

Can I Hear Me Now?

Echo Chambers in Online Discourse

For several years, many have discussed the idea that politics has become increasingly partisan and divided. Nowhere is this more clear than in politicized discussion of the coronavirus, particularly in the United States. People disagree not just on the correct response to the pandemic, but indeed even whether the pandemic is itself real. More and more it seems that ‘reality’ depends on the lens through which one views the world.

In seeking to explain this phenomenon, people have pointed to the communication ‘bubbles’ we all live in online, where algorithms feed us content specifically tailored to us. This bubble allows people to surround themselves with news and opinions that confirm their biases rather than challenge pre-conceptions and build compromise between groups. Rather than learning to get along, participants convene instead to agree that it is those in the other bubbles who must be wrong.

While most discussions have focused on American politics, as a follower of Canadian politics I have increasingly felt this trend sweeping through online discussion of day to day issues north of the border, especially since 2016. More specifically, I have seen this on Reddit, a popular multi-topic discussion forum encouraging the sharing and discussion of news and media. Users of Reddit congregate in topic-specific forum pages known as subreddits. For example, followers of Canadian news and current events like myself might wish to visit the subreddit found at reddit.com/r/Canada, or ‘r/Canada’ for short.

Consistent with the polarizing trends observed globally, I’ve observed discussion on Canadian political subreddits to be increasingly hostile to dissent. More pertinently, some subreddits have seemingly evolved into ‘bubbles’ with discussion now exhibiting a partisan divide between users of r/Canada and another subreddit called r/CanadaPolitics.

As I considered this in the context of an online ‘bubble’ I began to wonder if such a divide had indeed taken place between these communities, or if I was projecting onto them a divide that exists only in my head. Conversely, perhaps these communities had always been divided, and I had just been too idealistic to recognize it in the past? To paraphrase a popular meme, I was in effect asking myself “*Am I out of touch? Or is it the users who are wrong?*”

The most appropriate response to that may be another meme (*Why not both?*) – but it wouldn’t be a terribly satisfying answer. So, like any good data nerd, I decided to try to find out more definitively whether Canadian political communities on Reddit have become polarized.

This article is broken into six sections:

1. **Selection of communities:** A brief summary of the Reddit communities selected for analysis.
2. **How would polarized versus non-polarized communities interact?** A discussion of what we would expect to see in a polarized versus non polarized community.
3. **Initial data:** An initial review and description of the data used in this analysis
4. **Hints of polarization:** Potential insights that can be gleaned from the data before more rigorous analysis

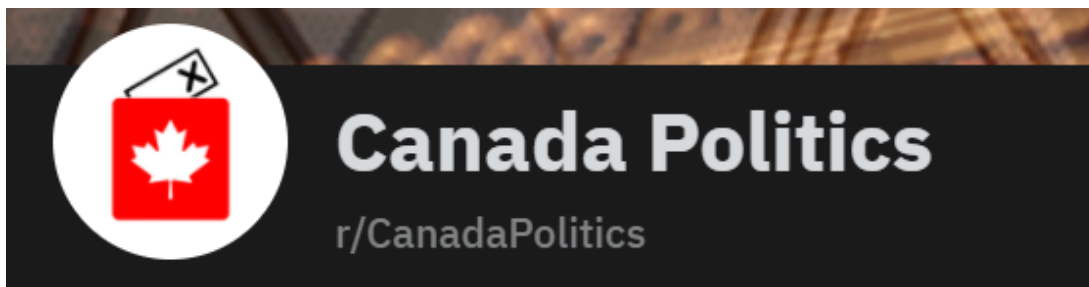
5. **Machine learning investigation:** Discussion of the results of applying a machine learning algorithm to the data to try to quantify polarization.
6. **Conclusions:** A discussion of the results and their implications.

Selected Communities

I chose to limit my analysis to just the two subreddits discussed previously, r/CanadaPolitics and r/Canada, since users of both subreddits post regularly, and neither have a declared partisan bias.

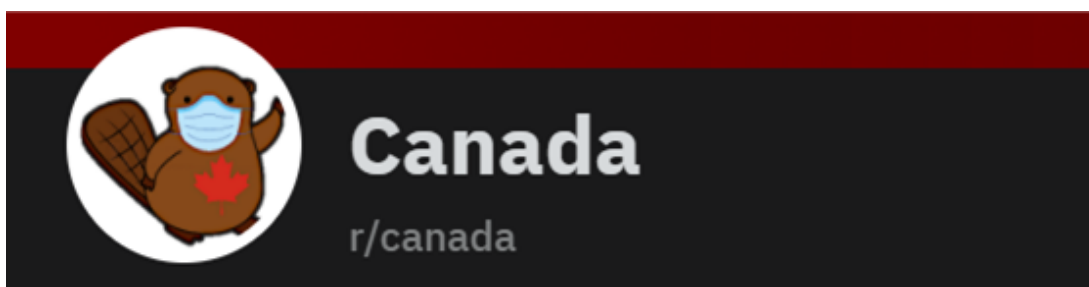
r/CanadaPolitics

The obvious first stop for Canadian political discussion on Reddit, r/CanadaPolitics, typically consists of news article links, followed by discussion. The topics must (unsurprisingly) be related to Canadian politics.



r/Canada

r/Canada consists of a similar style of content to r/CanadaPolitics; most posts are links to news articles followed by discussion. While the subreddit does not have a rule requiring the posts to be about politics, in practice many of the posts here are about politics or political news.



How would polarized versus non-polarized communities interact?

While participants in each subreddit would likely say the other has a specific partisan leaning, to a newcomer they all look like general discussion forums about the same topic without a specific tilt. Thus, you would expect to find the same content being discussed on each one – and in practice one often can find the same links posted to each of them at roughly the same time.

Where does that leave us in terms of polarization? Well, in an environment where these communities were not polarized, and instead represented a wide range of opinions, you would expect a new user to be relatively indifferent to participating in discussion on either. After all, the same content is generally being discussed on each one. You would also expect to find broadly similar points of discussion, and a similar range of views being expressed.

However, if these communities represent distinct viewpoints, a user may be more likely to gravitate towards the subreddit that most reflects their views, and we might expect to see qualitatively different comments being made on each one.

If such a polarization exists, one would need to examine the comments made in each community and determine if there were significant differences in tone or content. If one can easily distinguish which subreddit a comment comes from, based purely on the comment itself, then this is strong evidence that the communities are polarized.

