

Health and Politics:

Analyzing the Government of Alberta's COVID-19 Communications

Draft Multi-Part Paper for our Panel for CSDH-SCHN 2021

Rockwell, Geoffrey; Baumgart, Daniel; Goodman, Karen; Goebel, Randy; Barnard, Sara; Chartier, Ryan; Farruque, Nawshad; Ingram, Katrina; Tchoh, Bennett; O'Neil, Erin

1.0 Introduction

Geoffrey Rockwell

On March 17, 2020, the province of Alberta declared a state of emergency in the wake of the coronavirus pandemic that was sweeping the world, sending people into lockdown and shuttering economic activity. Alberta was particularly hard hit by this crisis. The Alberta economy, which relies heavily on the energy sector, was reeling from a global drop in energy demand and oversupply of oil that sent prices into negative territory. Politically, Premier Jason Kenney and the UCP government rose to power on the promise of creating jobs and fixing Alberta's economy. COVID-19 upended these plans and has led to a growing deficit, tens of thousands of additional lost jobs, a downgrading of the province's credit rating, and a net out-migration as people seek opportunities elsewhere. It is a grim economic picture and with the coronavirus pandemic far from over, stark tradeoffs persist in the potentially conflicting goals of safeguarding public health and economic well being.

Questions to be Addressed

Given the tension between ensuring public health and mitigating economic damage, what messages has the Government of Alberta delivered to the public and in what ways have these messages shaped the public discourse or been shaped by it? Has the COVID-19 crisis provided opportunities for the government to mobilize messages in pursuit of other political goals, such as moving towards privatization of healthcare options?

From April 2020 through to today, we have been tracking official communications of Premier Jason Kenney and Chief Medical Officer Deena Hinshaw along with Alberta Twitter discourse about the pandemic and a selection of local media coverage. This panel gathers a number of short papers that look at the background of the GuARD-AI (Guarding At-Risk Demographics with AI) project, the types of data gathered, the analytical techniques applied, and preliminary results. Our analysis of this data describes how the government attempted to balance public health and politics during these extraordinary times. Our panel is made up of short papers wrapped with an introduction and conclusion that will show the arc of a project that is ongoing, in the sense that we are still collecting data and will until the emergency is over.

The successful management of a public health emergency such as COVID-19 requires a risk communication approach that guides the official public health discourse so it can insert effective disease prevention messages into the public discourse that influences public behaviour. (Health Canada) The way the pandemic and its effects are discussed in public discourse tell us about how COVID-19 is affecting everything from our jobs to our mental health. How can we document and study the communications environment around the pandemic - an environment which includes different communication channels from government press briefings, news reports, social media, and regularly updated data visualizations? The GuARD-AI project is a partnership that gathers streams of public discourse and health data to develop new ways of guarding at-risk communities.

Why public discourse? We've seen how important public communication was in previous pandemics, such as H1N1 or SARS (Quammen 2012). We are now seeing how it plays out in the current pandemic (Duhigg 2020, Kim, 2020). We see that politicians and public health officials often have different agendas and communicate different priorities. When politicians and public health officials deliver misaligned messages, it creates confusion that may impact public behaviour in undesirable ways. (Kottsova, 2020) For example, the US White House Coronavirus Task Force briefings that often featured former President Trump regularly conflicted with what senior US health officials were communicating. We also see that different demographic groups pay attention to different media channels. For example, youth may pay more attention to social media or Ryan Reynolds than government press briefings or mainstream media (Espinoza 2020). Finally, we see that data visualizations have become a common way that complex health statistics are presented and consumed by the public. (Schneiderman, 2020)

The objectives of the GuARD-AI project are to use digital humanities methods like text scraping, analysis and visualization techniques to compare political and public health discourse in order to better understand how the coronavirus has affected us. Specifically we have been:

- Gathering data with a CIFAR grant that is scraping public discourse about COVID-19 from Alberta. This has allowed us to build a digital archive following best practices in the digital humanities that will be shared under a Creative Commons license (CC BY-NC-SA 2.5 CA).
- Analyzing the discourse using humanities text analysis techniques combined with emerging AI (Artificial Intelligence) techniques. This project brings together specialists in humanities analytics (Rockwell) and AI analytics (Goebel) along with health data specialists (Baumgart and Goodman) to develop innovative medical humanities analysis and visualization methods.

2. Covid-19 and Public Health

Daniel Baumgart and Karen Goodman

Pandemics like the COVID-2 pandemic threaten to overwhelm the capacity of health systems to meet the healthcare needs of populations. Logistic challenges to rapid development and

deployment of accurate diagnostic tests constrain the availability of testing on the scale required to detect and isolate cases to prevent contagion. The adequacy of public health resources determines whether contact tracing efforts occur on the scale required to prevent further spread. The healthcare system's facilities and workforce require dynamic adaptation so resources can be deployed to manage rapidly evolving care demands across populations, a particular challenge for populations that are widely dispersed geographically. Physical distancing measures that reduce person-to-person transmission force practitioners to rapidly adopt digital medicine and learn how to use novel virtual healthcare delivery models effectively. Ideally, all these challenges should be addressed in real time with evidence from data-driven tools. AI has the potential to rapidly transform large amounts of data into predictive models that can assist rapid decision making. In particular, the application of AI techniques to text mining has the potential to analyze messy public discourse data to give health care providers a catalog of health-related messages being disseminated to the public over time, along with a distillation of notions about the pandemic circulating in society. With such evidence derived from data, health care leaders can better identify which health services will be most beneficial; as well, they can develop guidelines for effective health care messaging practitioners can use.

From a public health perspective, this project aims to develop best practices for health analytics in situations where time is of the essence and action-based decisions are supported by extracting value from highly dynamic and time-sensitive data. We have used funding from CIFAR to:

1. identify existing data that can be ethically sourced for COVID-2 research;
2. experiment with candidate AI techniques that work with minimal training data.

3. Gathering Data

Sara Barnard

The compilation of data for the project began in March 2020 and consists of three separate corpora. These are:

- A collection of news articles about COVID-19 in Alberta,
- Public briefings by Alberta Premier Jason Kenney and the provincial Public Health Officer, Dr. Deena Hinshaw, and
- A collection of tweets with local Alberta pandemic-related hashtags (for example: #ABhealth, #ABdoctors, and #COVID19AB)

The first corpus, the News Articles, are being manually gathered as part of the process of building a narrative timeline of COVID in Alberta. This is being done in a few steps. Firstly, alerts have been set up on google so that we were fed key stories each week. This was done using the keywords "abtractogether," "Cargill alberta," "covid alberta," "first nations alberta covid," "long term care alberta," and "virtual care alberta." The result is about 25-30 alerts each week, of which each averaged 1-6 stories. The majority of the stories came from sources such as the Edmonton Journal, Calgary Herald, CBC, Global TV or smaller city newspapers.

When a story was saved, it is copied and pasted into Notepad as a text document. Each story is titled with a specific format for easy organization and access. This format is as YYYY-MM-DD followed by ARTICLETITLE.txt. The articles are initially saved locally, followed by a batch upload once per week to the shared drive. After the articles are saved, the metadata is also recorded in a spreadsheet. This includes the formatted title of the article along with the URL where the article was originally found.

