NRC·CNRC

NRC Publications Archive
Archives des publications du CNRC

Tweet Emotion Dynamics: emotion word usage in tweets from US and Canada

Vishnubhotla, Krishnapriya; Mohammad, Saif M.

This publication could be one of several versions: author's original, accepted manuscript or the publisher's version. / La version de cette publication peut être l'une des suivantes : la version prépublication de l'auteur, la version acceptée du manuscrit ou la version de l'éditeur.

Publisher's version / Version de l'éditeur:

Proceedings of the 13th Language Resources and Evaluation Conference (LREC 2022), pp. 4162-4176, 2022-06-25

NRC Publications Archive Record / Notice des Archives des publications du CNRC :
https://nrc-publications.canada.ca/eng/view/object/?id=7a1b2e6d-4a39-42c5-8d95-a24e3e5107da
https://publications-cnrc.canada.ca/fra/voir/objet/?id=7a1b2e6d-4a39-42c5-8d95-a24e3e5107da

Access and use of this website and the material on it are subject to the Terms and Conditions set forth at
https://nrc-publications.canada.ca/eng/copyright
READ THESE TERMS AND CONDITIONS CAREFULLY BEFORE USING THIS WEBSITE.

L'accès à ce site Web et l'utilisation de son contenu sont assujettis aux conditions présentées dans le site
https://publications-cnrc.canada.ca/fra/droits
LISEZ CES CONDITIONS ATTENTIVEMENT AVANT D'UTILISER CE SITE WEB.

Questions? Contact the NRC Publications Archive team at PublicationsArchive-ArchivesPublications@nrc-cnrc.gc.ca. If you wish to email the authors directly, please see the first page of the publication for their contact information.

Vous avez des questions? Nous pouvons vous aider. Pour communiquer directement avec un auteur, consultez la première page de la revue dans laquelle son article a été publié afin de trouver ses coordonnées. Si vous n'arrivez pas à les repérer, communiquez avec nous à PublicationsArchive-ArchivesPublications@nrc-cnrc.gc.ca.

National Research Council Canada | Conseil national de recherches Canada

Canada

Proceedings of the 13th Conference on Language Resources and Evaluation (LREC 2022), pages 4162–4176
Marseille, 20-25 June 2022
© European Language Resources Association (ELRA), licensed under CC-BY-NC-4.0

Tweet Emotion Dynamics: Emotion Word Usage in Tweets from US and Canada

Krishnapriya Vishnubhotla[*‡] **and Saif M. Mohammad**[†]
[*]Department of Computer Science, University of Toronto
[‡]Vector Institute for Artificial Intelligence
vkpriya@cs.toronto.edu
[†]National Research Council Canada
saif.mohammad@nrc-cnrc.gc.ca

Abstract

Over the last decade, Twitter has emerged as one of the most influential forums for social, political, and health discourse. In this paper, we introduce a massive dataset of more than 45 million geo-located tweets posted between 2015 and 2021 from US and Canada (*TUSC*), especially curated for natural language analysis. We also introduce *Tweet Emotion Dynamics (TED)* — metrics to capture patterns of emotions associated with tweets over time. We use TED and TUSC to explore the use of emotion-associated words across US and Canada; across 2019 (pre-pandemic), 2020 (the year the pandemic hit), and 2021 (the second year of the pandemic); and across individual tweeters. We show that Canadian tweets tend to have higher valence, lower arousal, and higher dominance than the US tweets. Further, we show that the COVID-19 pandemic had a marked impact on the emotional signature of tweets posted in 2020, when compared to the adjoining years. Finally, we determine metrics of TED for 170,000 tweeters to benchmark characteristics of TED metrics at an aggregate level. TUSC and the metrics for TED will enable a wide variety of research on studying how we use language to express ourselves, persuade, communicate, and influence, with particularly promising applications in public health, affective science, social science, and psychology.

Keywords: Emotions, Sentiment analysis, Tweets, Valence, Arousal, Dominance, Corpus Linguistics

1. Introduction

Over the last decade, Twitter has emerged not only as one of the most influential micro-blogging platforms, but also one of the most actively engaging (if sometimes polarizing) fronts for social, political, and even health discourse. Early work (Pak and Paroubek, 2010; Dodds et al., 2011) identified tweets as a crucial indicator of public sentiment. Since then, various samples of tweet data have been used to analyze a wide variety of phenomena, including the recent COVID-19 pandemic. However, past work largely uses topic-based keywords to obtain datasets of interest (often at the expense of geo-location information); for example, work that analyzes emotions in tweets that mention COVID-19-associated terms (Banda et al., 2020; Lwin et al., 2020). Further, very little work explores changes in patterns of emotions of individuals over time.

This paper introduces a new framework to analyze patterns of emotions associated with tweets over time, which we refer to as *Tweet Emotion Dynamics (TED)*. TED builds on ideas first introduced in Hipson and Mohammad (2021), and applies metrics such as *home base, variability,* and *rise rate* to tweets. We also introduce a new dataset of geo-located English *Tweets from US and Canada (TUSC)*. TUSC is not restricted to specific topics and so can be used to study tweets in general, as well as to study notable phenomena (such as a pandemic, climate change, or polarizing political events) on tweets at large (as opposed to examining tweets directly discussing those phenomena). TUSC also includes a subset, (*TUSC100*), made up of tweets from 170K tweeters who each posted at least a 100 tweets between 2020 and 2021. TUSC100 is especially well suited for longitudinal analysis. The creation of the datasets included careful post-processing to make the resource particularly suitable for textual analysis.

TUSC and TED can each be used, together or independently, to explore a wide range of research questions pertaining to tweets and emotions that may be of interest to researchers in Psychology, Affective Science, Social Science, Behavioural Science, Public Health, NLP, and Linguistics. In this paper, we use them to explore questions about how people use emotion-associated words in English tweets from US and Canada.[1] We record the common characteristics of emotion word usage from 2015 to 2021, with a special focus on 2020 — the year that the WHO declared the Novel Coronavirus Disease (COVID-19) outbreak to be a pandemic — and its adjoining years (2019 and 2021). Finally, we benchmark individual tweeter behaviour in terms of various TED metrics. Recording this information holds considerable promise in future work; for example, for studying the emotional impact of the pandemic, for helping clinicians and patients track emotional well-being before and after health interventions, studying emotion regulation and coping strategies, etc. The data (tweet IDs), Emotion Dynamics code, and visualizations are freely available through the project homepage.[2]

[1]Words have associations with emotions; e.g., *success* with pleasure, *illness* with displeasure, etc.

[2]https://github.com/Priya22/EmotionDynamics

2. Related Work

We group related work into two kinds: 1. psychological and psychology-inspired research on the theory of emotions and utterance emotion dynamics; and 2. NLP research in analyzing emotions in tweets.

2.1. Emotions

Several influential studies have shown that the three most fundamental, largely independent, dimensions of affect and connotative meaning are valence (V) (positiveness–negativeness / pleasure–displeasure), arousal (A) (active–sluggish), and dominance (D) (dominant-–submissive / in control–out of control) (Osgood et al., 1957; Russell and Mehrabian, 1977; Russell, 2003). Valence and arousal specifically are commonly studied in a number of psychological and neuro-cognitive explorations of emotion.

The NRC VAD Lexicon (Mohammad, 2018) contains about twenty thousand commonly used English words (lemmas and common morphological variants) that have been scored on valence (0 = maximally unpleasant, 1 = maximally pleasant), arousal (0 = maximally calm/sluggish, 1 = maximally active/intense), and dominance (0 = maximally weak, 1 = maximally powerful).[3] As an example, the word *nice* has a valence of .93, an arousal of .44, and dominance of .65, whereas the word despair has a valence of .11, an arousal of .79, and dominance of .25.