Initial data

To analyze whether such a difference exists, I downloaded all comments made to each of these subreddits between January 1st, 2015, and September 30th, 2020. Full details for how I acquired the data, and the steps I took to analyze it are available in the companion process article that accompanies this one.

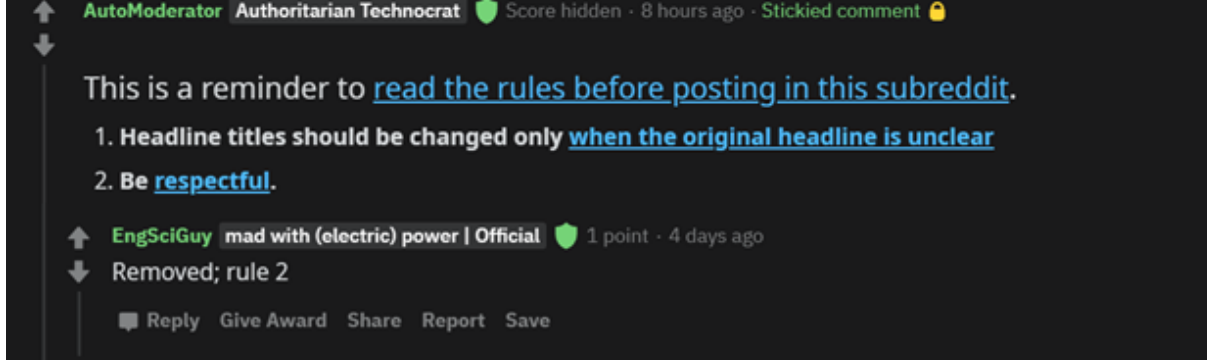
Specifically, my dataset includes the comments' text content, author, and publication-date, as well as details of the subreddit on which it was posted, the specific post to which it replied, and the media link that post contained.

Users can upvote and downvote both posts and comments. A comment's net score is total upvotes minus downvotes. For example, a comment with five upvotes and three downvotes would have a net score of two. This net score for each comment has also been added to the dataset.

I wanted to make sure that I was measuring differences in comments about identical topics, instead of capturing differences related to users discussing different events entirely. To that end, I restricted posts to ones which appeared in both subreddits. As many posts on both subreddits take the form of a link to a media article exclusively, with no commentary, I was able to accomplish this by ensuring my dataset contained only links which appeared on both subreddits. After performing this filtering, I was left with a dataset consisting of approximately 3.2 million comments.

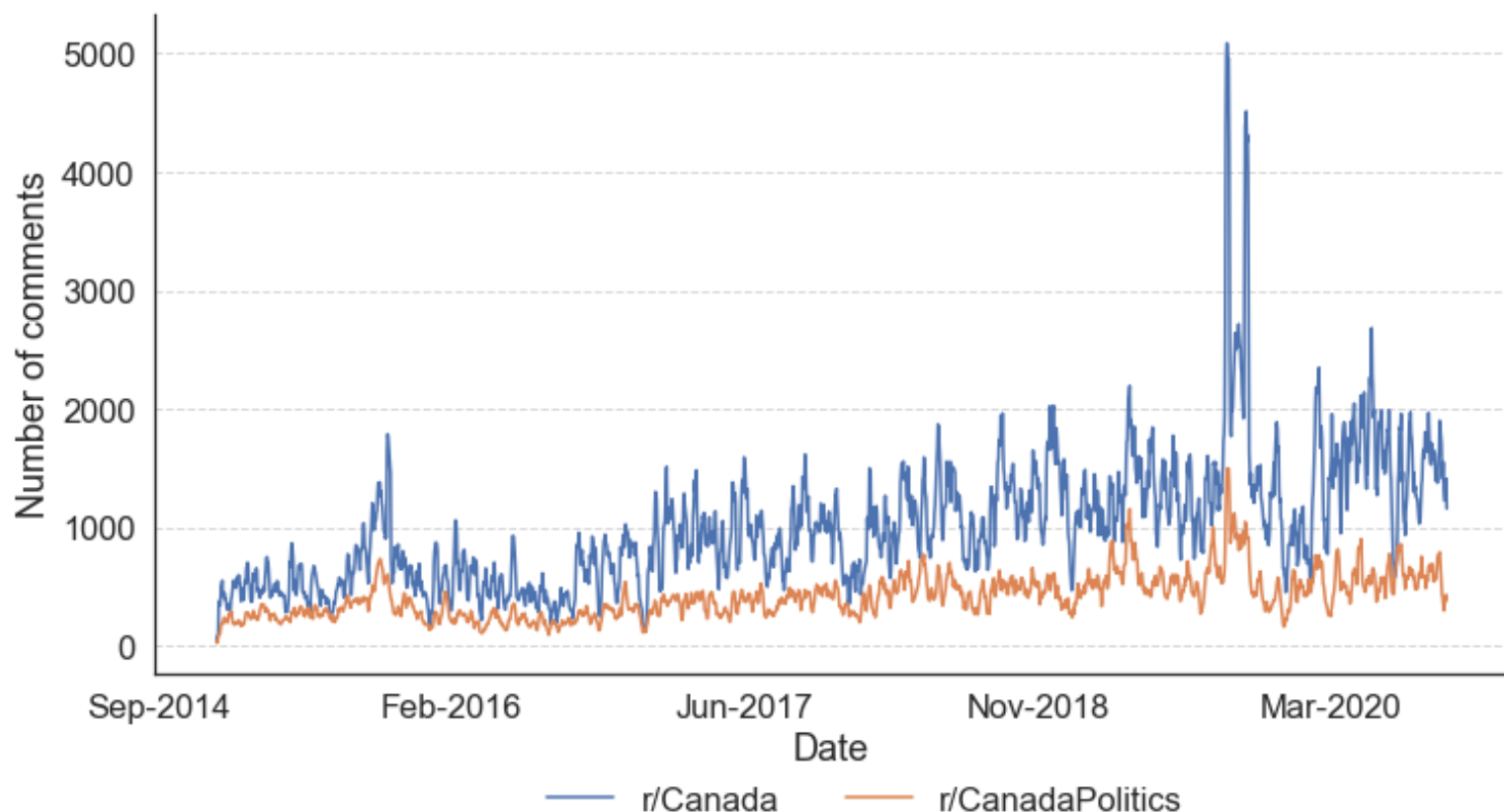
Next, I cleaned the data for content that was irrelevant to the discussion on these subreddits, and thus not indicative of the nature of the conversation. Examples include comments made by the Automatic Moderator bot, which reminds users of the rules when they make posts, or comments made by moderators explaining a specific action. Other examples might be spam, comments which parrot popular memes, or simple one-word comments like “agreed” which do not contribute ideas to the discussion.