After the articles are uploaded, the data is manually cleaned. During this process extraneous navigation content is removed, such as article sidebars, ads, and special characters which do not translate while gathering the information. However, information such as the author's name and date of publication are kept. Similarly, the titles are double checked to ensure that they conformed to our agreed format. Finally, the articles are then gathered into weekly datasets so that they could be easily accessed to create a timeline of events. We have over 1000 such web pages.

The second corpus is composed of the public Press Briefings of Premier Jason Kenney and Chief Medical Officer of Alberta Dr. Deena Hinshaw regarding the threat of and responses to COVID-19 in Alberta. The first briefing dates to March 6 2020, and we are still gathering them as new COVID press briefings are held. There are roughly 100 briefings from Kenney and 190 briefings from Hinshaw in the corpus.

The text of Dr. Hinshaw's speeches are gathered from the publicly available transcripts on the Government of Alberta's website. On the other hand, Premier Kenney's speeches are downloaded from SoundCloud and then we use speech recognition (Otter.ai) to translate them into plain text files. Finally, the transcript data is manually checked and cleaned to ensure accuracy. The transcripts from both speakers exclude question and answer periods from the briefings so that only the presented speeches are included. A Spyral (Voyant) notebook of the briefings from March to August 2020 with embedded analytics is available at <https://voyant-tools.org/spyral/CovidDiscourse2/>.

The Twitter corpus was gathered by scraping three hashtags: #abhealth, #albertadoctors and #covid19ab. The first tweets were scraped on March 17 2020, March 18 2020, and April 21 2020, respectively. While the data of this corpus is not cleaned, retweets are excluded from the corpus as analyzed. The text of the tweets is divided into weekly files to create a corpus that can be also analyzed using text analysis tools like Voyant..

4. Topic Modelling of the Data

Ryan Chartier

In analyzing the data we collected we first began by attempting to identify the basic structure of the collected Twitter data. To do this, as mentioned above, we divided the raw text of all the collected tweets, minus the retweets, into weekly files. Each weekly file can then be thought of as a single document that we can analyze using the Mallet LDA (latent dirichlet allocation) topic modelling engine (McCallum 2002). LDA topic modelling is a statistical algorithm that attempts to extract the underlying topics from a corpus of documents by clustering every word in that corpus into one of several topics (Blei 2001; Blei). The algorithm sorts words that are more commonly used together into the same topic.

Topic modelling does two important things for us. Since each document in our Twitter dataset represents a single snapshot of time, the proportion of words in each document assigned to each topic can give us an idea of how the prevalence of certain words has changed as the conversation has progressed. Secondly, we can model each document as a vector. These vectors can be directly compared and clustered giving us a rough metric of how quickly the conversation has changed from week to week. Documents that do not cluster very closely together represent periods in the conversation where topics are changing while documents that do cluster represent periods where the topics are more sustained.

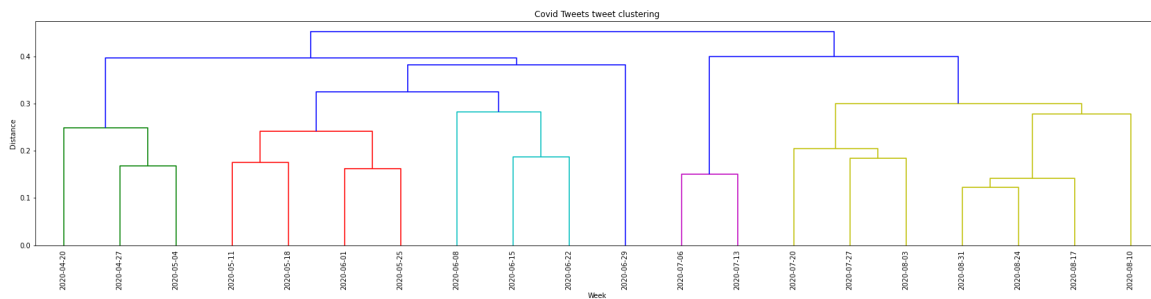


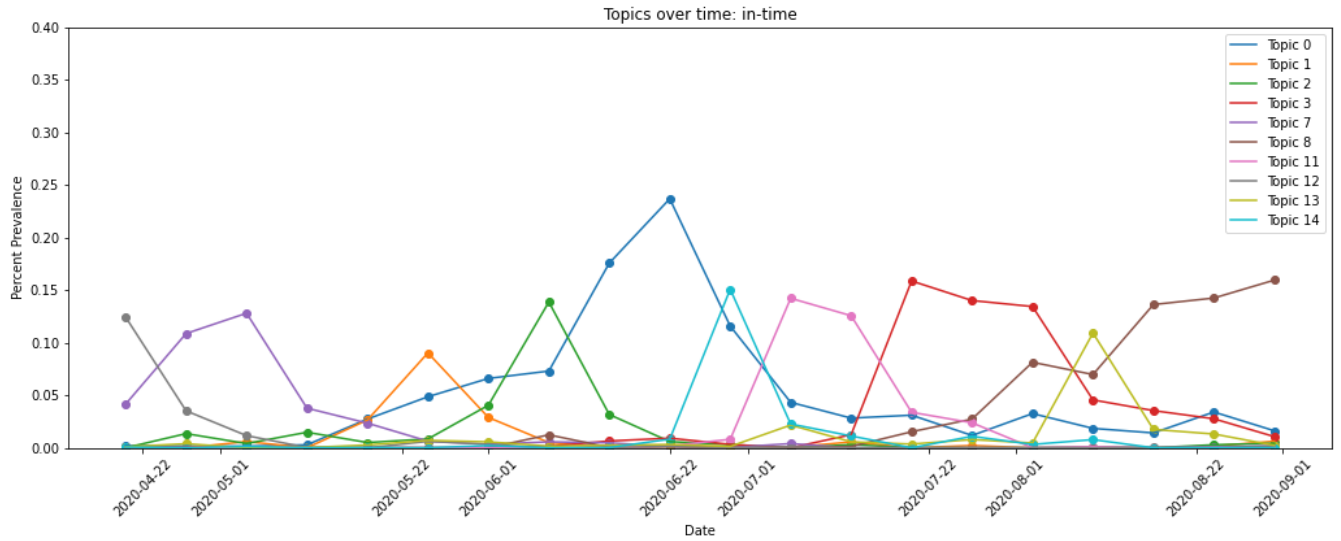
Figure 1 – Clustering of all weeks in the data set.

The first step was clustering. The main output of a Mallet topic model is a vector embedding for each document in probability space. Each number in this vector represents the proportion of all the words in a document that were assigned to each topic. Equivalently we can think of each vector as being a probability vector; the probability that a random word taken from each document will be assigned to a specific topic. Probability vectors are well understood mathematical objects which can be analyzed using basic information theory. We can calculate the relative distance between each document using the 'Jensen-Shannon Divergence' and then cluster the results using any hierarchical clustering algorithm. In this case we used the `scipy.hierarchy.linkage` function which joins nodes together sequentially starting with the two closest nodes and working until only one node remains.

One important result we were looking for in this process was that individual weeks would cluster closer to adjacent weeks. It is our hypothesis that conversations change gradually with topics spanning several weeks. The number of topics that Mallet generates is a number that we must set as a hyperparameter to the algorithm. If the number is set too high the conversation would change too much between each document and we risk entire documents being assigned their own topics. If set too low the topics would fail to reveal anything useful about the data. This issue actually came up with us during an earlier draft of this methodology. We originally used monthly datasets which resulted in uninteresting clustering because Mallet effectively assigned every month to its own cluster.

