environmental Change* 21(4): 1163–72.

Schumacher, Ingmar. 2014. "An Empirical Study of the Determinants of Green Party Voting." *Ecological Economics* 105: 306–18.