Figure 1: Examples of comments removed from the dataset



Once this ‘noise’ is removed, I am left with a dataset consisting of approximately 3 million comments across both subreddits. The evolution of those comments over time is displayed below.

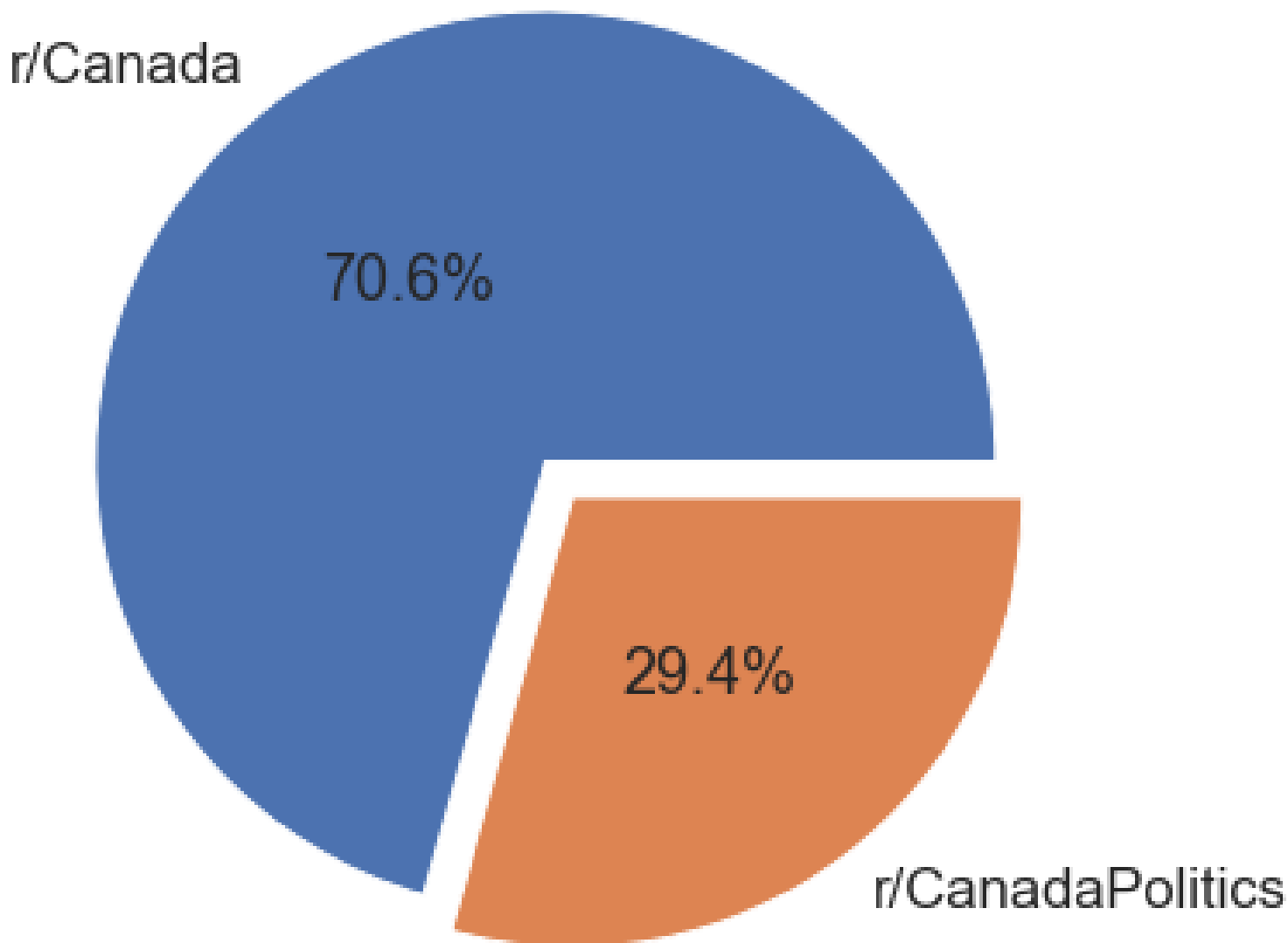
Figure 2: Number of comments over time



Here we see that the two subreddits display similar peaks and troughs, with spikes in October 2015 and October 2019 corresponding to the two elections held in Canada during the data collection period. There is also a larger uptick in discussion on r/Canada around March and April 2020, likely due to the coronavirus.

As is probably obvious from the graph above, most comments in the dataset are from r/Canada. The split of comments between subreddits is displayed below.

Figure 3: Comment count by subreddit



Hints of polarization

Previously, I noted that if the communities are polarized, we would expect users to gravitate towards the subreddit that shares their biases rather than being equally likely to participate in either one. Short of analyzing the text of all the comments on a given subreddit, might we be able to get some hints about whether they have become more (or less) polarized from the raw data itself? As it turns out, indeed we can.

Specifically, I hypothesize that a subreddit which is becoming **more** polarized should display at least one of the following four characteristics:

1. **An increasing average net upvote/downvote score on comments.** It is fairly uncontroversial to note that people downvote things they disagree with, and upvote things they agree with, even when a subreddit's rules discourage downvoting simply due to disliking something. If a subreddit is diverse and representative, we would expect that the

average net upvote score over time to increase at no more than the rate of increase in the subreddit's userbase, as the proportion of people who agree with the average comment should stay the same. If, however, the subreddit is becoming more polarized over time, we might expect the average net upvote score to increase more rapidly than the userbase of the subreddit, as people who agree with the average comment congregate in one place, and those who disagree leave the subreddit.