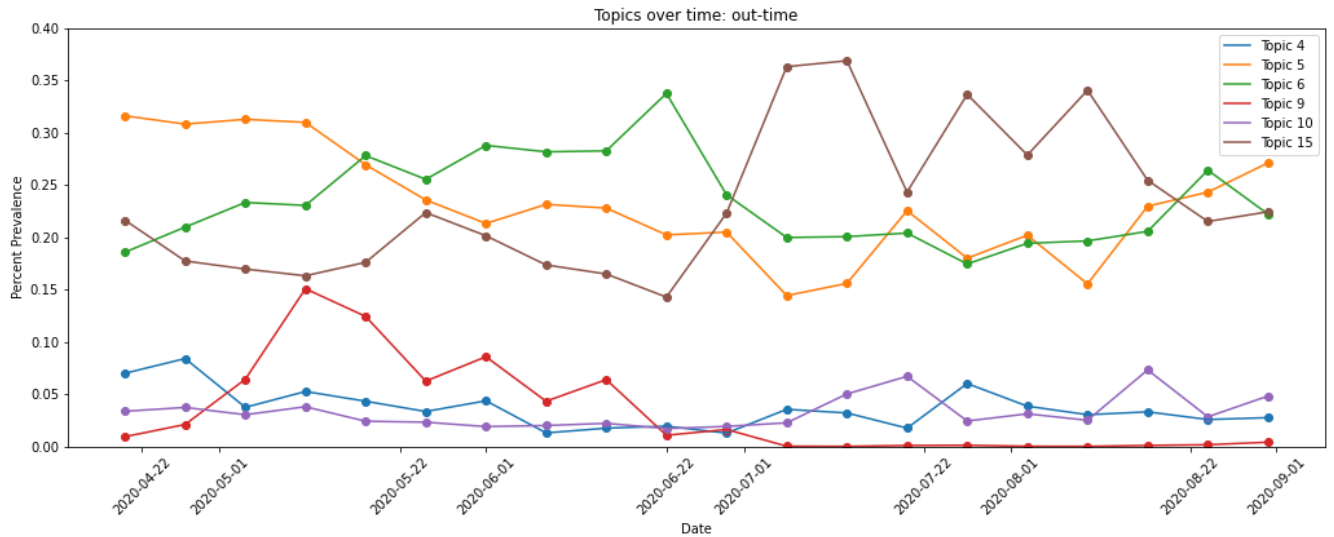
Figure 1 graphically represents this process with early joins being represented lower on the graph and the final join at the top of the graph. Three large clusters are apparent in this

approach. The first three months are the strongest cluster and unfortunately represent a systematic problem with our data. A number of extra hashtags were added to our search in the fourth month. This would appear in our analysis as a dramatic shift in the conversation as we are physically scraping conversations we were not before. The other two clusters, with 2020-06-29 as an outlier, are actual major phases identified by this method.



	12	7	1	2	0	14	11	3	13	8
0	rural	cargill	fixseniorscarenow	stage	yeg	ahs	ama	abed	private	abed
1	cdnpoll	workers	cdnpoll	relaunch	active	rachelnotley	doctors	masks	healthcare	school
2	workers	total	seniors	june	edmonton	plan	patients	mask	hospital	active
3	plant	plant	cdnhealth	health	mask	resign	masks	school	system	schools
4	april	recovered	abpoll	masks	hospital	leadership	bill	schools	facility	cmoh alberta
5	shandro	plan	kenney	phase	responsibilities	executives	ahs	mandatory	rdp	yeg
6	summer	app	jason	friday	duties	canada	healthcare	kids	patients	abpoll
7	cargill	relaunch	federal	services	services	ceo	rdp	plan	https	kids
8	communities	confirmed	services	les	family	rhl	system	adrianalagrange	tyler	teachers
9	cancelled	province	rural	cas	recovered	day	physicians	wear	surgical	safeseptemberab

Figure 2 – Time bound topics.



0	5	6	9	15	4	10
1	covid	covid	active	ableg	abdocs	virus
2	ableg	alberta	calgary	abpoli	patients	numbers
3	calgary	cases	relaunch	health	people	distancing
4	yeg	yyc	businesses	alberta	kenney	months
5	https	https	yeg	care	kenney	icu
6	hinshaw	canada	recovered	yeg	services	cases
7	update	people	brooks	shandro	residents	start
8	coronavirus	zone	care	ucp	great	it's
9	pandemic	deaths	deena	public	centre	political
9	cmoh_alberta	today	total	doctors	albertacares	public

Figure 3 – Global Topics

The topics themselves offer us a different angle into how the conversation changes over time. In this data set topics could be categorized into one of two groups. Some topics have distinct time periods within which they began and then ended; such as topic seven (Fig. 2) which spiked around when the outbreak happened at the Cargill meat packing plant and then mostly disappeared. Others stuck around longer and represent the conversation as a whole; such as topic ten (Fig. 3) which seems to represent the persistent conversation around case numbers.

Along with each topic we also get a list of the topics keys. These are words that most strongly correlate with a specific topic. The keys of a topic that spikes at a certain time period tell us a little about what words are most important in the conversation at that point. Some topics, like topic seven, are easy to identify because the words can be matched to real events that were happening at the time; words such as “cargill”, “workers” and “plant” make it easy to assess the topic as referring to the cargill meat packing plant. However, others, like topics five and six, share a lot of keys implying that the separation between them might not be that significant.

The relationship between keys and events can be further explored in a final tool we created. We wanted to superimpose graphs of more than one corpus and overlap known events onto our data in order to explore how the usage of important words correlates with actual events and how different discourses correlate. You can see in Fig. 4 the graphs for school-related words¹ for

¹ "school", "schools", "teacher", "teachers", "child", "children", "parent", "parents", "kid", "kids"

Kenney, Hinshaw and Twitter superimposed with two key dates, when the schools were closed in March 2020 and when they were reopened.