Hipson and Mohammad (2021) introduced Utterance Emotion Dynamics (UED), a framework to quantify patterns of change of emotional states associated with utterances along a longitudinal (temporal) axis. Specifically, they proposed a series of metrics, including:

1. *Density or Mean*: A measure of the average utterance emotional state. For example, the proportion of emotion words a person utters over a given span of time. If each word has a real-valued emotion score (say, for V, A, and D), then this is calculated as the mean of emotion scores of the words in the utterance window. This roughly captures the utterance emotional state.
2. *Variability*: The extent to which a speaker's utterance emotional state changes over time (measured as the standard deviation of the emotion states).
3. *Home Base:* A speaker's home base is the subspace of high-probability emotional states where they are most likely to be found. This is formulated as the range of values within one standard deviation of the average of the emotion states at each timestep.
4. *Rise and Recovery Rates:* Sometimes a speaker moves out of their home state, reaches a peak value of emotion state, before returning to the home state. The rise rate quantifies the rate at which a speaker moves towards the peak; recovery rate is the rate at which they go from the peak to the home state.

[3] http://saifmohammad.com/WebPages/nrc-vad.html

One can determine UED metrics using: 1. the utterances by a speaker, 2. the temporal information about the utterances, for e.g., time stamps associated with the utterances, or simply an ordering of utterances by time, and 3. features of emotional state drawn from text. The emotional state at a particular instant can be determined using simple lexical features (say, drawn from emotion lexicons), predictions of supervised machine learning systems, etc.

Hipson and Mohammad (2021) apply this framework to utterances from a corpus of movie dialogues, which are naturally ordered along a temporal axis. They represented emotional state in a two-dimensional valence–arousal space. The co-ordinates are determined by the average valence and arousal scores of the words (using the NRC VAD lexicon) in a small window of recent utterances (usually spanning 20 to 50 words). Rolling windows of words (moving forward one-word at a time) determine the sequence of emotional states. Here, we apply that framework to tweets. However, in this work we consider each of the valence, arousal, and dominance dimensions separately (separate one-dimensional axes).

2.2. Analyzing Emotions in Tweets

Dodds et al. (2011) analyzes large amounts of Twitter data to explore temporal patterns of 'societal happiness'. Larsen et al. (2015) show a correlation between patterns of emotional expression in tweets with WHO data on anxiety and suicide rates, across geographical location. Snefjella et al. (2018) analyze differences in language use in 40 million tweets from Canada and the USA, and find that the former tend to use more positive language, which correlates with national character stereotypes of Canadians being more agreeable and less aggressive. Twitter data has been used to study people's emotions during significant events, commonly revolving around certain tragedies and natural disasters, and significant political events. Doré et al. (2015) studied the changes in intensity of emotions of anxiety, anger, and sadness expressed on Twitter regarding the Sandy Hook Elementary School shooting. The 2016 US Presidential Election spurred several studies on the language used across geographical and political lines (Littman et al., 2016). Twitter was also used to measure the impacts of the COVID-19 pandemic on the emotional states and mental health of tweeters (Banda et al., 2020). Lwin et al. (2020), for example, looked at changes in the usage of tweets that expressed fear, anger, sadness, and joy in COVID-associated tweets from January 28 to April 9 2020. In our work, we focus on the emotion dimensions of valence, arousal, and dominance, rather than categorical dimensions such as anger, fear, sadness, etc. We also study these patterns of emotion usage across a large time period (2015–2021), and in geo-located tweets.

Dataset	Canada			USA		
	#tweets	# tweeters	Av.TpT	#tweets	# tweeters	Av.TpT
TUSC-Country						
2015	89,566	40,290	15.729	131,330	104,670	13.805
2016	93,280	40,994	16.164	133,413	109,110	14.305
2017	94,364	39,258	18.067	133,854	107,080	16.015
2018	95,403	38,866	21.763	133,066	105,227	19.394
2019	330,361	70,122	22.040	339,186	204,311	19.341
2015–2019	702,974	159,284	18.753	870,849	516,885	16.572
2020	321,176	57,465	22.123	503,976	250,080	19.698
2021	304,106	49,128	22.192	478,798	214,653	19.566
2015–2021	1,328,256	206,691	19.73	1,853,623	802,369	17.45
TUSC-City						
2020 (Apr–Dec)	15,039,503	716,063	19.275	23,470,855	2,669,081	17.556
2021	22,371,990	798,602	19.367	43,693,643	3,247,124	17.306
2020–2021	37,411,493	1,049,774	19.327	67,164,498	4,274,374	17.413

Table 1: #tweets, #tweeters, and Average number of token per tweet (Av.TpT) in the TUSC Datasets.

3. Tweets Dataset: TUSC

3.1. Sampling Tweets

Twitter's regular API allows one to obtain a random sample of tweets from the past week. However, the search is limited to only the tweets from the past week. The Academic search API provides access to historical tweets, but with a lower rate limit. To benefit from both APIs and to confirm that our results are consistent regardless of the API and search method, we compiled two separate tweet datasets using each of the APIs:

1. using Twitter's free API and its geo-location and random-sample switches to collect tweets from 46 prominent American and Canadian cities. Data collection began in April 1, 2020 and is ongoing. We refer to the dataset created with this method as *TUSC-City*.

2. using Twitter's Academic API to collect tweets emanating from US and Canada from Jan 2015 to Dec 2021. The Academic API provides switches to specify the country of origin and the time span of search. However, the sample of results it provides tend to be in reverse chronological order for the specified time span. Thus, to obtain a sample of tweets from various time spans across the various years of interest, we employed the following strategy: For each year of interest, we randomly generated a date time (using unix epoch seconds). We then specified a search interval of 8 hours starting from that date time. We repeated this procedure thousands of times for each year. Since we were especially interested in the years of 2019, 2020, and 2021, we collected more data from these years. We refer to the resulting dataset as *TUSC-Country*.

3.2. Tweet Curation

We curated the tweet collection to make it more suitable for computational natural language analyses by applying the following steps:

- Kept one tweet per user, per day. This mitigates the impact of highly prolific tweeters and commercial accounts on the dataset.
- Kept only English language tweets (since the English set is the focus of this project). These are identified by the `iso_language` tag provided by Twitter for each tweet.
- Removed all retweets.
- Removed all tweets containing a URL and/or links to media (to focus on textual tweets). This also limits tweets by commercial organizations.
- Discarded all tweets with less than three tokens. This eliminates certain formulaic tweets such as wishes for holidays. The tweet text is tokenized using the Python implementation[4] of the Twokenizer package (Gimpel et al., 2010; Owoputi et al., 2013).

We kept quotes and replies as they include new textual information.

3.3. Key Data Statistics and Distribution

We organize the TUSC tweets as per the sampling strategy used to obtain them (see TUSC-Country and TUSC-City in §3.1) as well as the year of posting (2015 through 2021), and country of origin (US, Canada). Table 1 shows the number of tweets, number of tweeters, and average number of tokens per tweet in each of these dataset groupings. (Table 3 in the Appendix shows a breakdown by city for TUSC-City.) It is interesting that an average Canadian tweet has about two more tokens per tweet than a US tweet (one possible explanation is the tendency of American tweeters to use more informal and non-standard language, as found in Snefjella et al. (2018)).

TUSC-City is the larger dataset, and contains millions of tweets for many of the 46 cities for 9 months in 2020 (Apr–Dec), and all the months of 2021. It is useful for analyzing trends at the city–level, and also at the user-level, since we are more likely to have a large number of tweets from the same user.

[4] `https://github.com/myleott/ark-twokenize-py`

4. Emotion Word Usage in American and Canadian Tweets

The TUSC datasets can be used to answer several important questions about emotion word usage in English tweets from US and Canada, including:

- Are there notable trends across years in the valence, arousal, and dominance of tweets? Are we tweeting with more positive words, more negative words, more high arousal words, etc. than in past years?
- How has the COVID-19 pandemic impacted the emotionality of our tweets? At what point of time in the pandemic did we use the most amount of words conveying a lack of control and uncertainty? How were individual cities impacted?
- How are Canada and US different in terms of emotion word usage? Did the pandemic impact the emotionality differently in the two countries?

We will explore these, and other, questions below.