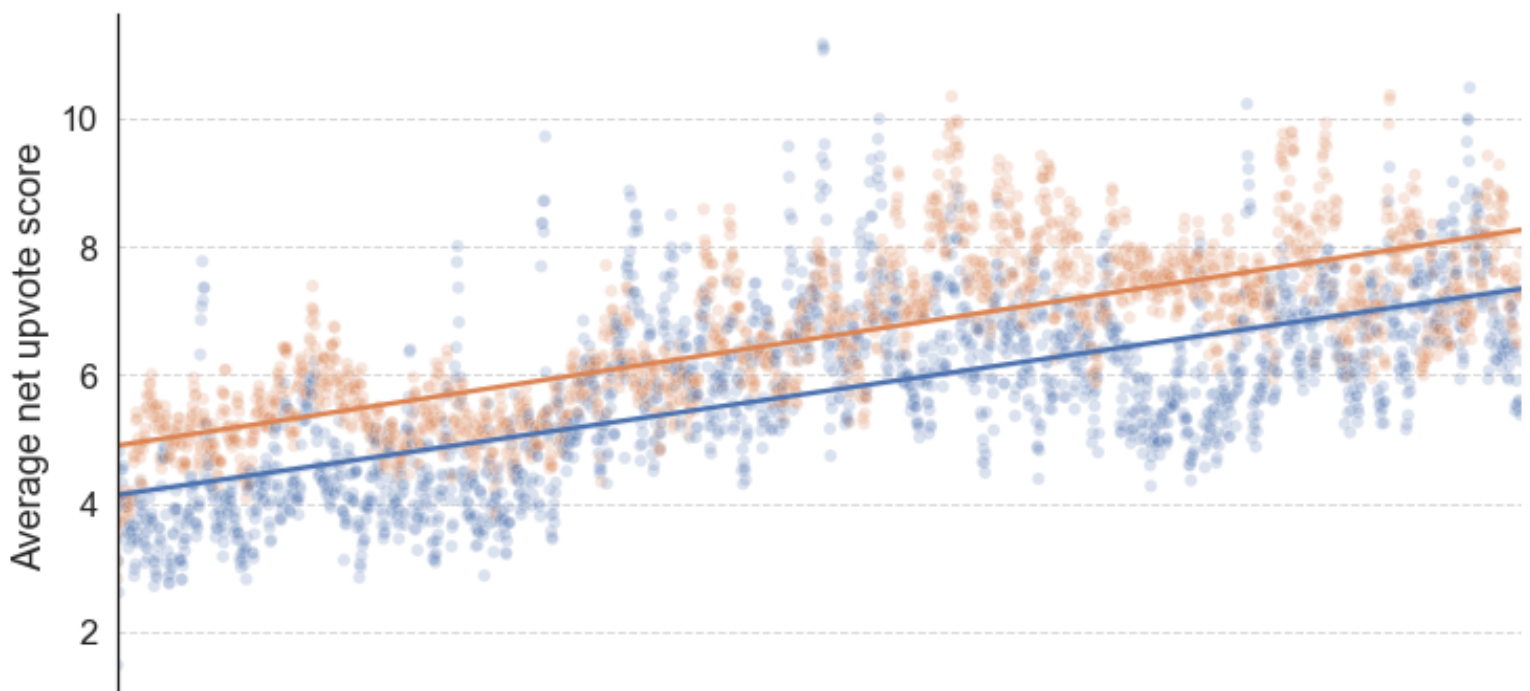
2. **An increasing rate of moderator removal.** Posts can be removed by moderators for a variety of reasons, but in general, the more posts that are removed, the more restrictive discussion becomes. While this may be desirable in certain instances – the removal of bullying content for example – given the political context of each of these subreddits, the judgement of what violates civility rules and what is simply an unpopular opinion is at risk of being quite subjective. In a more polarized community, mere unpopular opinions may be more likely to be viewed as offensive by some users and reported as candidates for removal.

3. **An increasing stratification of users between subreddits.** As both these subreddits discuss very similar topics and have no declared partisan bias, if there was very little polarization we would expect crossover between these communities to be quite high. If we observe that the user groups between subreddits have become less overlapping over time, this may be indicative of increased polarization between communities.

4. **An increasing divergence in key words and phrases.** If we were to construct a 'word cloud' from the comments of each subreddit, in a non-polarized environment we would expect to see similar sets of words and phrases contained within comments. In a polarized one, however, we would expect to see different words and phrases. For example, a subreddit that was supportive of the current government would be more likely to contain terms like "smart" or "proud". By contrast, an anti-current-government subreddit would likely be heavy on terms like "corrupt" or "laughingstock".

To be clear, the presence or absence of any of these things on their own (or even together) is **not** conclusive evidence to say that a subreddit has become more or less polarized. Instead, they serve as "hints" about the presence of polarization, and could suggest that further analysis may be fruitful. So what do we find when we go looking for these hints?

Figure 4: changes in the average net upvote score (hint 1)



