Forecasting and Understanding the 2021 Canadian Federal Election Using Twitter Conversations

Shainen M. Davidson[†,*], Vaibhav Kesarwani[‡], Kenton White[†]
[†] Advanced Symbolics Inc, 109 York Street, Ottawa, ON, Canada
[‡] Novacene AI Corp., 123 Slater Street, Suite 610, Ottawa, ON, Canada

Abstract

Previous studies using Twitter data to predict election outcomes have focused on sentiment and/or volume of tweets as predictive features. However, it is difficult to determine whether a tweet is expressing views in favour or counter to a political entity. By contrast, in this study deep learning is used to cluster related tweets together into "conversations", and then those conversations are classified. Classification is achieved by identifying who is involved in the conversation: if a majority of partisans of one political party are present, then the conversation may be classified as being in favour of that party. The conversation is then predicted to change public opinion, relative to size of the conversation. This method has the added explanatory benefit of linking moments of large public opinion change directly to the source conversation, answering "why" a change happened. Time-series public polling data from the 2021 Canadian federal election is used to validate the model, and the final result of the election was forecast with the model the day before the election. It is observed that changes in public opinion seen in polling data are seen earlier in time in Twitter "conversations".

Keywords: NLP, Twitter, Social Media, forecasting, politics, elections

1. Introduction

Using social media data to predict elections is a common, though elusive, goal ([1] identified 787 such studies since 2010). The majority of this work focuses on the sentiment and volume of tweets around a party or candidate, with mixed success [2]. A chief difficulty is determining the stance of a political tweet, i.e. is the user saying something positive or negative about the political entity. Political speech is often complex and context dependent, so sentiment is an unreliable indicator.

However, when seeking to understand how political opinion changes over the course of an election campaign, one naturally looks to issues and events, and how the public conversation around those events pushes people's views on those up for election. Social networks, especially Twitter, are a primary arena in which those conversations occur. In this work those conversations, extracted with unsupervised deep learning, are used as the basis to predict change in public opinion over time.

On a daily basis, tweets mentioning political entities are collected from a sample of Twitter users. These tweets are clustered into groups according to their textual contents using deep learning. Each of these clusters, i.e. "conversations", collect a similar way some people are talking about an issue or event and the political entity mentioned. The partisanship of a subset of users in the twitter sample is determined based on who they follow. The political implication of the cluster is then achieved via the amount of political partisans in the conversation: if enough partisans of another party are engaged, then the cluster is classified as positive for that party (and negative for the party being talked about). In this way, the difficulty of classifying an individual tweet is avoided. The magnitude of the

[*]shainen.davidson@advancedsymbolics.com

change in political opinion is proportional to the number of non-partisan users involved in the conversation[1].

A further advantage of this method is the explanatory benefit. To understand why a party gained or lost support on a particular day, the largest pertinent conversations can be investigated, either with NLP summary tools, or simply by qualitatively examining the text. This feature is especially valuable to political professionals attempting to change public opinion during an election (as well as media and researchers trying to understand why an election is going a certain way). This also provides a simple way to qualitatively check the direction of effect the model predicts a particular conversation will have on public opinion due to partisan involvement.

The method was validated by comparing to public polling data leading up to the 2021 Canadian federal election, and then used to predict the result. There was general agreement among numerous pollsters about the general trend of public opinion for the major two parties over the campaign, which provides a reliable "ground truth" against which the method may be validated. And indeed, the Twitter conversation model does show the same trajectory for the major political parties, in fact even showing the changes in opinion days in advance of when the same change appears in polling. The Twitter data was used to forecast the election result on September 19th 2021, the day before the election. After passing the popular vote prediction of the "conversaition" model through a seat allocation algorithm (see [3] for details) the model correctly predicted the winner (Liberal), but over predicted the Liberal seat count by 8 seats and under predicted the CON seat count by 21 seats.

Previous work understanding political opinion using clustering techniques used Latent Dirichlet allocation (LDA) topic modelling. [4] uses LDA to retrieve keyterms predictive of constitutional referendum in Italy, and [5] uses temporal multinomial LDA to investigate issues related to the 2012 Korean presidential election. LDA creates topics with a bag-of-words (BAG) approach, which means that clusters are defined by containing similar words, regardless of how the words are used in the text. For this study, the clusters are to be classified for the political position they imply, so a clustering method that is sensitive to how the words are used is required. Language representation models, such as BERT [6], use deep learning architecture trained on massive language datasets to become sensitive to word usage. They have some of the best performance on stance-detection tasks [7]. Here this technology is used to extract political "conversations" that are sensitive not just to the general topic, but to the position on that topic.