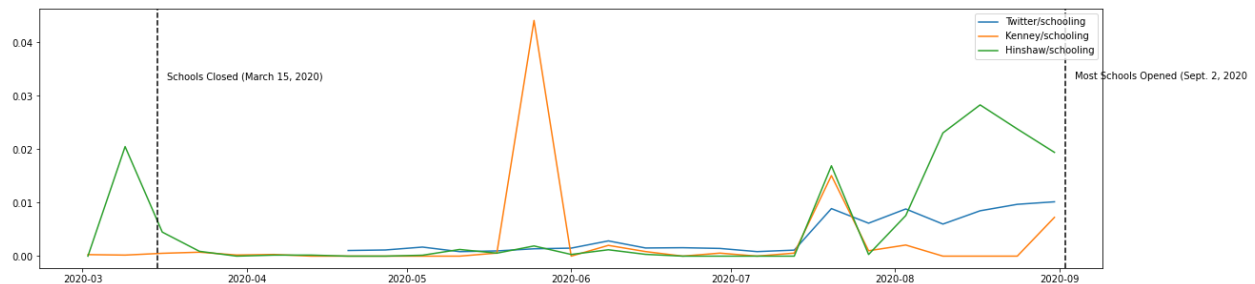


Figure 4: Superimposed Graphs and Events

5.0 Sentiment Analysis of the Data

For analyzing sentiment, we use the LIWC-2015 software. LIWC software makes use of a widely used Psycholinguistic Lexicon, named Linguistic Inquiry and Word Count (LIWC) (Pennebraker et al. 2015), developed over years with the help of a group of psycholinguists. This lexicon contains circa 100 psycholinguistic categories such as, personal pronouns, positive emotions, negative emotions, friends, health, work, past focus, and anxiety, to name a few. Given a particular blob of text, this software helps us find the percentage of words (or intensity scores) from each category present in that text blob out of all the words present in it. In our case, we fed LIWC-2015 with our weekified briefings for Hinshaw or Kenney between May - August, 2021, and we got intensity scores for three important categories of LIWC dimensions, such as: (1) **Affect** (containing Positive Emotion, Negative Emotion, Anxiety and Negate) (2) **Grammatical** (Personal Pronouns, Adjective, I-Pronoun and Verb) and (3) **Political** (Death, Health, Work and Family). We plotted the charts for Hinshaw and Kenney within the said period, over-laid on one another to help us compare and contrast between the discourse change and trends over these dimensions. Moreover, we employ an algorithm, called **Max-Dev**, through which we find the LIWC dimensions which have greatest fluctuations over the weeks and sort them based on highest to lowest of their standard deviation scores to get an essence of the sensitive LIWC dimensions for Hinshaw and Kenney. Following are the analyses based on (1) temporal fluctuations over three categories of LIWC dimensions and (2) Sensitivity or fluctuation analysis of LIWC dimensions under each of the categories.

Temporal fluctuations over three categories of LIWC dimensions: In this analysis we plot the weekly briefings of Hinshaw and Kenney within the period of March - August, 2021. We provide individual analysis of each of the LIWC dimensions below as well as over-all picture for a particular category.

- **Affect:** In this category we look into temporal fluctuations of LIWC scores or intensity for Positive Emotion, Negative Emotion, Anxiety and Negate. The **Positive Emotion** dimension consists of words related to positive emotions, such as, “fair, excite, favor, kind, etc”. We observe increasing trends for both Hinshaw and Kenney over weeks.

Negative Emotion dimension consists of words related to negative emotions, such as, “abandon, anguish, annoy, etc”. We observe a decreasing trend for both Hinshaw and Kenney for this category. For the **Negate** dimension (consisting of words, “ain’t, isn’t, aren’t etc”), we see an increasing trend for Hinshaw as opposed to Kenney with a decreasing trend. For **Anxiety** we observe, both Hinshaw and Kenney have an increasing trend. Summarily, except Negate dimension, Hinshaw and Kenney agreed upon all the other dimensions (increasing Anxiety and Positive Emotions with decreasing Negative Emotions) based on trend analysis. For Negate we see a contradiction between Hinshaw and Kenney, with Hinshaw’s briefings containing more negations over time compared to Kenney. See Figure 5 for a sample.

Anxiety Fluctuations over Weeks of March - August

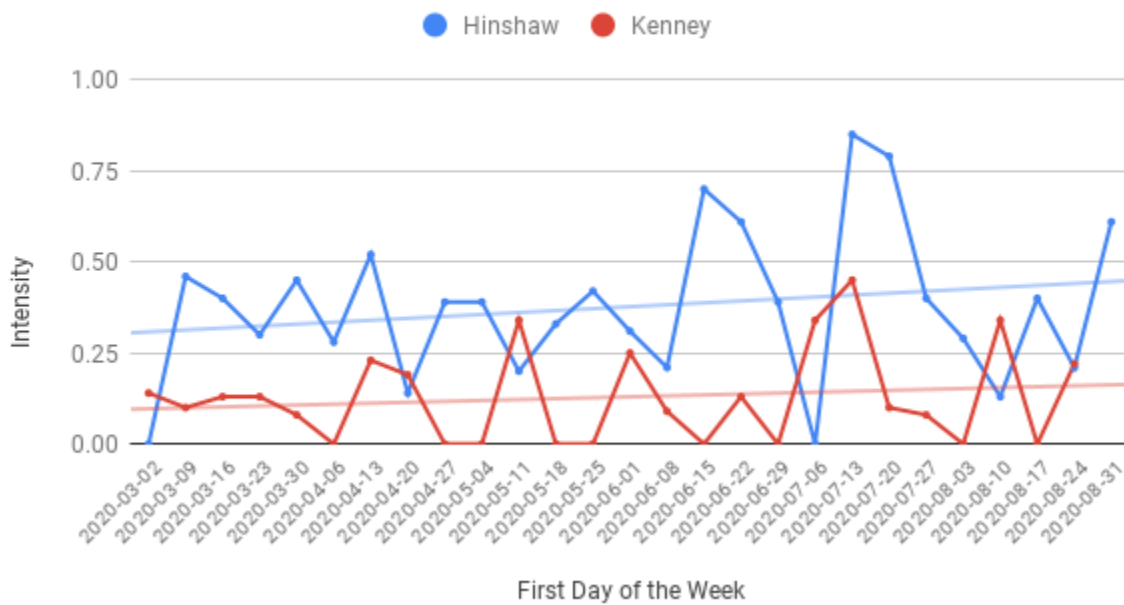


Figure 5: Anxiety Fluctuations

- Grammatical:** In this category we find out the fluctuations for different parts-of-speech categories, such as, **Personal Pronouns** (such as, “he, he’s, hers, etc”), **Adjectives** (such as, “awful, terrible, favorite, etc”), **Verb** (such as, “needs, notice, hope, etc”) and **I-pronouns** (such as, “I, I’d, I’m, etc”). For verbs, personal pronouns and I-pronouns we see, Hinshaw has an increasing trend but Kenney has a decreasing trend. However, for Adjectives, we see both Hinshaw and Kenney have an increasing trend. See Figure 6 for a sample.