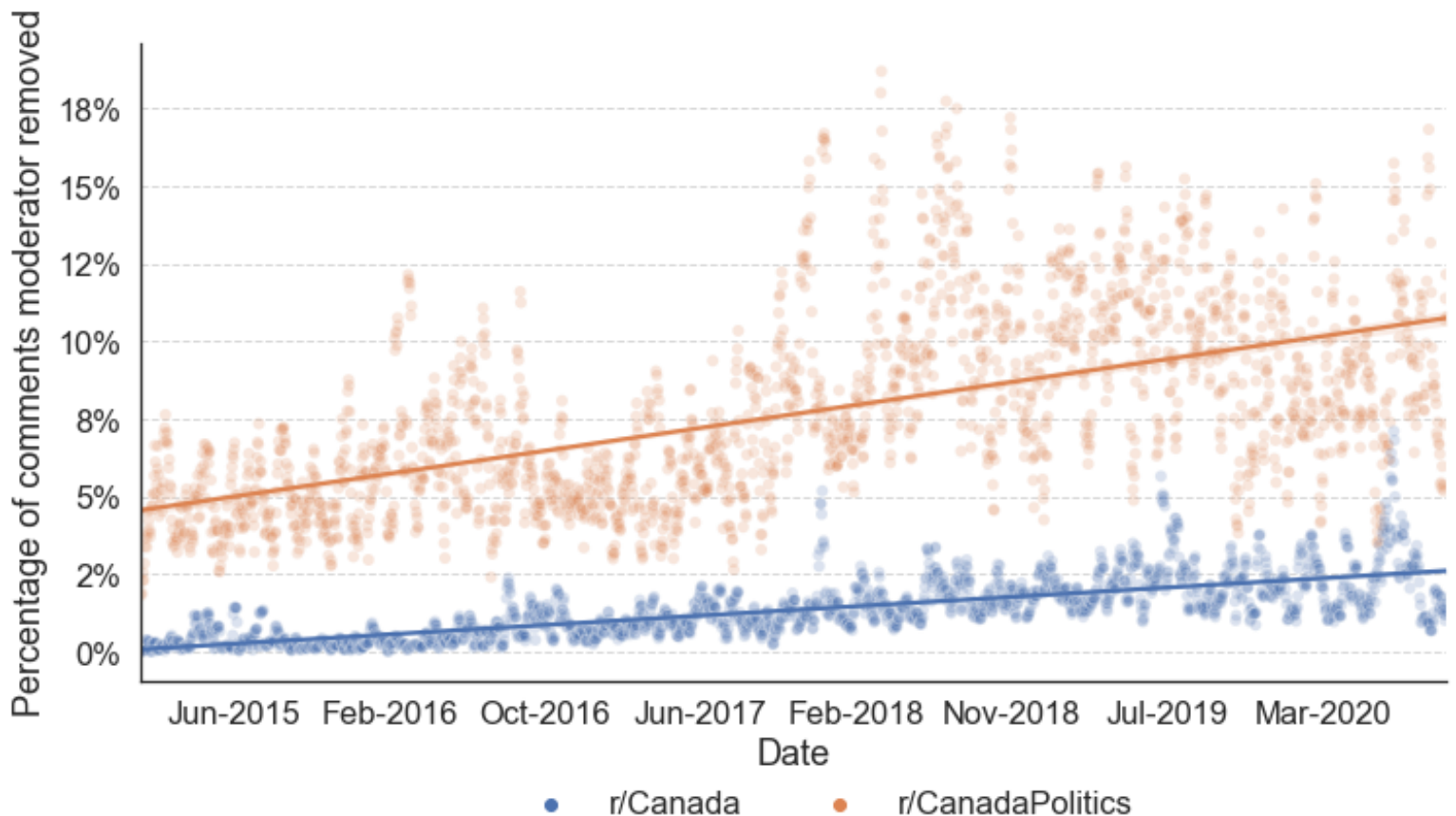


Here we see that across both subreddits, the average net upvote score has trended upwards over the course of the collection period. In the case of r/Canada and r/CanadaPolitics, the net score has gone from hovering around or just above 4 and 5 respectively, to being around 7 and 8, or a 60-75% increase depending on the subreddit.

For r/Canada, this rate of increase are not hugely out of line with the growth in the subreddit's userbase over time (discussed further below). However, r/CanadaPolitics grows at a much slower rate, but its net upvote/downvote score has increased at roughly the same rate; this implies that there may be more at work here than simple growth in the size of the community.

Given that it seems likely that increasing upvotes are driven by the subreddit's userbase increasingly agreeing with the comment in question, and they have grown at similar rates in terms of net upvote/downvotes, but not in terms of userbases, these trendlines imply that at least one of these communities may be more of an echo chamber today than it was several years ago.

Figure 5: Changes in the rate of moderator removal (hint 2)



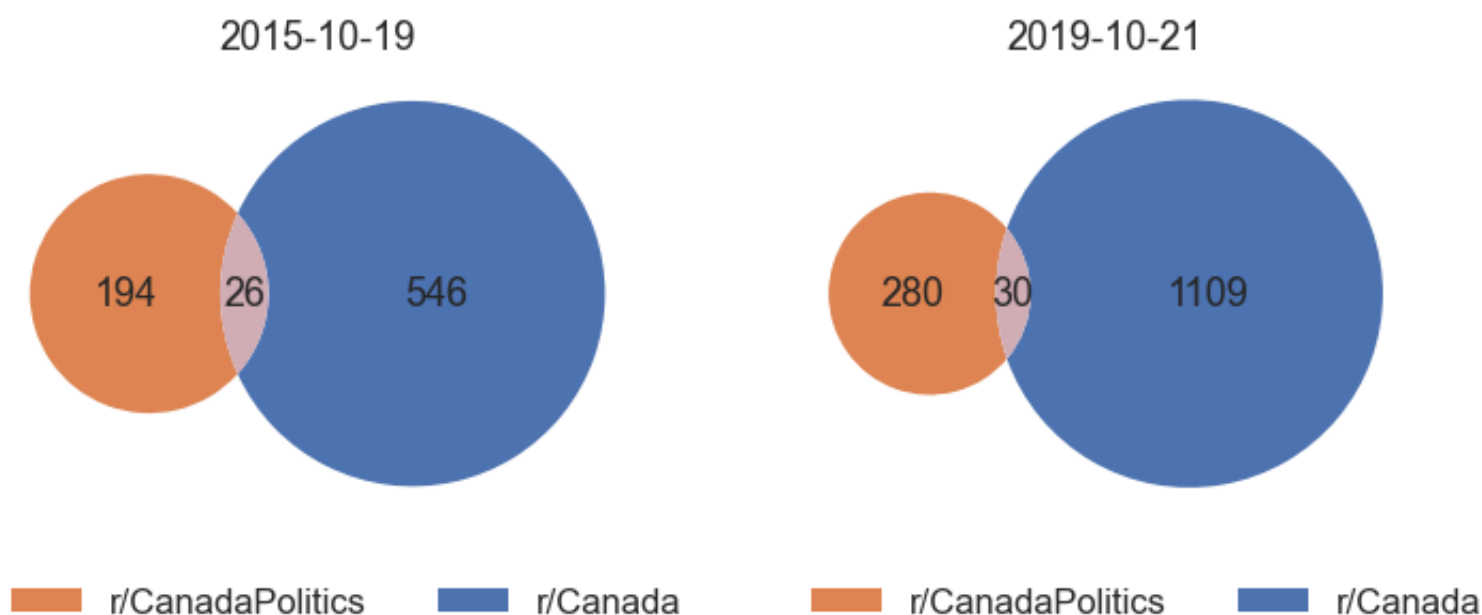
Here we see the different rates of moderator removal for comments in any seven-day period. It is immediately obvious that r/CanadaPolitics has a much more active moderation team in terms of removing comments than r/Canada. The higher

base count is not terribly surprising, as the method of removal differs between these subreddits. r/Canada's moderators will remove specific offending comments, while r/CanadaPolitics moderators remove the **entire thread** an offending comment creates. Thus, the higher **base** rate of removal does not imply that r/CanadaPolitics is polarized.