2. Twitter Data

2.1. Twitter sample

The Twitter data for the study is pulled from a sample of 290,641 Canadian Twitter accounts. The sample was created using the network sampling algorithm Conditional Independence Coupling (CIC) [8]. The sample is limited to Canadian accounts using the user reported location field: the text is sent to a geocoder, and only those which return a point within a province in Canada are retained in the sample. See White [9] for a discussion on the representativeness of such a sample. The tweets are collected for each user using the Twitter user history API. Using a sample means that the tweets gathered on any day of an election will be the full set from the users of the sample, and will thus represent an accurate accounting of the number of people engaged on the election. This has not been the case if

[1] Limiting the size of impact to non-partisans will avoid having big effects due to conversations where partisans are "preaching to the crowd" instead of talking about issues in a way swing voters would.

Party	Query Terms
LIB	Trudeau, JustinTrudeau, Liberal, LPC, liberal_party, parti_liberal, TeamTrudeau
CON	Erin O'Toole, Scheer, erinotoole, AndrewScheer, Conservative, CPC, CPC_HQ, PCC_HQ
NDP	Jagmeet Singh, theJagmeetSingh, NDP, New Democratic Party, NPD_QG, InItForYou
BQ	Yves-François Blanchet, yfblanchet, Bloc Québécois, Bloc Quebecois, BlocQuebecois, BlocQc, paysQC
GRN	Annamie Paul, AnnamiePaul, Green Party, GPC, CanadianGreens, LesVertsCanada, ElizabethMay, JenicaAtwin, PaulManly
PPC	Bernier, MaximeBernier, PPC, PPC, peoplespca, ppopulaireca, BernierNation, MadMax

Table 1. Search terms

the Twitter streaming API is used[2], as Twitter provides only a subset of relevant tweets, and the ratio of that subset will vary, leading to an inaccurate representation of the number of people engaged on a topic on any given day [10]. See [3] for further discussion of the creation of the Canadian CIC sample.

2.2. Election Twitter data

Tweets mentioning six political parties were collected for the 2021 Canadian federal election: the governing Liberal Party of Canada (LIB), the official opposition Conservative Party of Canada (CON), the Bloc Québécois (BQ), the New Democratic Party (NDP), the Green Party of Canada (GRN), and the People's Party of Canada (PPC). Only the tweets of the larger parties (LIB, CON, BQ, and NDP) were eventually included in the model, since the use of partisans to classify conversations means the method works best for well established parties. The smaller parties are mentioned in this section for completeness, and since the amount of tweets over time for the smaller parties is an interesting feature to report.

The sample is queried every day to create a daily dataset for each party from August 1 to September 19th, 2021 (the day before election day). For each of five parties tweets are collected using keywords in Table 1 per day. The datasets range in size up to 13,342 on September 9th for LIBs (see Figure 1), and in total 642,242 tweets are analyzed.

2.3. Partisan users

Twitter clusters will be identified by who is in the cluster, specifically how many "partisan" users; i.e., people who have a longstanding positive view of the party, and are motivated to improve the fortunes of that party. It is likely that public social media posts from these accounts about the partisan's party will be positive, while posts about the competing parties will be negative.

Partisanship is determined looking at the number of "guaranteed partisans" a user follows. The list of "guaranteed partisans" is collected manually from existing lists of twitter accounts belonging to users who work for a party (e.g. members of Parliament, chiefs of staff). The final list is 1,290 users and has members of GRN, BQ, CON, LIB, and NDP. Then users of

[2]At time of writing, a "100% firehose stream" option is listed as "Coming soon" on Twitter's API website, https://developer.twitter.com/en/docs/twitter-api/migrate/twitter-api-endpoint-map, so this may not be an issue in the future.

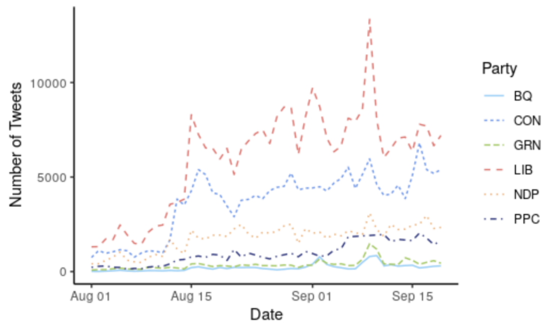

Figure 1. Number of tweets collected per party per day. Liberal peak on August 15th (when election was called) and September 9th (English Leaders' Debate).

the CIC sample are classified as partisans if they follow at least 5 "guaranteed partisans", and over 50% of them belong to a single party. For the case of LIB partisans, a more strict cut-off of greater than 75% LIB "guaranteed partisans" is used; since the LIB party has been in power since 2015, user are more likely to follow LIB party members for non-partisan reasons (e.g. for their ministerial role). Using the higher cut-off leads to a more balanced final number of partisans.

Using this criteria, there are 18,185 users classified as partisans in the sample: 6934 LIB, 5445 CON, 4709 NDP, 644 BQ, and 453 GRN. That this roughly matches the popular vote of the parties is an indication that the criteria is on the right track.