We used the NRC Valence, Arousal, and Dominance (NRC VAD) Lexicon (Mohammad, 2018) to determine the emotion associations of the words in tweets.[5] Specifically, we used the subset with entries for only the polar terms: i.e., only those valence entries that had scores $\leq$0.33 (negative words) or scores $\geq$0.67 (positive words).[6] Similarly, only those arousal and dominance entries were included that had scores $\leq$0.33 or $\geq$0.67. The entries with scores between 0.33 and 0.67 are considered neutral for that dimension.

Methodological Note 1: As is good practise in lexicon-based analysis (Mohammad, 2020), we removed lexicon entries for a small number of words that were highly ambiguous (e.g., *will, like*) or were expected to be frequently used in our tweets in a sense that is different from the usual predominant sense of the word (e.g., *trump*). The list of the 23 terms removed from the lexicon, and a description of the process of discovering them, is available in the Appendix.

Methodological Note 2: Similar analyses can also be performed using categorical emotions, such as joy, sadness, fear, anger, etc., using the NRC Emotion Lexicon (Mohammad and Turney, 2010; Mohammad and Turney, 2013).[7] See discussions on categorical and dimensional emotions in (Mohammad, 2021).

4.1. Average V, A, and D Across US–Canada

For each tweet, we take the average of the valence, arousal, and dominance values of each of the words in the tweet text. The averages are computed for TUSC-City over all tweets from each city, and for TUSC-Country at the country-level. We test whether the differences in values between countries and years are statistically significant by using the paired t-Test, with the significance threshold for the p-value set to 0.001.

Valence

Dataset	Year	Canada	USA
TUSC-Country	2015	0.675	0.644
	2016	0.672	0.644
	2017	0.666	0.639
	2018	0.669	0.640
	2019	0.668	0.638
	2015--2019	0.670	0.641
	2020	0.662	0.634
	2021	0.669	0.644
TUSC-City	2020	0.651	0.634
	2021	0.658	0.645

Arousal

Dataset	Year	Canada	USA
TUSC-Country	2015	0.461	0.465
	2016	0.464	0.469
	2017	0.465	0.472
	2018	0.463	0.472
	2019	0.461	0.476
	2015--2019	0.463	0.471
	2020	0.458	0.474
	2021	0.457	0.474
TUSC-City	2020	0.469	0.479
	2021	0.470	0.479

Dominance

Dataset	Year	Canada	USA
TUSC-Country	2015	0.564	0.535
	2016	0.567	0.543
	2017	0.568	0.546
	2018	0.576	0.553
	2019	0.577	0.553
	2015--2019	0.570	0.546
	2020	0.574	0.555
	2021	0.578	0.560
TUSC-City	2020	0.571	0.558
	2021	0.576	0.563

Figure 1: Average Valence, Arousal, and Dominance of words per tweet for each dataset (year).

Yearly Trends: Figure 1 shows the average V, A, and D scores of tweets when aggregated at the country and year level, for the various data subsets. The gradient bars at the top show the colors used to indicate where the values lie in the spectrum from lowest to highest.

Valence: Observe that the average valence of Canadian tweets is consistently higher (more positive) than the tweets from the US (statistically significant); the difference are steady across years. There is a slight downward trend for valence in both countries from 2015 to 2019. We see the lowest values of mean valence occur in 2020 for both TUSC-Country and TUSC-City (the year the pandemic hit) for both the US and Canada. Average valence rises back up in 2021.

[5] http://saifmohammad.com/WebPages/nrc-vad.html

[6] There is no "correct" threshold to determine these classes; different thresholds simply make the positive and negative classes more or less restrictive.

[7] http://saifmohammad.com/WebPages/NRC-Emotion-Lexicon.htm

Country	Month	2015--2019	2019	2020	2021
Canada	Jan	0.671	0.673	0.663	0.662
	Feb	0.671	0.665	0.667	0.667
	Mar	0.671	0.667	0.663	0.669
	Apr	0.670	0.667	0.664	0.665
	May	0.675	0.671	0.660	0.668
	Jun	0.673	0.671	0.656	0.669
	Jul	0.667	0.668	0.662	0.670
	Aug	0.668	0.666	0.660	0.668
	Sep	0.669	0.665	0.662	0.667
	Oct	0.662	0.662	0.663	0.673
	Nov	0.668	0.668	0.659	0.674
	Dec	0.675	0.671	0.669	0.676
USA	Jan	0.642	0.641	0.634	0.633
	Feb	0.645	0.644	0.641	0.643
	Mar	0.643	0.641	0.633	0.646
	Apr	0.644	0.639	0.637	0.641
	May	0.646	0.639	0.634	0.646
	Jun	0.642	0.641	0.625	0.642
	Jul	0.636	0.632	0.634	0.643
	Aug	0.640	0.636	0.634	0.643
	Sep	0.638	0.635	0.633	0.645
	Oct	0.634	0.635	0.632	0.643
	Nov	0.640	0.639	0.631	0.647
	Dec	0.643	0.640	0.640	0.652

Mean 0.625 0.676

Figure 2: Monthly trends in valence (TUSC-Country).

Arousal: Overall, tweets from Canada have lower average arousal (more calm, less active) than the US (statistically significant). Again, the difference in mean between the two countries remains relatively steady across years. Across years, arousal values for both countries increase from 2015 till 2017; they then drop steadily for Canada, while the USA sees a peak in 2019 followed by slight drops in subsequent years.

Dominance: Canada on the whole consistently has higher dominance values (greater feeling of control) than the US across the Both countries have the lowest dominance values in 2015, and the highest in 2021. For all three dimensions, we note that the yearly trends observed in TUSC-Country are largely also observed in the TUSC-City trends, across 2020–2021.

Monthly Trends: Figure 2 shows a breakdown of the average valence scores at the month level, across years. (Figures 6 and 7 in the Appendix show the breakdown for arousal and dominance.)

We immediately notice from the color shading that Canada consistently has higher valence (green), lower arousal (blue), and higher dominance (purple) than the US, across the months and years. June 2020 is particularly notable as it has the lowest values of valence for both USA and Canada; we hypothesize that this is an effect of both the COVID-19 pandemic (the seriousness of which was starting to become evident a couple of months earlier in March 2020) and the *black lives matter* protests (which peaked after the Dylan Roof shooting incident). By contrast, the final months of 2021 have the highest positivity. This could be attributed to feelings of a potential return to normalcy, and a general uptick in mood due to the holiday season (this was just before the Omicron variant took root in early 2022).

The dominance numbers indicate that April and May of 2020 for Canada and the USA are marked by some of the lowest scores, suggestive of a feeling of loss of control due to the onset of the global COVID-19 pandemic.

4.2. Tweets with Emotional Terms

The experiments above showed notable differences in the average V/A/D scores of tweets across US and Canada. However, they also lead to further questions such as whether the higher valence in Canadian tweets is because of a greater usage of positive words or a lower usage of negative words. To explore such questions, we determine how frequently people post tweets with at least one high valence word, how frequently people post tweets with at least one low valence word, how frequently people post tweets with at least one high arousal word, and so on. High and low categorization of a word is based on whether their score (for V/A/D) is $\geq$0.67 or $\leq$0.33, respectively. Figure 3 shows the results.

The darker shades of the color indicate a greater percentage of tweets had at least one of the relevant emotional words.

Observe that in both American and Canadian tweets:

- people post markedly more tweets with at least one positive word than tweets with at least one negative word (about 100% more).
- people post markedly more tweets with at least one low arousal word than tweets with at least one high-arousal word (about 40% more).
- people post markedly more tweets with at least one high dominance word than tweets with at least one low-dominance word (about 33% more).

In terms of their differences, we see that the tweets from Canada are marked by both a higher usage of high-valence words, as well as a lower usage of low-valence words, than the US (statistically significant). Tweets from Canada have a higher proportion of low arousal words, whereas high arousal word usage is similar in both countries. Canadian tweeters use about the same number of low dominance words as those in the US, but use a greater number of high dominance words.