Personal Pronouns Fluctuations over Weeks of March - August

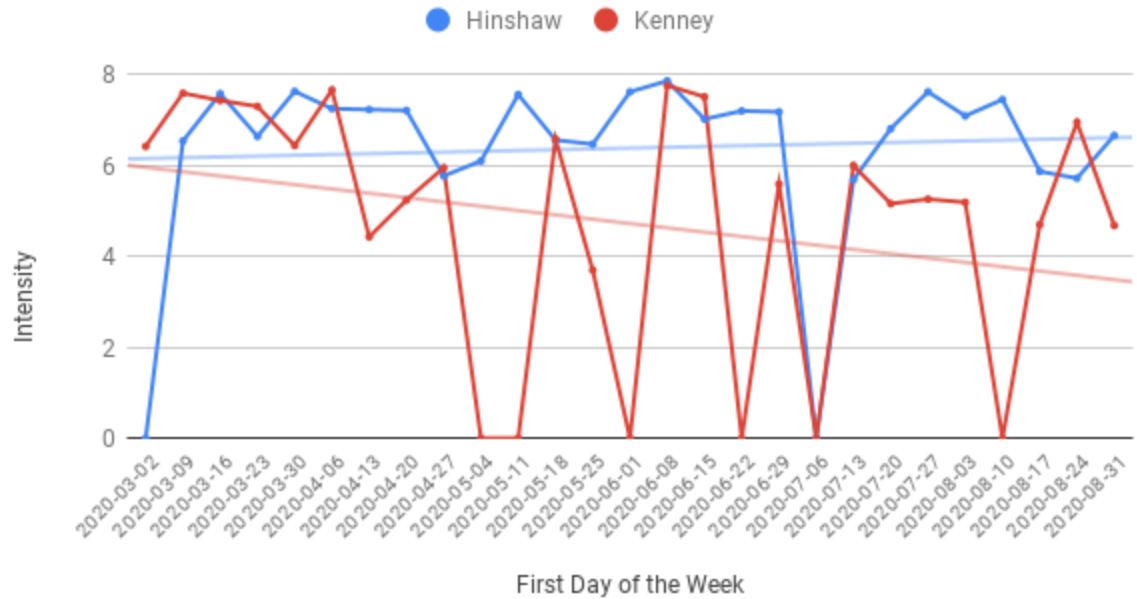


Figure 6: Personal Pronouns Fluctuations

- Political:** For political category analysis we focused on the following LIWC dimensions, such as, **Death** (contains words like, “death, decease, demise, etc”), **Health** (contains words such as, “nurse, pill, pain, physician, etc”), **Family** (contains, “family, brother, sister, sibling, etc”) and **Work** (contains, “work, income, highschool, academic, etc”). For Death dimension, we observe a decreasing trend albeit very slight, over the weeks for both Hinshaw and Kenney. However, we observe opposite fluctuations in the months of April, May and June for the same. For the Health dimension, we observe a slight increasing trend for both Hinshaw and Kenney and both have almost similar fluctuations except few opposing spikes. For Work, Kenney has a drastic decreasing trend over time as opposed to Hinshaw who has an increasing trend. For Family dimension, Hinshaw has a drastic decreasing trend as opposed to Kenney having a drastic increasing trend. See Figure 7 for a sample.

Work Fluctuations over Weeks of March - August

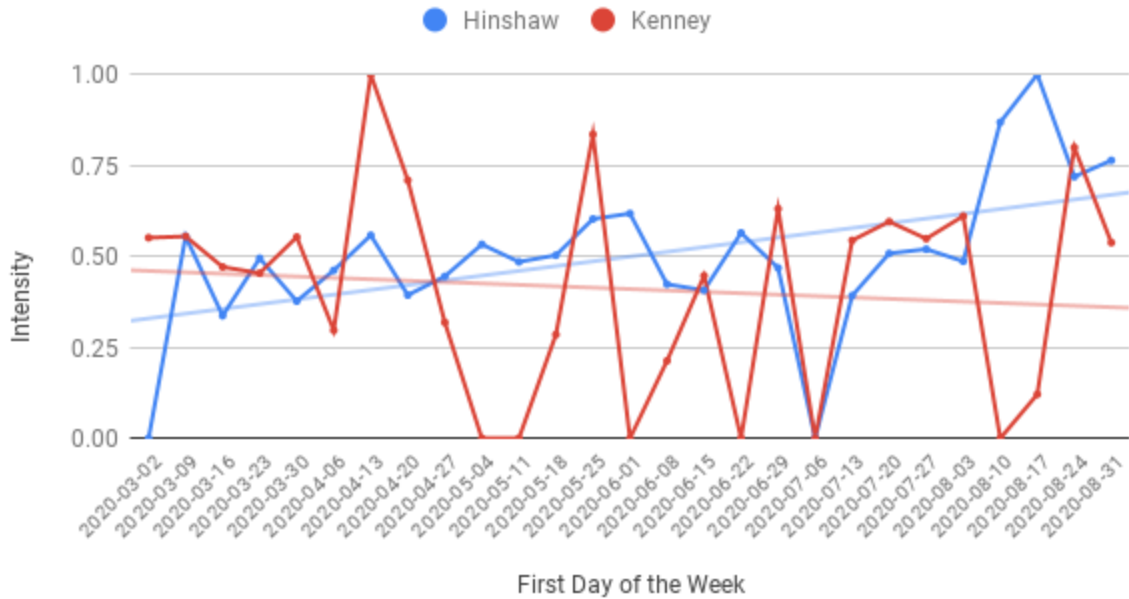
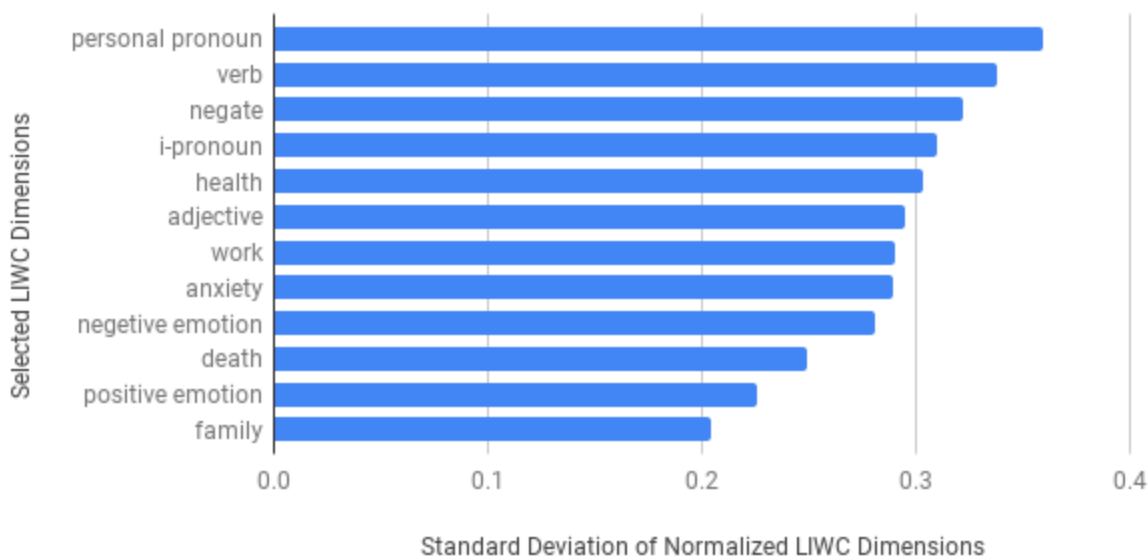


Figure 7: Work Fluctuations

Hinshaw has slightly more intensity scores for the LIWC dimensions may be because, Hinshaw consistently delivered lengthier briefings than Kenney and thus had more probability of having more words under different LIWC categories.

Feature Interestingness Analysis: We calculate the standard deviation of each LIWC dimensions stated above. Before calculating standard deviation, we normalized the dimensions based on min-max normalizer (Min-max Normalization) to limit the dimension values within the range of 0 and 1. Later, we sorted the features in descending order of their standard deviations to capture the features which are most sensitive or fluctuated often over the course of weeks. The following figures depict the important features across all discourses mentioned above for Hinshaw and Kenney.