3. Method

For a set of tweets (queried by party and date), "conversations" are extracted, i.e. clusters of tweets with similar text. Within each cluster, the number of tweets by partisans is used to determined whether cluster is good or bad for given party (i.e. is this cluster of tweets representing one party "winning" an aspect of an issue). Then the party who has "won" this conversation gains net public opinion proportional to the number of non-partisan users engaged in the conversation, at the expense of the party whom the conversation is about (i.e. public opinion here is a zero-sum game).

3.1. Clustering Tweets

Clustering involves the clustering or grouping of any pre-processed text in an unsupervised manner. The number of clusters or themes is unknown and varies from one dataset to another; therefore traditional supervised learning methods cannot be applied. A supervised algorithm is tailored to a specialised dataset, whereas an unsupervised one is good at approximating and extrapolating on unseen data.

Most clustering methods can be broken down into two steps. First, a method to convert text into a mathematical object such that it makes sense to talk about the "distance" or the similarity between text segments. This generally involves converting text to a vector, i.e. vectorization. The second step is a method to determine how to choose the clusters, i.e., what are the criteria such that two text segments are close enough that they belong in the same cluster.

While vectorizing text for clustering is a well-studied problem, much of the work has not been optimal for short text segments on social media, such as Tweets. [11] discusses the analysis of open-ended answers from two surveys: the "10 Years After 9/11 Survey" and the "National Cultural Building Survey". The 9/11 survey asks about specific rights and freedoms in the US that respondents would tell people from other countries, e.g. "Express yourself" or "Freedom to vote", etc. The clustering methodology used for this survey is K-means after removing common English stop words and specific keywords. This methodology does not provide an adequate solution to our problem as removing stop words would impede language skills for the vectorizer. Removing stop words will not impact this specific dataset as the responses are not conversational and contain broken phrases. Further, the authors have manually removed words like "freedom", which works for this specific dataset but cannot be generalized for other data. Instead of removing specific words, a more elegant approach would be to use a TF-IDF based word weighing algorithm that automatically does this. Further the length of this dataset is just 1-2 words. This algorithm would not generalize for longer sentence data. [12] introduces three new methods for topic modelling of open-ended survey responses: LF-LDA, BTM, and WNTM. The authors also use LDA as a baseline against which they compare these methods. The initial part of the paper explains the methods in depth and provides justification for their use. The authors conclude that the methods do not lead to optimal results but rather serve as a starting point for further research. The quantitative evaluation done in this paper does not include metrics that are common for unsupervised analysis like silhouette coefficient or Fowlkes-Mallows score despite having 5,000 labelled instances. This paper provides context but not a foundation onto which to build a solution. [13] compares 14 different vectorization methods including TFIDF, Word2Vec, GloVe, ELMo, BERT, ALBERT, XLNET, S-BERT, S-RoBERTa, and more. The paper tests these methods on two different Twitter datasets. The authors postulate that for noisy datasets like social media content, models trained on standard English corpora do not perform well and some pre-training on Twitter data should be performed in order to achieve better performance. This paper provides a good starting point to carry out further experimentation using the BERT, RoBERTa and XLNET algorithms, which provided the best performance amongst all evaluated methods. For this study, BERT vectorization [6] was chosen for further experimentation on clustering method.

For the clustering methods, baseline experimentation was done on K-means [14], Birch [15], and Agglomerative [16], but all of these algorithms need an input cluster number before they cluster the results. The challenge with providing a cluster number is that even after applying a cluster number deduction algorithm like the Elbow method, the clusters formed were not representative of the data density (or distribution), instead dependent on the count distributions. Further, DBScan [17] was applied, which is a density-based algorithm and does not require the cluster count as input. However, the results with DBScan were not balanced and the clusters formed varied vastly in size and composition. To overcome this, a custom multi-pass hierarchical clustering application was implemented that iteratively forms clusters based on the density of the data. Table 2 shows a comparative evaluation of various clustering algorithms with the hierarchical clustering algorithm. It was evaluated on Q12 dataset, that is a labelled dataset and therefore we used Fowlkes-Mallows[18] index to compare them.

The hierarchical clustering algorithm first runs a BERT [6] vectorizer on the preprocessed text and converts it into vectors. Semantic similarity is computed on these vectors by doing a dot product on the matrices.

$$a.b = \sum_{i=1}^{n} a_i b_i = a_1 b_1 + a_2 b_2 + ... a_n b_n \tag{3.1}$$

Then in the first pass of the clustering, it groups all vectors based on semantic similarity distances into approximate clusters. This pass results in a significantly unbalanced distribution of the elements where cluster sizes vary vastly and there is considerable overlap between the clusters. Hence, subsequent passes iteratively smooth down the cluster distribution. Moreover, all elements in the boundary zone are candidates for reallocation. Since