Across years, low valence words increase in usage relatively steadily until 2020, and drop in 2021. For all dimensions, the sharpest rise in usage occurs from 2016 to 2017. When comparing TUSC-Country 2020 with 2021, observe that the higher number of low valence words used is more prominent than the lower number of high valence words — thus, the drop in average valence in 2020 (Figure 1) is because people *tweeted more negative words* (and not because people *tweeted less positive words*).

Dataset	Year	Low Valence		High Valence	
		Canada	USA	Canada	USA
TUSC Country	2015	36.3	38.4	80.2	75.4
	2016	37.6	39.6	80.8	76.7
	2017	40.8	43.2	82.9	79.0
	2018	42.7	45.7	83.2	80.2
	2019	42.9	45.3	82.7	79.4
	2015--2019	40.0	42.4	82.0	78.1
	2020	43.8	46.1	82.3	79.2
	2021	42.6	44.5	82.5	79.7

Dataset	Year	Low Arousal		High Arousal	
		Canada	USA	Canada	USA
TUSC Country	2015	55.8	51.1	40.9	38.1
	2016	56.4	52.0	42.1	40.0
	2017	59.4	55.6	45.0	43.5
	2018	61.4	57.8	47.1	46.0
	2019	60.8	56.7	46.5	46.1
	2015--2019	58.8	54.6	44.3	42.7
	2020	60.8	57.0	45.7	45.9
	2021	61.1	57.2	45.6	46.1

Dataset	Year	Low Dominance		High Dominance	
		Canada	USA	Canada	USA
TUSC Country	2015	38.6	38.4	53.1	45.7
	2016	39.2	39.1	55.1	48.5
	2017	42.3	42.2	58.4	52.6
	2018	44.1	44.6	61.5	56.0
	2019	43.8	44.1	61.4	55.6
	2015--2019	41.6	41.7	57.9	51.7
	2020	44.4	44.3	61.0	56.1
	2021	43.8	43.4	61.5	56.5

Figure 3: Percentage of tweets with at least one — low valence word, high valence word, low arousal word, high arousal word, low dominance word, high dominance word — across datasets (years).

See Vishnubhotla and Mohammad (2022) for an analysis of the topics driving the differences across V, A, and D: across the years, and across US and Canada.

5. Tweet Emotion Dynamics

While the previous section looked at samples of tweets emanating from countries and cities as a whole, in this section we explore individual tweeter behaviours and *metrics* of emotion word usage over time. Specifically, we apply the framework of Utterance Emotion Dynamics (UED) (Hipson and Mohammad, 2021) to tweets to explore a number of questions, such as:

- What is the usual range and distribution of various metrics of tweet emotion dynamics (TED), such as mean and recovery rate, for American and Canadian tweeters? Establishing benchmarks for these metrics is crucial for subsequent studies that may explore, for example, the impact of a health intervention on one's TED metrics.
- Are there notable differences in the distributions of TED metrics across American and Canadian tweeters?
- Are there notable differences in the distributions of TED metrics across 2020 and 2021?

Recall that computing UED metrics requires: 1. a set of texts associated with each speaker, 2. a temporal ordering of these texts, and 3. a way to determine emotions associated with the texts. The timestamp of a tweet provides the temporal order. The tweets from each speaker are then concatenated together and tokenized to obtain a ordered list of tokens. Next, a rolling window of 20 tokens is considered to determine the average V, A, and D scores of the words in that window. These scores are a representation of utterance emotional state corresponding to that window. The rolling window is moved forward one word at a time to determine the subsequent averages.[8] In the rest of this section, we use the term *mean* to refer to the mean of all the rolling window averages for a speaker.

We re-implemented the Hipson and Mohammad (2021) R code in Python. Also, rather than using the valence–arousal two-dimensional (ellipse-based) representation of emotions explored in their work, we analyse the dynamics of each emotion dimension separately along one-dimensional axes. This eases the interpretation of the UED metrics, and allows for considering more dimensions such as dominance. We also break down the rise and recovery rates into separate rates corresponding to when one is moving from the home base to higher emotion values (Hm–Hi), from the highest value to home (Hi–Hm), from the home base to lower emotion values (Hm–Lo), and from the lowest value to home (Lo–Hm). This allows us to explore whether, for example, some tweeters have close-to-median Hi–Hm rates, but markedly low Lo–Hm rates (indicating that once they start uttering negative words, then they tend to dwell in the negatives and only gradually return to their home state).

We use the code to determine TED metrics for the tweeters in in the TUSC-Country dataset. Only tweeters with at least 100 tweets in a year were considered, since drawing inferences about one's tweeting behavior requires a sufficient sample size. There were about 40K such tweeters in the 2020 subset and about 130K such tweeters in the 2021 subset. We refer to their tweets (5.6M from 2020 and 19M from 2021) as the *TUSC100-2020* and *TUSC100-2021* datasets, respectively. Average number of tweets by a tweeter in these datasets is 153 (no tweeters had more than 365 tweets due to our earlier stated 'one tweet per user per day' pre-processing policy). See Table 4 in the Appendix for detailed statistics.

Figure 4 plots the distributions of some of the metrics for the joint set of 2020 and 2021 tweeters. The plots in (a) are distributions of the mean values for the three emotion dimensions (V, A, D). The x-axis is made up of bins of size 0.005 (from 0–0.005 to 0.995–1). The

[8] Variations of this approach that do not use rolling windows across tweet boundaries produce similar results.

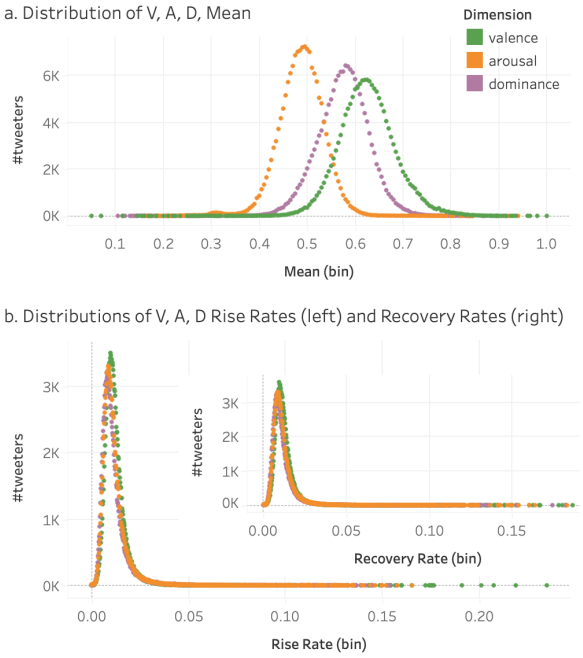

Figure 4: Distributions of Means, Rise Rates, and Recovery Rates for Valence, Arousal, and Dominance (TUSC100).

y-axis indicates the number of tweeters with mean values in each of the bins. Observe that the means for V, A, and D all follow a near-normal distribution. Mean valence and dominance values are more spread out compared to arousal values. For V, most fall between 0.5–0.8 , with a median value of around 0.65, though we can see that there is a long tail of outliers. Dominance scores are spread around a median score of 0.6, and the median is even lower for arousal (0.49).

Figure 4(b) shows distributions for rises rates and recovery rates. Observe that these have a much narrower spread, and the distributions for all three dimensions are roughly the same.

Figure 5 shows box and whisker plots of the same three metrics: mean, rise rate, and recovery rates. However, separate plots are shown for tweeters from US and Canada, and across 2020 and 2021. The shaded region (the box) indicates the "middle portion" of the data distribution, i.e, the range covered between the first quartile (the 25% mark) and the third quartile (the 75% mark), with the median (50% mark) lying at the border of the light and dark shaded regions. The whiskers, the lines on either end of the plot, are at a distance of 1.5 times the inter-quartile length (inter-quartile length is the distance between the first and third quartiles). Points beyond the whiskers are considered outliers. Additionally, the average value (mean) is indicated with the pink horizontal dashed line.