LIWC Dimensions with Highest Fluctuations over Weeks of March - August for Kenney



LIWC Dimensions with Highest Fluctuations over Weeks of March - August for Hinshaw

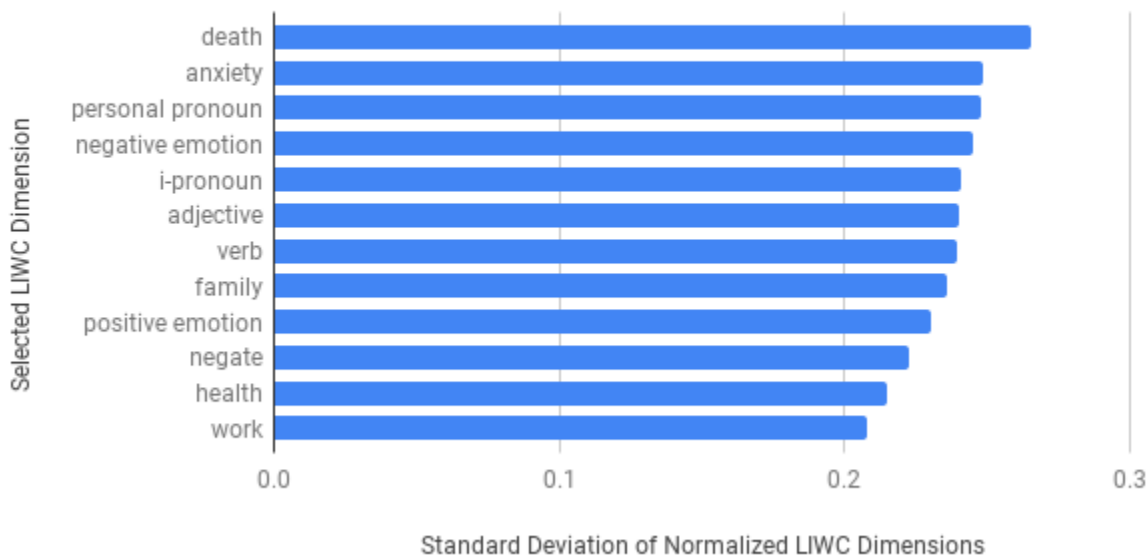


Figure: 8 Max-Dev Analysis for Kenney (above) and Hinshaw Figure: 9 (below)

It is interesting to note that, Kenney has the most deviations in grammatical category (personal pronouns and verbs are top-2s), whereas Hinshaw has most deviations in political categories (death and anxiety are top-2s). See Figures 8 and 9 for sample.

6. Conclusions and Next Steps

Communicating timely, relevant and accurate information about COVID-19 has been difficult to navigate for many countries across all levels of government. In Alberta, the tension between ensuring public health and safety while also attempting to balance economic considerations has been particularly challenging. Polls indicate that the Kenney government missed out on the early “COVID bump” that other politicians in Canada experienced. By early September, Premier Kenney had the second lowest approval rating for a Canadian Premier, marking the lowest level of public support since he took office. (CBC, 2020) The official public health messaging in Alberta, led by Dr. Deena Hinshaw, played out against this political backdrop. The Alberta government seems to have been driven by an ideological approach to managing the pandemic that’s often misaligned with public health messaging. (Bratt and Young, 2020) This was exhibited by fights with doctors, the push to reopen businesses quickly, tensions with the federal government over contact tracing apps, confusion around school reopenings, a reluctance to take swift action to enforce stricter measures during the 2nd wave and a lack of provincial guidance on masks. Our data highlights key inflection points in the public discourse that illustrates these issues. For example in the graph below you can see how in the combined Kenney and Hinshaw briefings we see different issues rise and fall as tracked by patterns like “case*”, “school*”, and “travel*”.

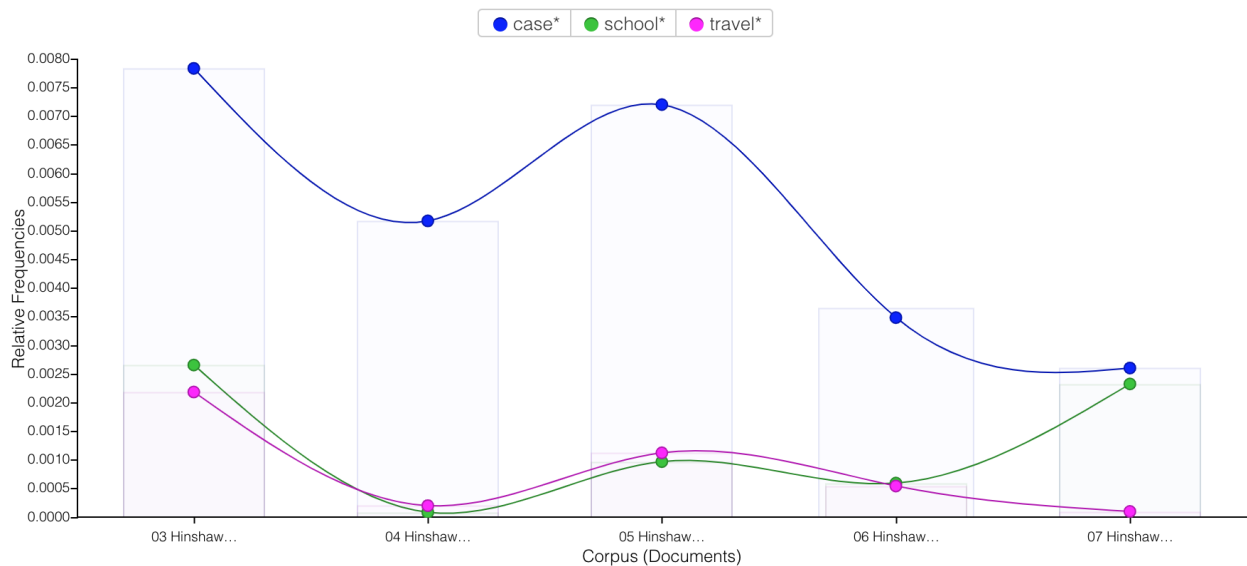


Figure 10: Case, School, and Travel over the months in Hinshaw

We can also see when comparing Hinshaw’s discourse to Kenney’s the difference between them in emphasis. It is not surprising that Kenney is interested in jobs, the economy and energy while Hinshaw is talking about COVID cases.