Algorithm	FM Score
K-Means	0.166
LDA*	0.202
DBScan	0.344
HC BERT Sim	0.368
HC BERT Sim Lem	0.380
HC BERT Pass 1	0.452
HC BERT Pass 1 v2	0.596
HC BERT Pass 1 v3	0.574
HC BERT Pass 2	0.599
HC BERT Pass N	0.625
HC BERT Pass N v2	**0.631**

Table 2. Results for various clustering algorithms tested on labelled survey Q12 dataset in ascending Fowlkes-Mallows scores. * Except LDA, all the algorithms use the same BERT vectorizer to facilitate equitable comparison.

this is a density-based scan, there is an unclustered elements list that contains all elements that do not fall in any cluster. Including them in the final distribution results in a significant decline in the Fowlkes Mallows score [18]:

$$\text{Fowlkes-Mallows index} = TP \div \sqrt{(TP + FP) \times (TP + FN)}$$

where TP is the number of True Positive
FP is the number of False Positive
and FN is the number of False Negative.

Fowlkes Mallows index is used for external evaluation and Silhouette Coefficient [19] for internal evaluation:

$$\text{Silhouette Coefficient} = \frac{(b-a)}{max(a,b)}$$

where a is the mean intra-cluster distance
and b is the mean nearest-cluster distance.

Since the density-based hierarchical clustering algorithm does not take cluster number as input before processing, there is no implicit control on the size of the clusters formed. In algorithms that take cluster number as input, increasing the cluster count results in finer clusters and vice versa. To enable this functionality for hierarchical clustering, a "cluster resolution" parameter was introduced which varies from 1 to 4. At 1, the algorithm forms a few big clusters, and at 4 it forms many finer clusters. Some tweets which do not meet the clustering condition remain unclustered; the amount of unclustered tweets will vary by resolution (see Figure 2).

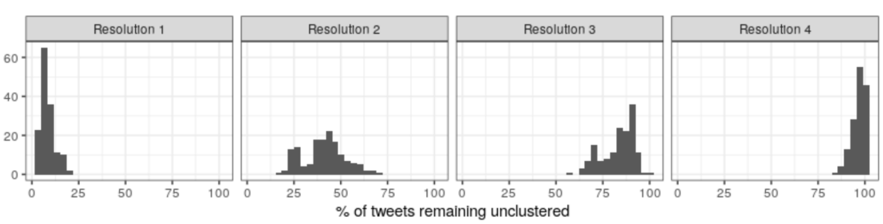

Figure 2. Histograms of the ratio of tweets remaining unclustered in daily political datasets by resolution parameter. For resolution 1, the conditions for tweets to be grouped together are permissive, and thus few large clusters are formed and very few tweets remain unclustered. For resolution 4, the tweets must be close in the vector space to be clustered together, thus small clusters are formed and many tweets remain unclustered.

To determine which resolution is optimal, recall the goal is to classify clusters by the participation of partisan actors. We wish to maximise the number of clusters which can be clearly labelled as representing the viewpoint of one party, i.e., the majority of partisans are of a single party. Resolution 3 has the most clusters which can be classified this way (81%), so this is the clustering resolution used in this study.

3.2. **Classifying conversations and connecting public opinion**

On each day, the number of non-partisan users who engage in topics good/bad for each party is used as a proxy for how public opinion has changed positively/negatively for that party.

For each tweet cluster, the ratio of partisan tweets

$$R^c_{P=p} = \frac{NT^c_{P=p}}{\sum_k NT^c_{P=k}} \tag{3.2}$$

is calculated, where $NT^c_{P=p}$ is the number of tweets of party p in cluster c. For clusters where

$$R^c_{P=w} > 0.5, \tag{3.3}$$

for a party w, the cluster is considered to be a conversation "won" by party w. This threshold allows only one party (or none) to be the winner.

The value of a cluster c (i.e. how much an effect it should have on public opinion) is

$$V^c = \frac{NU^c_{NP}}{NU_{day}} \tag{3.4}$$

where NU^c_{NP} is the number of users in cluster c not identified as partisan and N^U_{day} is the number of users in total talking about politics that day (this scaling is to prevent over reliance on days when Twitter traffic is high).

The value of the conversation is added to the total of the "winning" party w, while the value is subtracted from the party whom the conversation is about q (i.e. the party keyterms queried for the dataset):

$$V^c_{P=p} = (\delta_{pw} - \delta_{pq})\, V^c. \tag{3.5}$$

Thus the change in conversation is a zero-sum game among the parties (and a conversation "won" by a party about itself does not affect public opinion). If we begin with all the parties at 0 on August 1st, the change in pubic opinion, $O^T_p(d)$ for party p on day d is

$$O^T_{P=p}(d) = \sum_{t\leq d,c} V^c_{P=p}(t), \tag{3.6}$$

given in Figure 3. The most noticeable trend is a sudden increase in public opinion for CON at the expense of LIB starting August 12th, followed eventually by the reverse beginning around September 15th.