What **does** perhaps hint at increased polarization is the **substantial** increase in the rate of removal for r/CanadaPolitics. While both subreddits have increased (r/Canada goes from removing almost no comments in 2015 to removing around 2% of all comments today), the rate at r/CanadaPolitics has gone from around 5% in 2015 to around 10% in 2020, with wide variation depending on the week. In some weeks almost one in five of all comments made are removed by moderators.

While this **may** be a hint of increased polarization in both communities, and particularly in r/CanadaPolitics, changes in removal rates may also be reflective of changes in the rules of the subreddits themselves. Given the stark difference between the rate of increase, it seems unlikely that increasing removals are accounted for by changes in sitewide policy from Reddit itself.

Figure 6 – Changes in stratification between subreddits (hint 3)



It appears that both subreddits have grown substantially between two key events in the noted time period: the 2015 and 2019 federal elections. The above venn diagrams are reflect election day in each year, when we would expect high traffic on both subreddits. Here, each circle denotes the number of unique commenters on each subreddit averaged across a seven-day period. The crossover regions are the users who posted in **both** subreddits.

Crossover in 2015 is quite small, with just under 12% of r/CanadaPolitics users commenting on r/Canada and those same users also made up just over 4.5% of r/Canada users. The userbases of r/Canada appears to be about two and a half times as large as r/CanadaPolitics. By late in the 2019 election however, things have changed. While roughly 10% of r/CanadaPolitics users still post in r/Canada, the reverse relationship is only just over 2.5%, or just about half of what it was before.

Figure 7: Evolution of user crossover

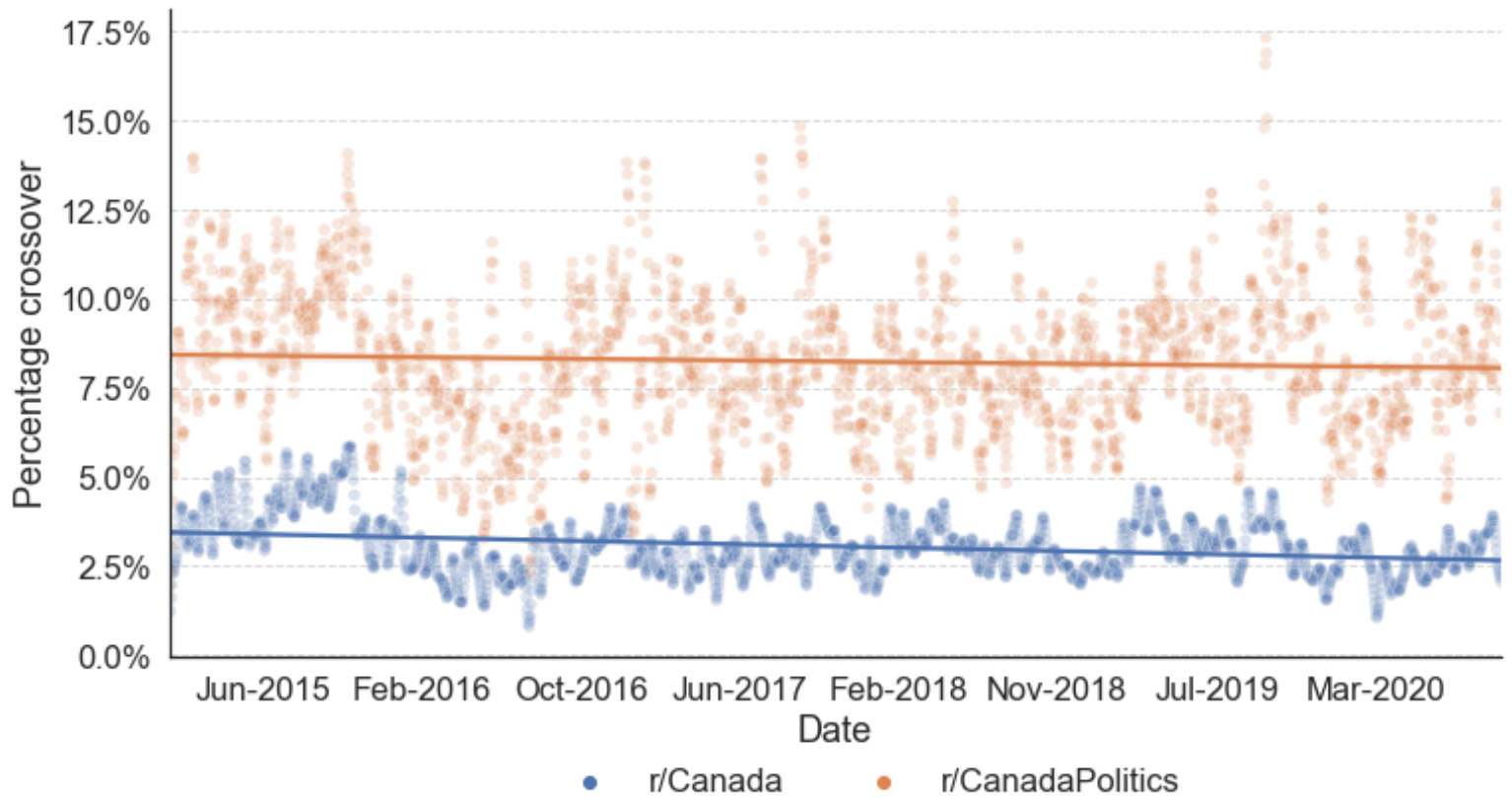
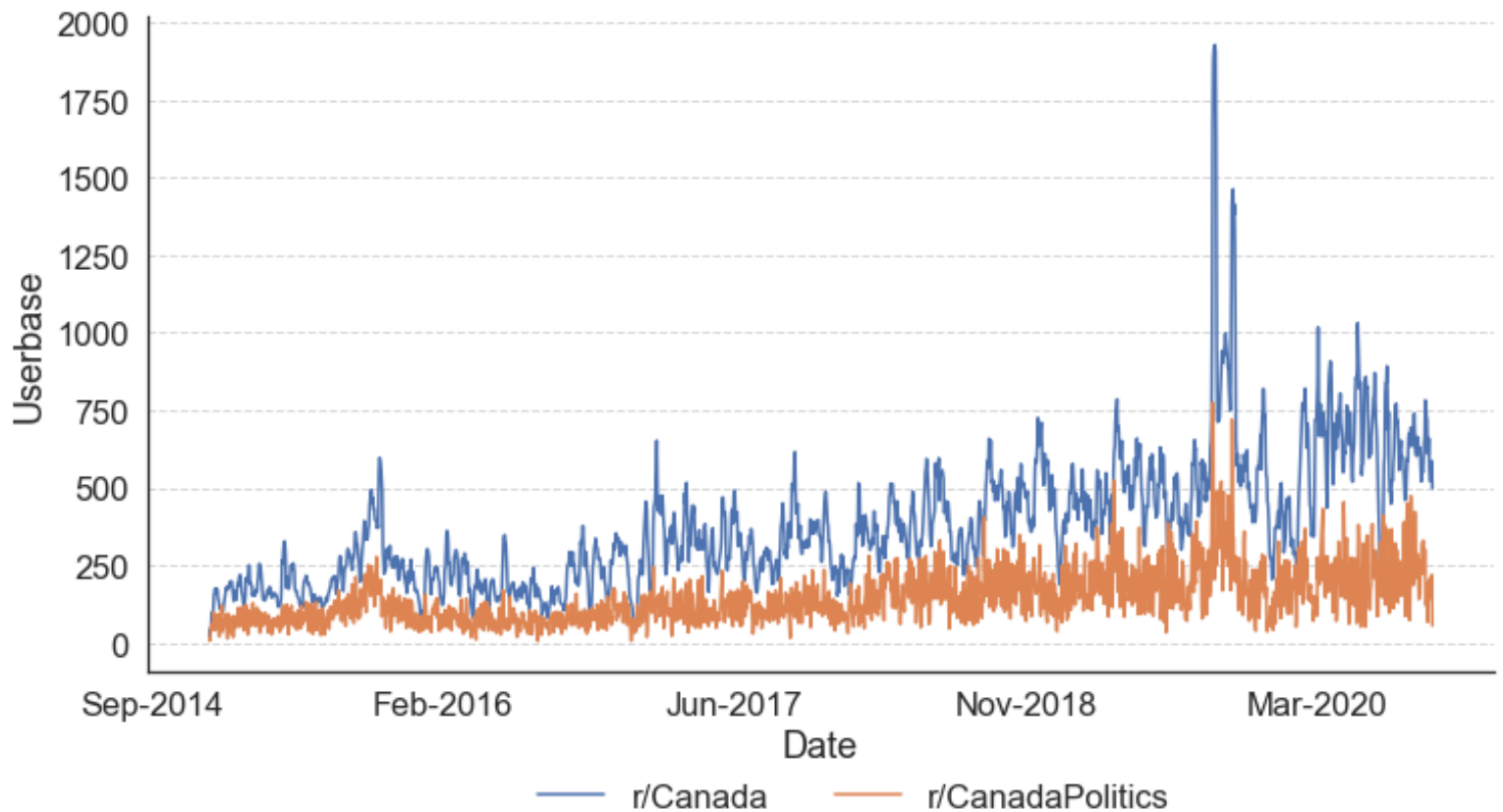