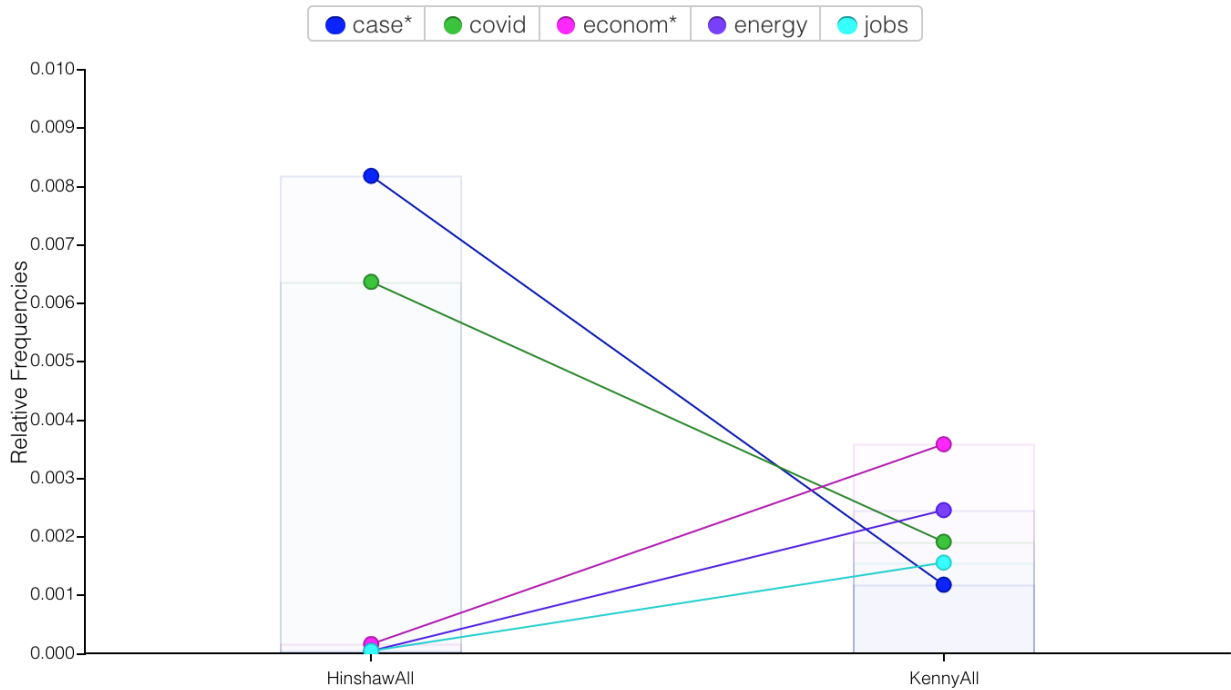


Figure 11: Comparing Hinshaw to Kenney

Some of the major themes discussed in the different corpora include:

Schooling. In Figure 4 above you can see the way schooling was discussed first by Hinshaw before schools were closed and then a build up over the summer before the Sept. 2nd reopening of schools.

Responsibility. One of the words that stands out in Kenney’s briefings is responsibility, especially compared to the Twitter discourse where it isn’t discussed as much.

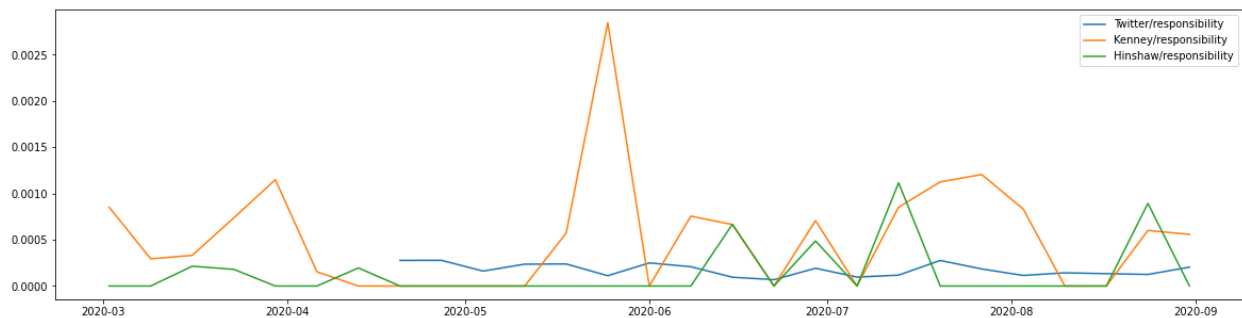


Figure 12: Responsibility(ies) in Kenney, Hinshaw, and Twitter

When we look at collocates we see words like “personal”, “great”, “culture”, and “care.” In Figure 13 you can see a KWIC for Kenney’s use of “responsibility. He is fond of phrases like “personal and civic responsibility”.

Left	Term	Right
i think it is our	responsibility	as leaders to offer a
have a role and a	responsibility	to stop the spread and
on the future and the	responsibility	that our company has to
canada has the highest social	responsibility	or what is now known
to the world the of	responsibility	for others and in as
having continued to show great	responsibility	and care for others through
of the culture of personal	responsibility	and care for others which
of school choice and the	responsibility	of parents to direct their
and choice as well as	responsibility	because we believe that parents
our great civic and personal	responsibility	and our infection and hospitalization
help them continue to personal	responsibility	for their own safety and
commitment to personal and civic	responsibility	is what got us to
relaunch it with the same	responsibility	and resilience there will be
to repeat we've shown and	responsibility	at every level of our
have with great resilience and	responsibility	throughout the pandemic alberta had
part of us taking back	responsibility	for this area of the
to the culture of personal	responsibility	and care for others that
sense of personal and civic	responsibility	that we did so well
we all have a personal	responsibility	as albertans parents friends and
in our province through personal	responsibility	so let me be if
the pandemic albertans have great	responsibility	doing their part to contain
shown great personal and community	responsibility	in following the basic health
treatment at their own greater	responsibility	in the community along the
more flexibility to let their	responsibility	responsibly in local parks let
and thanks to albertans personal	responsibility	and our for the our
culture of personal and civic	responsibility	that and culture of enterprise
your care for your personal	responsibility	for your and your worship
have great personal and civic	responsibility	by following the public health

Figure 13: KWIC of “responsibility” in Kenney

Masks have also been an important issue and we can see words like “mask” and “masks” rising in frequency across all Hinshaw as she calls for mask wearing and Twitter as people discuss it. Note how the discussion drops after Edmonton and Calgary pass mandatory mask laws.

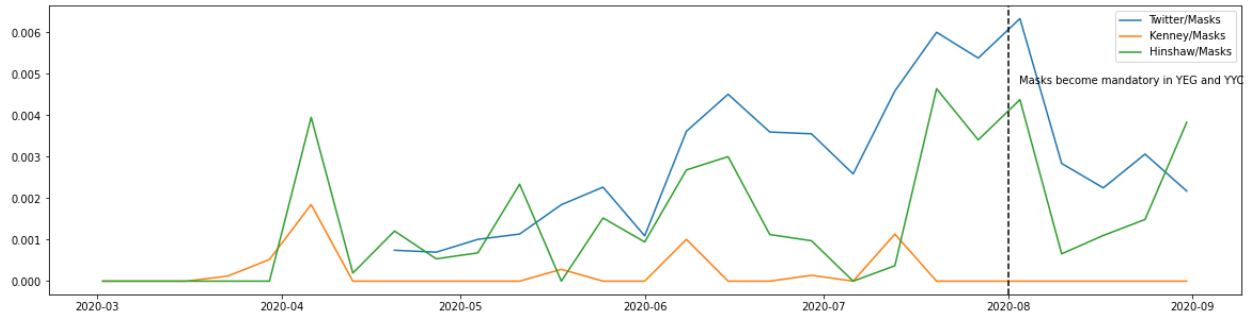


Figure 14: Mask and Masks

Regulations, Laws, Measures. It is interesting to follow the discourse around different measures and regulations.