The change in opinion $O^T_p(d)$ has no meaningful scale. Since it is only a measure of *change* of opinion, there is no information on the opinion baseline to deviate from. The unit of the change is a ratio (non-partisan users / total users per day), so a scaling factor needs to be applied to convert the data to a form to compare to external data for validation.

4. **Results: 2021 Canadian Federal Election**

Longitudinal data is needed to validate the relationship between the twitter conversation model and public opinion. The public opinion poll with the largest sample size, the actual election, unfortunately only gives one time point in the election campaign. Publicly available polling data using traditional techniques is the next best option. It may be that there are biases in the various polling methods; nonetheless, the changes over time using the same

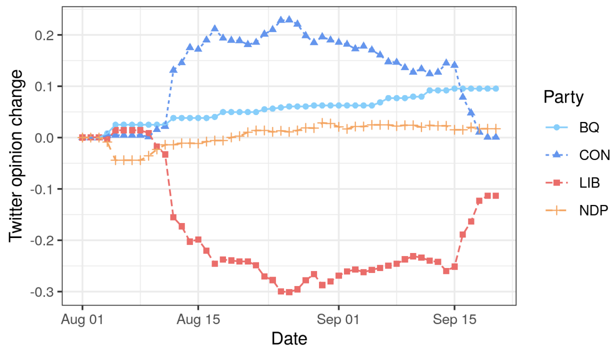

Figure 3. Public opinion change predicted with the Twitter conversation model.

method should give reliable information within the bounds of sample size errors. To give a sense of what this looks like, the results of the six polling firms with the highest frequency of polls during the campaign (out of 14 collected) are plotted in Figure 4. Looking at the polls together, only a few major trends in the major parties[3] are consistent across pollsters: Prior to the election, LIB is ahead of CON. Then sometime around the beginning of the election period (August 15) LIB begins falling and CON rising until about the end of August. Then the reverse occurs, and the election ends with the LIB and CON close together. To make this precise, a loess filter is applied to the polling data to smooth out the difference in the polls and only leave the consistent features (Figure 4).

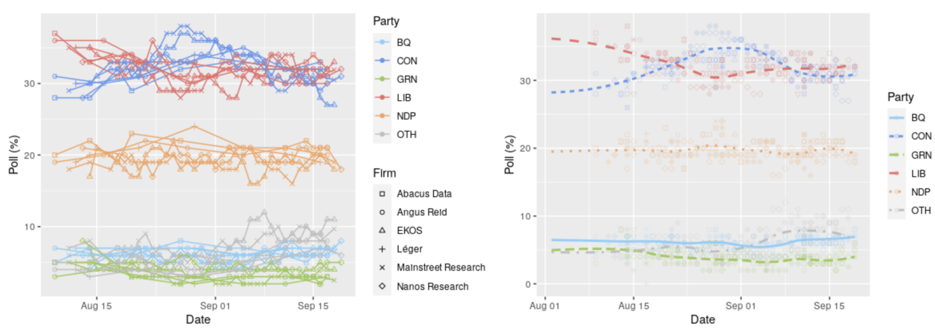

Figure 4. Public polling for the six polling firms with the highest frequency of polls during the campaign to illustrate change over time within each firm (left) and what the change looks like after applying a loess filter to extract common features (right).

The filtered polling data is used to provide a baseline and scaling to the Twitter data $O_p^T(d)$. $O_p^T(d)$ is scaled to match the mean and the variance of a smoothed version of publicly available polling:

$$O_p^S(d) = (O_p^T(d) - \mu_T)\frac{\sigma_p}{\sigma_T} + \mu_p, \tag{4.1}$$

where μ_T and σ_T are the mean and standard deviation of $O_p^T(d)$ over d, and μ_p and σ_p are the same for the smoothed polling average.

[3]There is also agreement that the OTH (Other) vote rises over the course of the campaign. This is the PPC, and their polling support trends similar to number of tweets about them (perhaps because they are a smaller party). As they are a smaller less established party, without obvious partisans at this point, the present method is not well suited for them.

These common trends in the polling data are manifested in the twitter data (Figure 5). In fact, the majority of the loss of the Liberal lead happens quite suddenly, with the biggest change occurring on August 12th, three days before the election is officially called. This loss of support from LIB to CON continues until August 26th, when the reverse trend begins (although not as sudden).