Figure 8: Evolution of subreddit active userbases

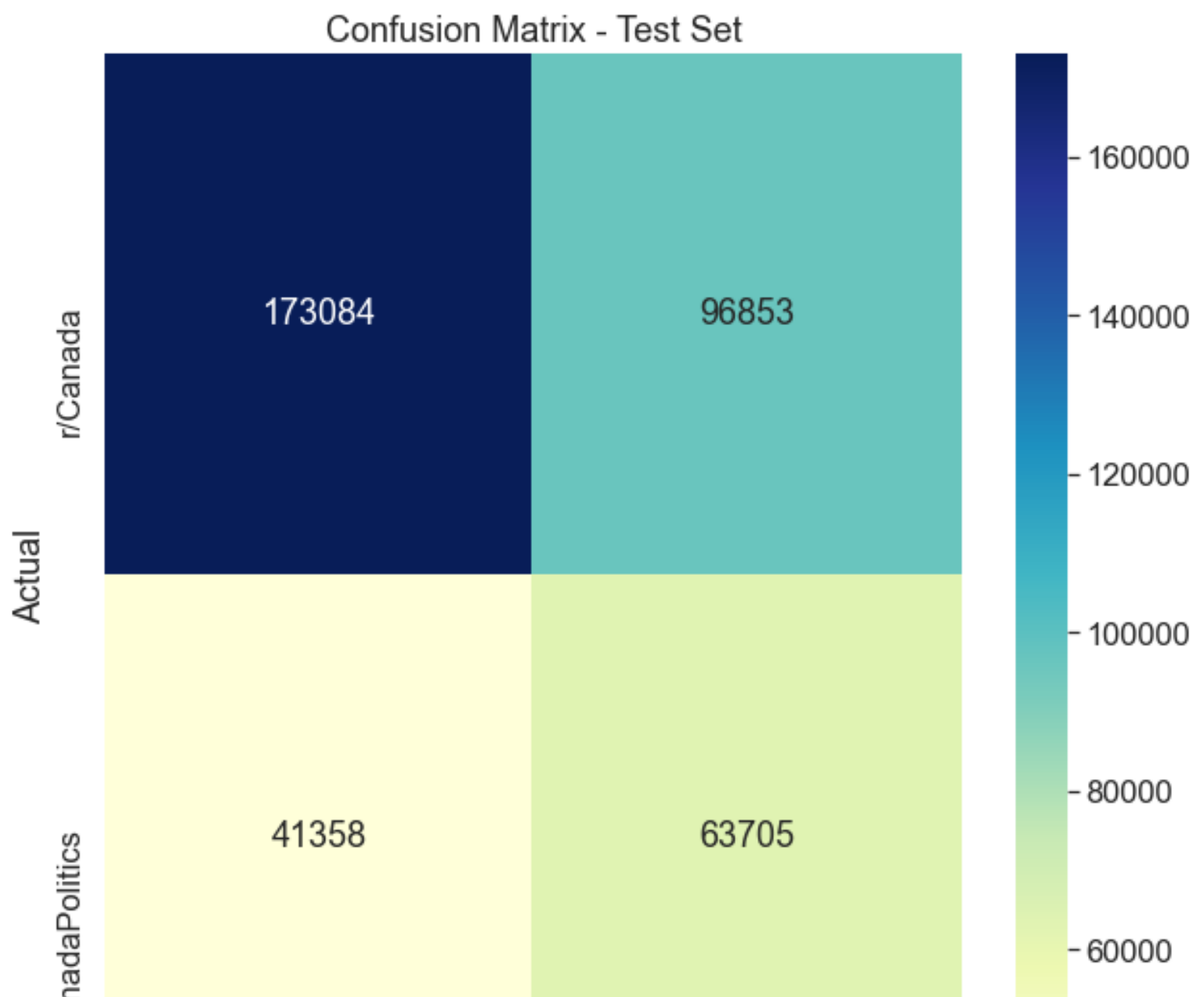


If an algorithm can, in later years, more easily classify a particular comment as belonging to r/Canada or r/CanadaPolitics, I consider this evidence that the communities have become more distinct over time. For instance, if the machine learning algorithm was able to accurately classify 90% of comments made in 2020, but only 50% of comments made in 2015, the implication would be that comments in 2015 are much harder to distinguish between subreddits. Thus, the accuracy of the technique itself over time is the metric by which I will be evaluating how distinct the communities are. A better performing model implies more mutually distinct (and internally homogeneous) communities.

To briefly summarize, I considered three different classification algorithms: a Linear Support Vector Machine, Naïve Bayes, and Logistic Regression. Details on the steps I took to select a model are available in the accompanying process article. For the purposes of this summary article, I have presented the results of the model which performed best: the Logistic Regression.

A key output for this kind of classification problem is the “confusion matrix”, a grid representation of the accuracy of the algorithm’s classifications.

Figure 10: Confusion matrix



r/Canada

r/CanadaPolitics

Predicted

This is an example of a confusion matrix. The top left box records the number of comments which were from r/Canada that the model **correctly** predicted were from r/Canada. The bottom right is the same, but for r/CanadaPolitics. The top right box has comments that the model **incorrectly** predicted were from r/CanadaPolitics, but were actually from r/Canada, and the bottom left is the reverse. Thus, all other things being equal, we hope to see high numbers in the top left and bottom right corners, and low numbers in the top right and bottom left corners.

There are several ways to evaluate a model. One such way is **accuracy**, which calculates the portion of all predictions that are correct. In the case of the above confusion matrix, that would be:

$$\frac{173,084 + 63,705}{173,084 + 63,705 + 96,853 + 41,358} = 63\%$$

However, as we saw in the previous section, over two thirds of the comments in the dataset come from r/Canada. Thus, a model which predicted 100% of comments as belonging to r/Canada would achieve an accuracy score of 70%; this is better accuracy than the example confusion matrix above, but clearly not a useful model. Thus, accuracy on its own can be a misleading metric.

As a result, I have measured model performance using an alternative metric: the “balanced accuracy”. This metric is defined as the average of the “recall” for both r/Canada, and r/CanadaPolitics. The “recall” is intuitively the algorithm’s ability to find all positive instances of a given category. Because the sample is imbalanced between the two subreddits, considering the model’s overall recall is not desirable, as out performance on r/Canada comments could cloud the overall results. The “balanced accuracy” approach helps to alleviate this problem by taking the average between two recall scores, thus penalizing the model if it underperforms on comments from r/CanadaPolitics. A more detailed discussion of evaluation metrics is available in the companion methodology document.

When evaluating these metrics, it is important to avoid “overfitting” one’s model. An overfit model is one which is too well calibrated to the data used to train it, and so performs very well on that data, but performs very poorly on unseen or “test” data. Checks for goodness of fit for this model are present in the accompanying process article, but for the purposes of this article, the model performs at very similar levels for both training and test data, and so is not overfit.