Observe that the mean valence is lower in 2020 than in 2021, and Canadian tweeters on average use more positive words than their US counterparts.[9] The distributions for mean arousal are quite similar across 2020 and 2021, but US tweeters have slightly higher mean arousal values. Canadian tweeters have a slightly higher median of dominance scores than US tweeters; whereas the US tweeters tend to have a wider range of dominance values. The difference in the distributions of the mean values for Canada and US is statistically significant for all three dimensions (p-values < 0.001).

The median rise rates and recovery rates do not differ markedly across countries or years. However, there is a notably large range of the third quartile (the quartile above the median) for Canadian tweeters in 2021. These are tweeters who are quicker to jump in and out of their home base. Tables 5, 6, 7 of the Appendix report mean scores for all of the TED metrics, averaged across all tweeters by country and year. This includes a breakdown of the rates into Hm-Hi, Hi-Hm, Hm-Lo, and Lo-Hm. Notable trends there are that the average rise and recovery rates on the high side of the home state (Hm-Hi, Hi-Hm) are lower than for the low side of the home state (Lo-Hm, Hm-Lo), for the valence and dominance dimensions. This says that tweeters are slower to rise to more positive and more dominant states, but quicker to both descend to more negative and less dominant states, and recover from them; similarly, they are slower to transit to and from states of high activity (high dominance). This difference between Hi and Lo rates is reversed for arousal. Thus, we are quicker to rise to states of high arousal, and come back down from them to the home state.

We also noticed in our analyses that there exist several tweeters that have very high rise rates but normative recovery rates, and also tweeters that have very high rise rates but normative recovery rates. Identifying such characteristics and tracking them in the context of health interventions is particularly promising future work. However, it should be noted that we strongly encourage such studies, when conducted, to be led by clinicians and psychologists, with appropriate consent and ethics approvals.

5.1. City as Speaker

An interesting variation of the experiments above, is to consider each city as a 'speaker', rather than individual tweeters. Figure 8, in the Appendix, shows the average TED metrics for each of the 46 cities in TUSC-City. The color gradients make it easy to spot which cities have had markedly high V/A/D means across 2020 and 2021. Consistent with some of the earlier country-level results, we see that the Canadian cities tend to have higher valence, lower arousal, and higher dominance, than the US cities. London, Ottawa, Halifax, and Victoria have the highest valence (most positive). From the set of Canadian cities, Windsor stands out as an anomaly with valence close to many US cities. Detroit, Houston, Los Angeles, and Philadelphia have some of the lowest valence values of all cities. All cities improve from 2020 to 2021, some more drastically than

[9] These trends align with the trends observed in Table 2.

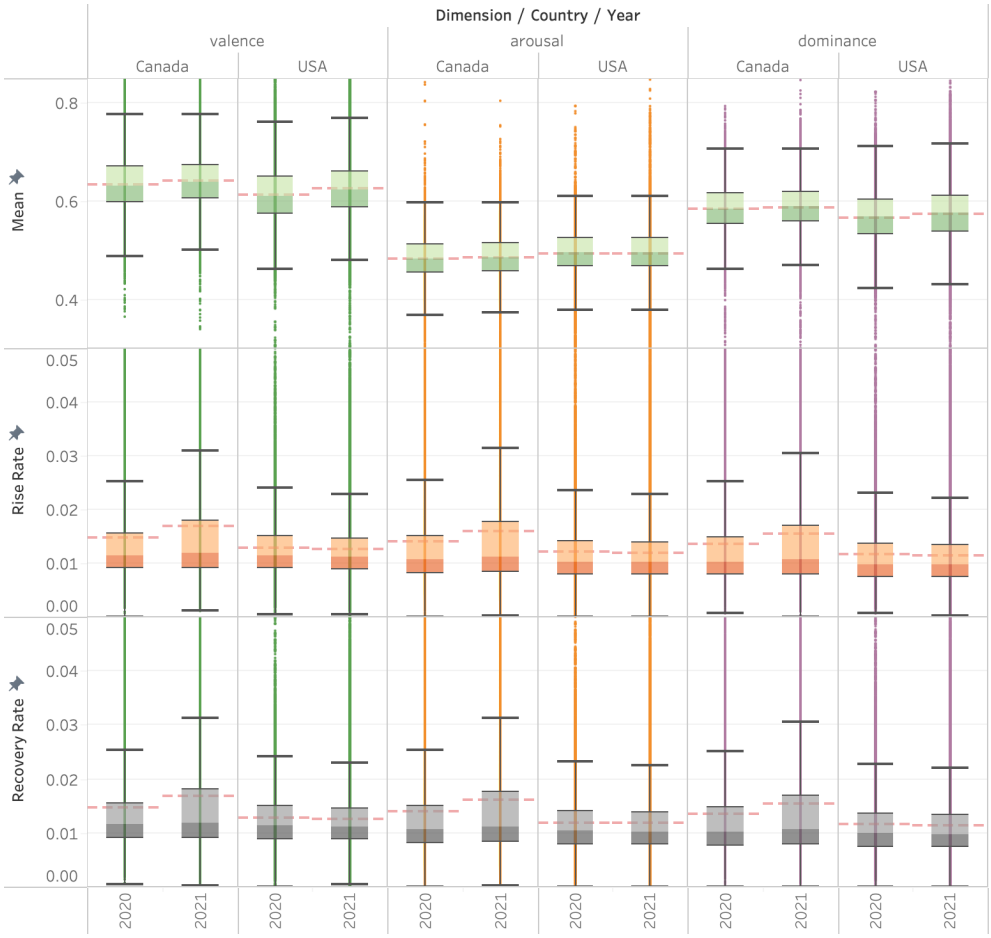

Figure 5: Box plots of means, rise rates, and recovery rates of Valence, Arousal, and Dominance of tweeters in 2020 and 2021 (TUSC100-2020 and TUSC100-2021).

others (Boston, Indianapolis, San Jose, for example). Quebec City and Windsor have the highest arousal rates in Canada; in the US, El Paso is at the top for both years. Nashville, San Francisco, San Jose, and Seattle have lower arousal rates (more in line with the average Canadian city). Washington, San Jose, and Boston also show markedly high dominance, as well as San Francisco. Among Canadian cities, Ottawa and Victoria have the highest dominance scores for 2020 and 2021, and Windsor again the lowest.

Figure 9, in the Appendix, shows values of the variabilities, rise rates, and recovery rates for the valence dimension. Looking at the column for variability, Windsor jumps out among the Canadian cities for having comparatively higher variability. Washington and Phoenix in 2020 have relatively high variability. Moving to the next columns, Windsor again has the highest rise and recovery rates among Canadian cities; US cities are the on the whole quicker to rise and fall.

The various metrics listed for various cities should be useful to those interested in the tweets from particular cities. Future work will drill down further into the data for individual cities to determine the factors driving the emotion word usage.

6. Conclusion

We introduced the Tweet Emotion Dynamics (TED) framework to quantify changes in emotions associated with tweets over time. We also released TUSC — a large collection of English geo-located tweets from Canada and the USA that were posted between 2015 and 2021. We studied emotion word usage in this data, using multiple metrics, for the primary dimensions of valence, arousal, and dominance. Our results showed interesting trends in the emotions expressed by tweeters from the two countries across different years, and also uncovered contrasts between Canadian and US tweeters. An expanded version of this paper presents further experiments, including a deeper analysis of the words and topics driving the TED metrics (Vishnubhotla and Mohammad, 2022). Future work will explore tweets from other countries and also tweets in languages other than English. We will also examine how the TED framework can be useful to clinicians and psychologists for measuring mental health outcomes, at an aggregate level, from social media data. Additionally, we will study the applications of Utterance Emotion Dynamics in other contexts such as novels, personal diaries, forum posts, and speech.

Acknowledgements

We thank Will Hipson for early discussions on his R code for utterance emotion dynamics. The first author is supported by funding from the Natural Sciences and Engineering Research Council of Canada, and resources provided by the Vector Institute of Artificial Intelligence.