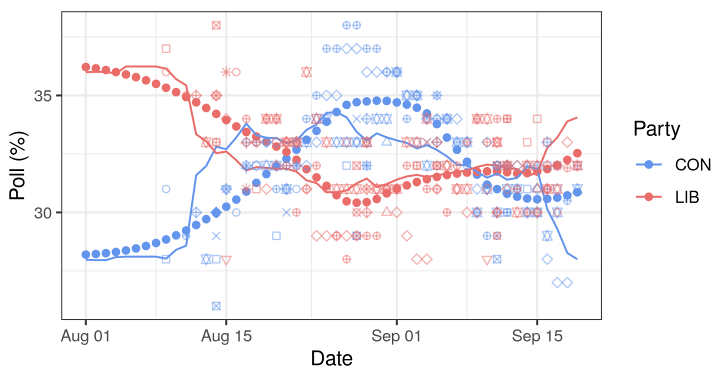

Figure 5. A comparison of the opinion change from the Twitter conversation model $O_p^S(d)$ (solid) with the loess polling average (dots). Individual polling firm results are marked as well.

A final feature we see in the twitter data which is not born out by all the polling is a sudden shift from CON to LIB in the final few days of the campaign (more on that later).

Unlike traditional polling, as the data comes from opinionated text segments an examination of *why* the change in opinion occurred is possible. In Table 3 the conversations causing the largest changes in public opinion are listed with summaries of conversation: one generated by summarising the tweets with Google's T5 Text-to-Text Transfer Transformer [20], and another written manually using the output of T5 and checking with the contents of the tweets. Note they are all negative, as the zero-sum game rules requires that a positive conversation (i.e. a conversation where the majority of partisans are talking about their own party) cannot create a flux of public opinion from one party to another.

The largest single change occurs on August 12th, and that conversation was concerning the LIB party calling a snap election that was seen as politically motivated (similar conversations occur over the next few days). This is perhaps surprising, as the snap election was not yet triggered. However, it turns out that there were enough "leaks" to the media about the coming election decision that it was being reported before the Prime Minister visited the Governor General to officially dissolve parliament. The fall in LIB numbers appears in the Twitter data before it shows up in any public polling. This could be for a number of reasons: Public polling results are collected over a several days so will lag public opinion, while the Twitter data is instantaneous to the day. It is also possible that Twitter data is pulling from a segment of the electorate that is more engaged and thus their opinion precedes the opinion formed by people who participate in polls. To tease apart the cause is difficult without access to the micro-data of public polling broken down by date.

The slow decline in CON numbers, from August 26 to September 10, similarly precedes the average decline in public polling, although in this case not as definitively. The data does not show a single dominant theme. The biggest single conversation is on August 28th, dealing with protesters at LIB events which used violent language and made the CONs look bad.

The final large shift in the numbers from September 15 to September 19 is a sharp decrease in CON numbers to the benefit of the LIBs. Here the subject is COVID: First, a

Date	Party	Δ (%)	T5 Summary	Manual Summary
2021-08-05	NDP	-0.41	Andrea Horwath backtracks on her comments to CBC News yesterday. she now supports mandatory vaccina NEW: NDP Leader after her Charter rights statement on yesterday:I fully support mandatory vaccin Two groups opposed to mandatory vaccination: 1) Ontario Nurses Association 2).	Horwath backtracks on opposing mandatory vaccines
2021-08-12	LIB	-1.70	cdnpoli Trudeau is expected to call snap election for sept. 20. it's time we had a Prim Global News can confirm the Reuters, CBC and La Presse reporting that prime minister is likely to visi he is planning an election in the middle of the pandemic because his focus is on politics.	Calling a snap election for political reasons
2021-08-13	LIB	-0.52	a few people who proclaim one should Never Vote Conservative! are apparently voting for the Liberal party. the first thing you should ask is: Is this about you, or is it about Voting for anyone but CPC ensures an Liberal majority? Ivison: Trudeau is about to call an election with no purpose other than to free him from parliament's fetters.	Trudeau is about to call an election with no purpose
2021-08-14	LIB	-0.74	cdnpoli's Justin Trudeau will visit Rideau Hall and meet with Gov. Gen. Mary Simon on Sunday, according to his official itinerary. he is expected to request parlia Yup tomorrow morning, paving the way for a expected federal election.	Trudeau calling an unnecessary election
2021-08-17	LIB	-0.41	canada has 'no plans' to recognize the Taliban as the legitimate government of Afghanistan, says Trudeau. cdnpoli says he will wait and see how the thugs work out o' toole, a recognized terrorist organization under Canadian law, to get people out of the Afghanista.	Fallout from Afghanistan
2021-08-18	CON	-0.44	this is re NEW - The Ontario Liberal party: Ontario's Conservative Party is mailing fake invoices to households, with a demand Another apparent donation request disguised as an invoice. this time for $1000. I fear thi So Doug Ford, Conservatives are apparently sending out donation requests disguises as invoice for $300 now.	Doug Ford fake invoices
2021-08-28	CON	-0.50	ag Erin O' Toole warns supporters that they're not welcome on the Conservative campaign if they engage in harassment and intimidation of other party leaders. he strongly condemns the actions of demonstrators who forced the cancellation of the liberal campaign Conservative party members vote down resolution to officially recognize climate change - national make Canada Great Again red hats causing an suspension of campaign event.	O'Toole warns supporters not to harass other politicians
2021-09-16	CON	-0.52	cdnpoli Conservative leader Erin O' Toole asked repeatedly about Jason Kenney's handling of the pandemic in Alberta refuses to add Hi do you still believe the Alberta government has had the best response, one worthy of praise and emul.	O'Toole asked about Jason Kenney's handling of the pandemic in Alberta
2021-09-17	CON	-0.74	Erin O' Toole stonewalls questions on Alberta/ Ken So why was CPC candidate Troy Myers name dropped from the ballot? no other party's supporters would do such a thing, especially not Liberals. it seems very strange that they won't talk to the media, go on r.	O'Toole dodges questions on past support of Kenney's handling of COVID in Alberta
2021-09-18	CON	-1.30	85% of federal Conservative candidates won't disclose COVID-19 vaccination status elxn44 Only 15% of conservative candidates say they're fully vaccinated, according to the Globe. Erin O' Toole wants 90% of C Here's candidate saying the quiet part loud. Don'T believe in mandatory vaccinations.	Conservative candidates won't disclose vaccination status