Below I have displayed the evolution of balanced accuracy score on a six-month rolling average to help smooth volatility and discern trends.

Figure II: Model performance over time



The first thing to note is that while the model is not perfect, as its balanced accuracy is in the low sixties in percentage terms, it is not an entirely **unskilled** model because its balanced accuracy score is above 50%.

A higher balanced accuracy score tells us that the model is better able to categorize comments into the appropriate subreddit based upon their text alone. Intuitively speaking, this would then imply that where the balanced accuracy score goes up, comments on the two subreddits are more distinct, despite the sample being limited to posts discussing identical topics.

Thus, the movement of the balanced accuracy line in the above graph can tell us a story about differences between r/Canada and r/CanadaPolitics. In 2016, comments became much easier to distinguish between the two subreddits, implying that they were diverging in terms of discussion and content. But then, in 2017 and 2018, comments became steadily harder to distinguish, implying that the communities were becoming more similar. This difficulty peaked in early autumn 2018 before reversing. Comments were easier to distinguish in 2019, though not as distinct as in 2016.

Conclusions

At the beginning of this project I set out to try and measure if these two communities were becoming polarized. I would have liked to have been wrong, as I do not view such polarization as a positive development. Unfortunately, the model is not entirely unskilled, which implies that there is indeed polarization between the two subreddits.

Given that 2019 was a year in which a major government scandal was revealed and an election was held, it is especially concerning that the model implies increasing divergence compared to the previous year. Of all times, an election year

is one when it is most important for people to discuss and share ideas with their broader community as individuals decide how they wish to vote. However, it appears that during this time, the content of discussion on the two subreddits, which had been growing more similar, instead started to diverge again, implying that there may have been notable partisan differences between the two communities.

It is disheartening to note that, as we saw when we examined the stratification of users across these two subreddits, there is extraordinarily little crossover between the two communities. As such, the increased distinctiveness of each community seems less likely to support each as a platform for communicating their perspective to others, and more likely to suggest each may be becoming an echo-chamber that reinforces the homogeneity of the community.

It would therefore appear that Canadian political discourse – on Reddit at least – is not immune from the balkanization that has been observed in other contexts.

However, the results are not all gloomy. Were the communities **hyper** distinct, the model should be able to achieve a much higher balanced accuracy score than it has. Additionally, while the downward trend in polarization did reverse in 2019, it has not reached the heights demonstrated in 2016. While it is possible that a more sophisticated technique (e.g. deep learning) may be able to detect differences with greater accuracy, it is also possible that there simply isn't enough polarization to make this task easy, telling us that while these communities may have their differences, they also still have much in common.

If differences continue to grow, however, that commonality may shrink. In the increasingly charged political climate we are all living in, results like this can serve as a useful warning to politicians, community leaders, and active participants in online debate, perhaps helping us to combat – or potentially reverse – any growing divides between fellow citizens.