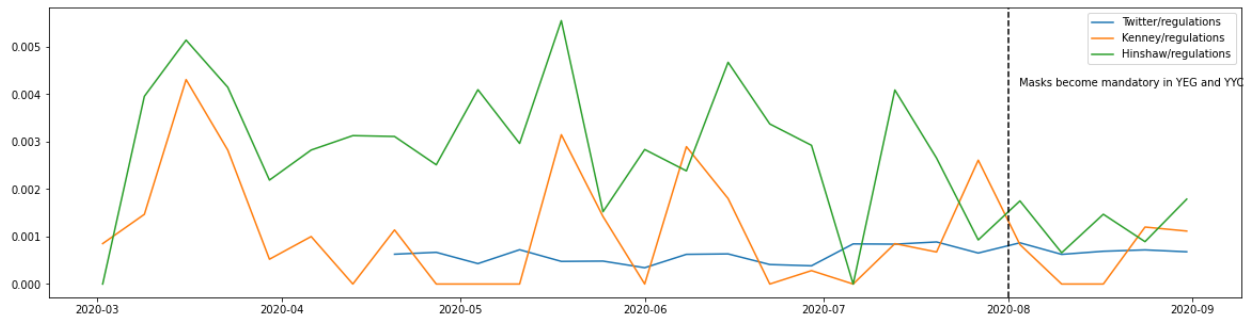


Figure 15: Regulations, Laws, Measures

Cities. We can see different cities getting attention at different times as the caseloads rise and fall with successive waves.

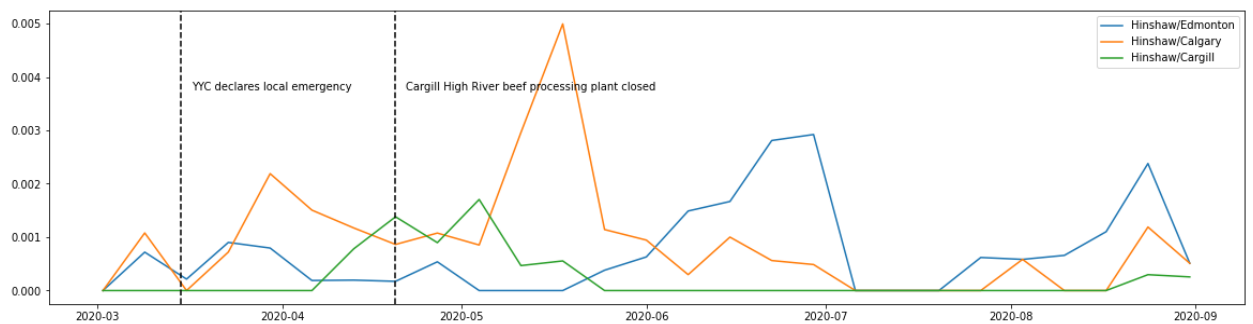


Figure 15: Hinshaw on YEG, YYC, and Cargill

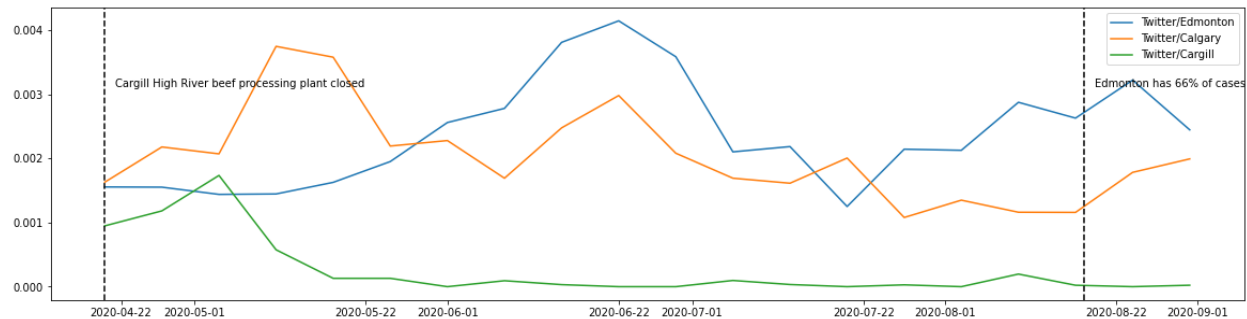


Figure 16: Twitter on YEG, YYC, and Cargill

Next Steps

The next steps involve expanding the window studied as we continue to gather data and will continue until the pandemic is declared over. This first study focused on the March to August 2020 window and led to creating custom tools like the tool for graphing corpora with an overlay of key events. We now expect to expand the window and reapply the tools. Once we believe we have analytics that produce useful results we then plan to prototype real-time tools that can gather data and track issues as they unfold.

7.0 References

- Blei, D. M. (2001). "Topic Modeling and the Digital Humanities." *Journal of Digital Humanities*. Vol. 2, Issue 1.
<http://journalofdigitalhumanities.org/2-1/topic-modeling-and-digital-humanities-by-david-m-blei/>
- Blei, D. M., et al. (2003). "Latent Dirichlet Allocation." *Journal of Machine Learning Research*. 3: 993-1022.
- Bratt, Duane & Lisa Young. (2020). "The politics behind Jason Kenney's 'tepid' response to COVID-19." CBC
<https://www.cbc.ca/news/canada/calgary/road-ahead-jason-kenney-alberta-covid-politics-1.5818058>
- CBC News (2020). "Jason Kenney has second lowest approval rating of all premiers, Angus Reid poll suggests." CBC
<https://www.cbc.ca/news/canada/calgary/poll-angus-reid-premiers-covid-approval-kenney-1.5706236>
- Duhigg, C. (2020, May 4). "Seattle's Leaders Let Scientists Take the Lead. New York's Did Not." *The New Yorker*.

<https://www.newyorker.com/magazine/2020/05/04/seattles-leaders-let-scientists-take-the-lead-new-yorks-did-not>>

Espinoza, J. (2020, Aug. 15) "Ryan Reynolds and Seth Rogen Urge British Columbians to Stop Partying Amid Pandemic." Complex.
<<https://www.complex.com/pop-culture/2020/08/ryan-reynolds-seth-rogen-urge-british-columbians-avoid-large-gatherings-amid-pandemic>>

Health Canada (n.d.) A Framework for Strategic Risk Communications Within the Context of Health Canada and the PHAC's Integrated Risk Management.
<<https://www.canada.ca/en/health-canada/corporate/about-health-canada/reports-publications/strategic-risk-communications-framework-health-canada-public-health-agency-canada.html>>

Kim, P. (2020, April 27) "During COVID-19 data journalism is taking center stage." Broadcast Dialogue.
<<https://broadcastdialogue.com/during-covid-19-data-journalism-is-taking-centre-stage/>>

Kottasova, I. (2020, July 16). "The muddled public message on coronavirus isn't just confusing. It's harmful." CNN.
<<https://www.cnn.com/2020/07/16/health/coronavirus-pandemic-communication-intl/index.html>>

McCallum, A. K. (2002). "MALLET: A Machine Learning for Language Toolkit." from
<<http://mallet.cs.umass.edu>>.

Quammen, D. (2012). Spillover: Animal Infections and the Next Human Pandemic. New York, W.W. Norton & Company.

Schneiderman, B. (2020, April 30). "Data Visualization's Breakthrough Moment in the COVID-19 Crisis." Nightingale.
<<https://medium.com/nightingale/data-visualizations-breakthrough-moment-in-the-covid-19-crisis-ce46627c7db5>>

Min-max Normalization,
<<https://scikit-learn.org/stable/modules/generated/sklearn.preprocessing.MinMaxScaler.html>>

Pennebaker, J. W., Boyd, R. L., Jordan, K., & Blackburn, K. (2015). *The development and psychometric properties of LIWC2015*.