Table 3. The individual Twitter conversations with the largest impact on public opinion. Δ is the change in the % of popular support for the party. The T5 summary is generated by inputting the text of all the tweets concatenated together, preceded by the "summarize:" prefix (see [20] for details on how the T5 language model creates summaries). The manual summary was written after consulting the T5 summary and looking at the content of the tweets.

COVID outbreak in the province of Alberta blamed on the poor provincial leadership of the Conservative hurts the federal CON leader as he had defended this Premier's handling of COVID in the past (and refused to go back on that support). Then as the election neared, the COVID vaccination status of CON members was under scrutiny, with the CON leader refusing to disclose who in his caucus was vaccinated.

Such a shift in public opinion is not seen in the average of the polling data (although there is a similar trend in the polling companies who use IVR rolling polls, albeit not as sharp). It is not possible to see if this is again the Twitter data anticipating the polls, as after the election on September 20th, polling companies cease polling. However, the actual election outcome is available for comparison, and the CONs end up winning 33.74% of the popular vote to the LIB's 32.62%. This makes it unlikely that such a large change in public opinion did occur prior to the election. The final forecast of popular vote from the "converstations" model is 34.1% LIB to 28.0% CON. While this still predicted the correct overall winner in seats (due to the distribution of the vote, the LIBs still one the plurality of seats, and hence won the election, see [3] for details of the seat model), it is the case that forecasting the vote based on the Twitter data a few days earlier would have yielded a more accurate popular vote. It may be the case that the final few days were picking up opinion change in those more political engaged which had not the time to percolate into the general public. The way the Twitter data anticipated the other large changes in public opinion seen in polling data gives support to this hypothesis. It is also possible that the data was biased by partisan attacks playing out on social media: perhaps in the final days before an election, social media is weaponized by partisans in a way that makes Twitter data during this period behave in a different way.

5. Conclusion and Future Work

A method of tracking changes in political opinion using unsupervised learning was presented. The time-series data matches well with polling data, capturing the major trends in the campaign especially well for the major parties.

The method is tested on the 2021 Canadian Federal Election. Twitter data anticipates the major public opinion trends observed in public opinion polls, and is able to explain why those changes happened. However, the divergence in public opinion seen in the Twitter conversation model between LIB and CON in the final days of the campaign is not observed by all public polls and in the final election result. Further work applying this method to other elections is needed to determine whether this anomaly was due to capturing true shifts in public opinion among the politically engaged (and thus perhaps anticipation true opinion change which simply occurred after the election), or whether this is a shift in the nature of Twitter data in the final days of an election (e.g. due to partisans using social media as an election tool right before the final vote).

As the data is based on extracted "conversions", this method allows an understanding of the reasons for the changes in fortunes of the parties. In the case 2021 Canadian federal election, not only is the large decline in LIB support early in the campaign observed in the conversation model before its seen in polling data, but the conversations driving the change make it clear that the *reason* for the decline is the triggering of the election itself. The ability to understand the causes of changes in public opinion as they are happening makes this kind of analysis a valuable tool for those wishing to shape public opinion, and those wishing to understand it.

References

[1] A. Khan, H. Zhang, N. Boudjellal, A. Ahmad, J. Shang, L. Dai, and B. Hayat. "Election Prediction on Twitter: A Systematic Mapping Study". In: *Complexity* 2021 (2021).

[2] P. Chauhan, N. Sharma, and G. Sikka. "The emergence of social media data and sentiment analysis in election prediction". In: *Journal of Ambient Intelligence and Humanized Computing* 12.2 (2021), pp. 2601–2627.

[3] S. M. Davidson and K. White. "Forecasting Seat Counts in the 2019 Canadian Federal Election Using Twitter". In: *Advances in Artificial Intelligence*. Ed. by C. Goutte and X. Zhu. Cham: Springer International Publishing, 2020, pp. 151–162. ISBN: 978-3-030-47358-7.