7. Bibliographical References

Banda, J. M., Tekumalla, R., Wang, G., Yu, J., Liu, T., Ding, Y., and Chowell, G. (2020). A large-scale COVID-19 twitter chatter dataset for open scientific research - an international collaboration. *CoRR*, abs/2004.03688.

Dodds, P. S., Harris, K. D., Kloumann, I. M., Bliss, C. A., and Danforth, C. M. (2011). Temporal patterns of happiness and information in a global social network: Hedonometrics and twitter. *PloS one*, 6(12):e26752.

Doré, B., Ort, L., Braverman, O., and Ochsner, K. N. (2015). Sadness shifts to anxiety over time and distance from the national tragedy in newtown, connecticut. *Psychological Science*, 26(4):363–373. PMID: 25767209.

Gimpel, K., Schneider, N., O'Connor, B., Das, D., Mills, D., Eisenstein, J., Heilman, M., Yogatama, D., Flanigan, J., and Smith, N. A. (2010). Part-of-speech tagging for twitter: Annotation, features, and experiments. Technical report, Carnegie-Mellon Univ Pittsburgh Pa School of Computer Science.

Hipson, W. E. and Mohammad, S. M. (2021). Emotion dynamics in movie dialogues. *PLOS ONE*, 16:1–19, 09.

Larsen, M. E., Boonstra, T. W., Batterham, P. J., O'Dea, B., Paris, C., and Christensen, H. (2015). We feel: mapping emotion on twitter. *IEEE journal of biomedical and health informatics*, 19(4):1246–1252.

Littman, J., Wrubel, L., and Kerchner, D. (2016). 2016 United States Presidential Election Tweet Ids.

Lwin, M. O., Lu, J., Sheldenkar, A., Schulz, P. J., Shin, W., Gupta, R., and Yang, Y. (2020). Global sentiments surrounding the covid-19 pandemic on twitter: Analysis of twitter trends. *JMIR Public Health Surveill*, 6(2):e19447, May.

Mohammad, S. M. and Turney, P. D. (2010). Emotions evoked by common words and phrases: Using Mechanical Turk to create an emotion lexicon. In *Proceedings of the NAACL-HLT Workshop on Computational Approaches to Analysis and Generation of Emotion in Text*, LA, California.

Mohammad, S. M. and Turney, P. D. (2013). Crowdsourcing a word-emotion association lexicon. *Computational Intelligence*, 29(3):436–465.

Mohammad, S. M. (2018). Obtaining reliable human ratings of valence, arousal, and dominance for 20,000 english words. In *Proceedings of The Annual Conference of the Association for Computational Linguistics (ACL)*, Melbourne, Australia.

Mohammad, S. M. (2020). Practical and ethical considerations in the effective use of emotion and sentiment lexicons. *arXiv:2011.03492*.

Mohammad, S. M. (2021). Sentiment analysis: Automatically detecting valence, emotions, and other affectual states from text. In Herbert L. Meiselman, editor, *Emotion Measurement (Second Edition)*, pages 323–379. Woodhead Publishing, second edition edition.

Mohammad, S. M. (2022). Ethics sheet for automatic emotion recognition and sentiment analysis. *To Appear in Computational Linguistics*, June.

Osgood, C. E., Suci, G. J., and Tannenbaum, P. (1957). *The measurement of meaning*. University of Illinois Press.

Owoputi, O., O'Connor, B., Dyer, C., Gimpel, K., Schneider, N., and Smith, N. A. (2013). Improved part-of-speech tagging for online conversational text with word clusters. In *Proceedings of the 2013 conference of the North American chapter of the association for computational linguistics: human language technologies*, pages 380–390.

Pak, A. and Paroubek, P. (2010). Twitter as a corpus for sentiment analysis and opinion mining. In *Proceedings of the Seventh International Conference on Language Resources and Evaluation (LREC'10)*.

Russell, J. A. and Mehrabian, A. (1977). Evidence for a three-factor theory of emotions. *Journal of research in Personality*, 11(3):273–294.

Russell, J. A. (2003). Core affect and the psychological construction of emotion. *Psychological review*, 110(1):145.

Snefjella, B., Schmidtke, D., and Kuperman, V. (2018). National character stereotypes mirror language use: A study of canadian and american tweets. *PloS one*, 13(11):e0206188.

Vishnubhotla, K. and Mohammad, S. M. (2022). Tweet emotion dynamics: A framework to analyze emotions over time. In *ArXiv*.

Appendix

8. Ethics Considerations

Emotions are complex, private, and central to our experience. There is also tremendous variability in how we express emotions through language. Thus several ethical considerations are relevant to textual analysis of emotions. Some that we would particularly like to highlight are listed below:

- We only release the tweet IDs for each tweet, which will need to be hydrated by users of our dataset with the Twitter API. If any of the tweets are deleted by the associated tweeter, it will no longer be accessible.
- Our work on studying emotion word usage should not be construed as detecting how people feel;

rather, we draw inferences on the emotions that are conveyed by users via the language that they use.

- The language used in an utterance may convey information about the emotional state (or perceived emotional state) of the speaker, listener, or someone mentioned in the utterance. However, it is not sufficient for accurately determining any of their momentary emotional states. Deciphering true momentary emotional state of an individual requires extra-linguistic context and world knowledge. Even then, one can be easily mistaken.
- The inferences we draw in this paper are based on aggregate trends across large populations. We do not draw conclusions about specific individuals or momentary emotional states.
- We do not recommend the use of TED metrics to draw inferences about individuals, unless: 1. it is exercised with extreme caution, 2. for the express benefit, and with consent, of the people whose data is used, 3. the work is led by sbject-matter experts such as psychologists or clinicians, and 4. automatically drawn information is used as one source of information among many by human experts.
- Any information drawn from these metrics regarding one's language use should not be used to negatively impact the individual.

See Mohammad (2022) for a detailed discussion on the ethical considerations of automatic emotion recognition and Mohammad (2020) for practical and ethical considerations in the effective use of emotion lexicons.

9. Additional Emotion Word Usage Statistics from TUSC

We present, in this section, additional tables and figures that record details of emotion word usage broken down by city (for the 46 cities considered) and by month of year. Table 4 shows the number of tweets and tweeters in each subset of the TUSC100 dataset. Tables 5, 6, and 7 tabulate the numbers corresponding to the plots in Figure 5 in the main paper.

10. Modified NRC VAD Lexicon

When applying lexicon-based analyses to datasets from a specific domain, Mohammad (2020) recommends updating the emotion lexicons to remove terms that can be used in a sense different from the predominant word sense. Since manual examination of all the words in a large dataset is difficult, this step is recommended for at least the frequent terms.

For our analyses, we first compiled a list of all the terms from the NRC VAD lexicon that occurred in at least 0.1% of the tweets from either Canada or the USA, for any of the years in the TUSC-Country dataset (2015–2021). Both of the authors of this work examined the list and identified words that were highly ambiguous or occurred in the tweets predominantly in a sense different from what would be expected if people were shown the word out of context (as was the case of the original annotations in the NRC VAD lexicon). In all, 23 such words were identified (shown in Table 2). The entries for these words were removed from the lexicon before conducting the experiments described in the paper.

have	will	one	high
may	way	kind	be
thing	things	number	seem
do	look	three	third
five	senate	say	talk
president	trump	like	

Table 2: List of terms that were removed from the NRC VAD Lexicon.