[4] S. Fano and D. Slanzi. "Using Twitter Data to Monitor Political Campaigns and Predict Election Results". In: *Trends in Cyber-Physical Multi-Agent Systems. The PAAMS Collection - 15th International Conference, PAAMS 2017*. Ed. by F. De la Prieta, Z. Vale, L. Antunes, T. Pinto, A. T. Campbell, V. Julián, A. J. Neves, and M. N. Moreno. Cham: Springer International Publishing, 2018, pp. 191–197. ISBN: 978-3-319-61578-3.

[5] M. Song, M. C. Kim, and Y. K. Jeong. "Analyzing the Political Landscape of 2012 Korean Presidential Election in Twitter". In: *IEEE Intelligent Systems* 29.2 (2014), pp. 18–26. DOI: 10.1109/MIS.2014.20.

[6] J. Devlin, M.-W. Chang, K. Lee, and K. Toutanova. "Bert: Pre-training of deep bidirectional transformers for language understanding". In: *arXiv preprint arXiv:1810.04805* (2018).

[7] S. Ghosh, P. Singhania, S. Singh, K. Rudra, and S. Ghosh. "Stance Detection in Web and Social Media: A Comparative Study". In: *Experimental IR Meets Multilinguality, Multimodality, and Interaction*. Ed. by F. Crestani, M. Braschler, J. Savoy, A. Rauber, H. Müller, D. E. Losada, G. Heinatz Bürki, L. Cappellato, and N. Ferro. Cham: Springer International Publishing, 2019, pp. 75–87. ISBN: 978-3-030-28577-7.

[8] G. Li. *Sampling Graphical Networks via Conditional Independence Coupling of Markov Chains*. 2016. DOI: 10.1007/978-3-319-34111-8_36. URL: http://link.springer.com/10.1007/978-3-319-34111-8_36.

[9] K. White. *Forecasting Canadian Elections Using Twitter*. Ed. by R. Khoury and C. Drummond. 2016. DOI: 10.1007/978-3-319-34111-8_24. URL: http://link.springer.com/10.1007/978-3-319-34111-8http://link.springer.com/10.1007/978-3-319-34111-8_24.

[10] F. Morstatter, J. Pfeffer, H. Liu, and K. Carley. "Is the sample good enough? Comparing data from twitter's streaming API with Twitter's firehose". English (US). In: *Proceedings of the 7th International Conference on Weblogs and Social Media, ICWSM 2013*. 7th International AAAI Conference on Weblogs and Social Media, ICWSM 2013 ; Conference date: 08-07-2013 Through 11-07-2013. AAAI press, 2013, pp. 400–408.

[11] F. Wang and E. Mulrow. "Analyzing Open-Ended Survey Questions Using Unsupervised Learning Methods". In: *JSM Proceedings*. 2014.

[12] A.-S. Pietsch and S. Lessmann. "Topic modeling for analyzing open-ended survey responses". In: *Journal of Business Analytics* 1.2 (2018), pp. 93–116.

[13] L. Wang, C. Gao, J. Wei, W. Ma, R. Liu, and S. Vosoughi. "An empirical survey of unsupervised text representation methods on twitter data". In: *arXiv preprint arXiv:2012.03468* (2020).

[14] J. A. Hartigan and M. A. Wong. "Algorithm AS 136: A k-means clustering algorithm". In: *Journal of the royal statistical society. series c (applied statistics)* 28.1 (1979), pp. 100–108.

[15] T. Zhang, R. Ramakrishnan, and M. Livny. "BIRCH: an efficient data clustering method for very large databases". In: *SIGMOD '96*. 1996.

[16] W. H. Day and H. Edelsbrunner. "Efficient algorithms for agglomerative hierarchical clustering methods". In: *Journal of classification* 1.1 (1984), pp. 7–24.

[17] M. Ester, H.-P. Kriegel, J. Sander, and X. Xu. "A density-based algorithm for discovering clusters in large spatial databases with noise". In: AAAI Press, 1996, pp. 226–231.

[18] E. B. Fowlkes and C. L. Mallows. "A Method for Comparing Two Hierarchical Clusterings". In: *Journal of the American Statistical Association* 78 (1983), pp. 553–569.

[19] P. J. Rousseeuw. "Silhouettes: a graphical aid to the interpretation and validation of cluster analysis". In: *Journal of computational and applied mathematics* 20 (1987), pp. 53–65.

[20] C. Raffel, N. Shazeer, A. Roberts, K. Lee, S. Narang, M. Matena, Y. Zhou, W. Li, and P. J. Liu. "Exploring the Limits of Transfer Learning with a Unified Text-to-Text Transformer". In: *CoRR* abs/1910.10683 (2019). arXiv: 1910.10683. URL: http://arxiv.org/abs/1910.10683.