To examine the impact of the above lexicon update, we repeated all the experiments with the unmodified lexicon as well. We observed that while the numerical values of the UED metrics changed slightly (as would be expected), all relative trends remained the same, across countries and across years. The interested reader can find the scores for the UED metric and the complete set of experiments with the unmodified lexicon in version 2 (an older version) of the paper on ArXiv.[10]

[10] https://arxiv.org/abs/2204.04862

	2020		2021	
City	**# tweets**	**# tweeters**	**# tweets**	**# tweeters**
Canada				
Brampton	1,436,865	159,974	2,430,329	188,216
Calgary	294,911	31,988	503,173	39,416
Edmonton	806,116	43,427	1,319,950	49,058
Etobicoke	1,318,119	157,429	2,379,928	191,653
Halifax	572,562	23,733	678,033	23,541
Hamilton	446,038	37,023	702,761	43,537
Laval	453,670	48,145	733,844	58,344
London	298,615	16,977	428,929	18,928
Mississauga	450,835	97,328	977,517	142,817
Montreal	627,159	52,396	1,048,093	64,363
North York	1,274,462	148,271	1,685,201	152,105
Okanagan	30,771	1,814	37,424	1,813
Ottawa	1,055,035	55,430	1,332,680	56,621
Quebec	284,665	16,380	377,342	18,100
Scarborough	720,165	108,498	710,181	86,328
Surrey	1,115,467	84,001	1,679,642	94,177
Toronto	2,058,494	182,730	2,557,606	182,792
Vancouver	402,418	53,655	634,307	63,561
Victoria	340,720	14,787	436,905	15,565
Windsor	443,712	58,545	893,922	72,975
Winnipeg	608,704	27,954	824,223	29,365
US				
Austin	1,244,776	102,841	2,242,561	125,526
Boston	764,257	100,276	1,641,142	130,145
Charlotte	997,197	76,528	1,566,062	86,892
Chicago	721,075	142,591	1,652,701	194,565
Columbus	809,160	69,931	1,445,275	81,135
Dallas	674,613	129,304	1,671,887	181,895
Denver	1,198,813	98,697	1,712,785	106,959
Detroit	749,506	77,560	1,418,484	94,202
El Paso	692,705	38,096	781,937	37,335
Fort Worth	1,649,842	188,443	2,794,053	215,169
Houston	1,557,488	195,548	2,358,286	209,520
Indianapolis	710,808	65,665	1,287,399	78,214
Jacksonville	723,513	48,230	1,000,620	53,257
Los Angeles	1,028,102	246,491	2,470,750	337,004
Memphis	876,988	49,835	1,263,728	54,254
Nashville	584,997	68,306	963,947	79,006
New York	1,079,557	281,109	2,288,500	361,670
Philadelphia	1,142,818	127,257	2,579,605	161,491
Phoenix	531,390	81,093	1,454,042	117,145
San Antonio	1,213,537	91,427	1,835,376	98,868
San Diego	1,080,361	101,554	1,940,120	122,278
San Francisco	1,123,356	132,023	2,179,371	159,727
San Jose	790,511	72,049	1,184,999	78,750
Seattle	940,092	114,848	1,968,319	141,016
Washington	585,393	123,576	1,991,694	189,873
All	**38,510,358**	**3,332,189**	**66,065,633**	**3,976,481**

Table 3: The number of tweets and tweeters in TUSC-City for 2020 and 2021.

Country	Month	Year 2015--2019	2019	2020	2021
Canada	Jan	0.462	0.459	0.462	0.459
	Feb	0.463	0.462	0.462	0.455
	Mar	0.460	0.461	0.449	0.457
	Apr	0.467	0.462	0.451	0.451
	May	0.463	0.462	0.458	0.460
	Jun	0.464	0.463	0.463	0.456
	Jul	0.464	0.462	0.457	0.455
	Aug	0.464	0.461	0.463	0.458
	Sep	0.460	0.459	0.460	0.458
	Oct	0.464	0.463	0.461	0.456
	Nov	0.463	0.461	0.455	0.461
	Dec	0.461	0.460	0.457	0.455
USA	Jan	0.471	0.474	0.479	0.478
	Feb	0.469	0.476	0.476	0.473
	Mar	0.467	0.469	0.466	0.468
	Apr	0.470	0.477	0.467	0.475
	May	0.470	0.475	0.474	0.471
	Jun	0.471	0.474	0.475	0.476
	Jul	0.472	0.475	0.474	0.476
	Aug	0.470	0.477	0.474	0.474
	Sep	0.474	0.477	0.478	0.475
	Oct	0.474	0.479	0.475	0.475
	Nov	0.471	0.476	0.476	0.476
	Dec	0.470	0.477	0.471	0.471

Mean 0.449 0.479

Figure 6: Monthly trends in **Arousal** of tweets across years (TUSC-Country).

Country	Month	Year 2015--2019	2019	2020	2021
Canada	Jan	0.569	0.580	0.572	0.576
	Feb	0.569	0.575	0.577	0.575
	Mar	0.571	0.579	0.575	0.578
	Apr	0.571	0.576	0.569	0.576
	May	0.573	0.580	0.571	0.577
	Jun	0.575	0.583	0.578	0.576
	Jul	0.569	0.576	0.571	0.576
	Aug	0.569	0.574	0.574	0.577
	Sep	0.569	0.574	0.574	0.582
	Oct	0.568	0.578	0.578	0.579
	Nov	0.573	0.577	0.571	0.581
	Dec	0.570	0.573	0.575	0.579
USA	Jan	0.546	0.556	0.557	0.562
	Feb	0.548	0.559	0.556	0.558
	Mar	0.547	0.556	0.551	0.561
	Apr	0.549	0.558	0.551	0.561
	May	0.548	0.553	0.553	0.557
	Jun	0.546	0.554	0.559	0.560
	Jul	0.543	0.547	0.555	0.558
	Aug	0.546	0.550	0.554	0.560
	Sep	0.543	0.548	0.556	0.559
	Oct	0.545	0.551	0.556	0.559
	Nov	0.545	0.552	0.553	0.563
	Dec	0.546	0.551	0.556	0.566

Mean 0.543 0.583

Figure 7: Monthly trends in **Dominance** of tweets across years (TUSC-Country).

Dataset	# tweets	# tweeters
2020		
Canada	3,038,530	20,887
USA	2,641,694	19,709
2021		
Canada	7,467,446	45,573
USA	11,675,372	76,223

Table 4: Number of tweets and tweeters in the TUSC100 dataset.

Data	Year	Canada	USA
Mean	2020	0.6320	0.6132
	2021	0.6387	0.6257
Variability	2020	0.0708	0.0714
	2021	0.0700	0.0705
Rise Rate	2020	0.0121	0.0128
	2021	0.0117	0.0123
Recovery Rate	2020	0.0120	0.0127
	2021	0.0118	0.0123
Hm-Hi Rate	2020	0.0118	0.0129
	2021	0.0113	0.0121
Hi-Hm Rate	2020	0.0118	0.0129
	2021	0.0115	0.0122
Hm-Lo Rate	2020	0.0143	0.0149
	2021	0.0140	0.0145
Lo-Hm Rate	2020	0.0141	0.0148
	2021	0.0139	0.0144

Table 5: Tweet **Valence** dynamics metrics of tweeters in TUSC100. Averaged across all tweeters (not considering the cities they came from).

Data	Year	Canada	USA
Mean	2020	0.4828	0.4935
	2021	0.4854	0.4932
Variability	2020	0.0599	0.0593
	2021	0.0599	0.0595
Rise Rate	2020	0.0116	0.0120
	2021	0.0113	0.0117
Recovery Rate	2020	0.0115	0.0119
	2021	0.0113	0.0116
Hm-Hi Rate	2020	0.0129	0.0130
	2021	0.0125	0.0129
Hi-Hm Rate	2020	0.0127	0.0130
	2021	0.0126	0.0128
Hm-Lo Rate	2020	0.0121	0.0127
	2021	0.0118	0.0123
Lo-Hm Rate	2020	0.0120	0.0125
	2021	0.0117	0.0121

Table 6: **Arousal** dynamics metrics of tweeters in TUSC100. Averaged across all tweeters (not considering the cities they came from).

Data	Year	Canada	USA
Mean	2020	0.5821	0.5671
	2021	0.5873	0.5733
Variability	2020	0.0573	0.0556
	2021	0.0569	0.0557
Rise Rate	2020	0.0113	0.0116
	2021	0.0109	0.0113
Recovery Rate	2020	0.0112	0.0117
	2021	0.0109	0.0112
Hm-Hi Rate	2020	0.0114	0.0118
	2021	0.0111	0.0115
Hi-Hm Rate	2020	0.0114	0.0119
	2021	0.0111	0.0114
Hm-Lo Rate	2020	0.0127	0.0129
	2021	0.0124	0.0126
Lo-Hm Rate	2020	0.0126	0.0128
	2021	0.0123	0.0125

Table 7: Tweet **Dominance** Dynamics metrics of tweeters in TUSC100. Averaged across all tweeters (not considering the cities they came from).

Country	City	valence 2020	valence 2021
Canada	Brampton	0.634	0.642
	Calgary	0.636	0.647
	Edmonton	0.643	0.649
	Etobicoke	0.637	0.643
	Halifax	0.649	0.655
	Hamilton	0.644	0.651
	Laval	0.638	0.645
	London	0.649	0.656
	Mississauga	0.635	0.643
	Montreal	0.634	0.644
	NorthYork	0.634	0.639
	Okanagan	0.637	0.642
	Ottawa	0.651	0.656
	Quebec	0.638	0.647
	Scarborough	0.632	0.633
	Surrey	0.641	0.648
	Toronto	0.634	0.640
	Vancouver	0.638	0.645
	Victoria	0.649	0.655
	Windsor	0.613	0.623
	Winnipeg	0.644	0.650
USA	Austin	0.627	0.639
	Boston	0.625	0.641
	Charlotte	0.627	0.638
	Chicago	0.617	0.629
	Columbus	0.625	0.637
	Dallas	0.617	0.630
	Denver	0.624	0.636
	Detroit	0.614	0.626
	ElPaso	0.629	0.638
	FortWorth	0.620	0.630
	Houston	0.613	0.620
	Indianapolis	0.633	0.644
	Jacksonville	0.626	0.638
	LosAngeles	0.615	0.630
	Memphis	0.621	0.630
	Nashville	0.631	0.644
	NewYork	0.616	0.630
	Philadelphia	0.614	0.627
	Phoenix	0.617	0.633
	SanAntonio	0.624	0.634
	SanDiego	0.620	0.636
	SanFrancisco	0.626	0.640
	SanJose	0.630	0.645
	Seattle	0.628	0.639
	Washington	0.626	0.635
	Mean		

Country	City	arousal 2020	arousal 2021
Canada	Brampton	0.483	0.485
	Calgary	0.483	0.480
	Edmonton	0.478	0.480
	Etobicoke	0.485	0.485
	Halifax	0.472	0.473
	Hamilton	0.482	0.481
	Laval	0.488	0.488
	London	0.477	0.477
	Mississauga	0.485	0.483
	Montreal	0.488	0.489
	NorthYork	0.483	0.485
	Okanagan	0.478	0.478
	Ottawa	0.476	0.477
	Quebec	0.494	0.492
	Scarborough	0.485	0.487
	Surrey	0.480	0.480
	Toronto	0.484	0.485
	Vancouver	0.480	0.480
	Victoria	0.475	0.474
	Windsor	0.497	0.496
	Winnipeg	0.479	0.480
USA	Austin	0.489	0.489
	Boston	0.488	0.486
	Charlotte	0.490	0.493
	Chicago	0.496	0.493
	Columbus	0.492	0.491
	Dallas	0.498	0.495
	Denver	0.491	0.490
	Detroit	0.497	0.495
	ElPaso	0.502	0.504
	FortWorth	0.496	0.496
	Houston	0.498	0.499
	Indianapolis	0.492	0.490
	Jacksonville	0.492	0.492
	LosAngeles	0.498	0.497
	Memphis	0.493	0.494
	Nashville	0.486	0.488
	NewYork	0.494	0.493
	Philadelphia	0.493	0.494
	Phoenix	0.499	0.497
	SanAntonio	0.496	0.499
	SanDiego	0.496	0.495
	SanFrancisco	0.484	0.484
	SanJose	0.483	0.483
	Seattle	0.485	0.483
	Washington	0.491	0.487
	Mean		

Country	City	dominance 2020	dominance 2021
Canada	Brampton	0.593	0.599
	Calgary	0.599	0.604
	Edmonton	0.594	0.598
	Etobicoke	0.595	0.600
	Halifax	0.591	0.595
	Hamilton	0.599	0.605
	Laval	0.593	0.599
	London	0.597	0.604
	Mississauga	0.594	0.601
	Montreal	0.592	0.599
	NorthYork	0.593	0.599
	Okanagan	0.588	0.590
	Ottawa	0.605	0.610
	Quebec	0.592	0.598
	Scarborough	0.594	0.599
	Surrey	0.597	0.602
	Toronto	0.594	0.600
	Vancouver	0.595	0.601
	Victoria	0.606	0.610
	Windsor	0.576	0.579
	Winnipeg	0.591	0.595
USA	Austin	0.587	0.594
	Boston	0.592	0.599
	Charlotte	0.585	0.590
	Chicago	0.580	0.584
	Columbus	0.583	0.589
	Dallas	0.579	0.584
	Denver	0.589	0.593
	Detroit	0.577	0.581
	ElPaso	0.570	0.578
	FortWorth	0.577	0.583
	Houston	0.570	0.573
	Indianapolis	0.589	0.594
	Jacksonville	0.590	0.597
	LosAngeles	0.579	0.585
	Memphis	0.578	0.582
	Nashville	0.590	0.596
	NewYork	0.584	0.591
	Philadelphia	0.576	0.581
	Phoenix	0.585	0.591
	SanAntonio	0.571	0.579
	SanDiego	0.584	0.592
	SanFrancisco	0.593	0.601
	SanJose	0.595	0.605
	Seattle	0.591	0.596
	Washington	0.597	0.603
	Mean		

Figure 8: TED: Tweet Valence Means (left), Arousal Means (centre), and Dominance Means (right) across American and Canadian cities in 2020 and 2021 (using tweets from TUSC-City).

Country	City	valence 2020	valence 2021
Canada	Brampton	0.0788	0.0776
	Calgary	0.0793	0.0785
	Edmonton	0.0787	0.0773
	Etobicoke	0.0803	0.0776
	Halifax	0.0780	0.0766
	Hamilton	0.0802	0.0785
	Laval	0.0788	0.0768
	London	0.0790	0.0777
	Mississauga	0.0789	0.0774
	Montreal	0.0790	0.0776
	NorthYork	0.0791	0.0778
	Okanagan	0.0809	0.0785
	Ottawa	0.0785	0.0770
	Quebec	0.0800	0.0783
	Scarborough	0.0796	0.0787
	Surrey	0.0785	0.0772
	Toronto	0.0792	0.0777
	Vancouver	0.0796	0.0773
	Victoria	0.0785	0.0770
	Windsor	0.0820	0.0803
	Winnipeg	0.0787	0.0772
USA	Austin	0.0795	0.0777
	Boston	0.0801	0.0776
	Charlotte	0.0809	0.0787
	Chicago	0.0816	0.0788
	Columbus	0.0806	0.0789
	Dallas	0.0815	0.0791
	Denver	0.0803	0.0782
	Detroit	0.0818	0.0801
	ElPaso	0.0819	0.0798
	FortWorth	0.0808	0.0791
	Houston	0.0816	0.0804
	Indianapolis	0.0809	0.0782
	Jacksonville	0.0813	0.0795
	LosAngeles	0.0812	0.0791
	Memphis	0.0814	0.0799
	Nashville	0.0798	0.0779
	NewYork	0.0810	0.0788
	Philadelphia	0.0810	0.0797
	Phoenix	0.0827	0.0793
	SanAntonio	0.0816	0.0794
	SanDiego	0.0815	0.0786
	SanFrancisco	0.0798	0.0774
	SanJose	0.0796	0.0772
	Seattle	0.0820	0.0788
	Washington	0.0848	0.0787

Variability

Country	City	valence 2020	valence 2021
Canada	Brampton	0.0107	0.0105
	Calgary	0.0107	0.0104
	Edmonton	0.0105	0.0104
	Etobicoke	0.0106	0.0105
	Halifax	0.0104	0.0102
	Hamilton	0.0106	0.0104
	Laval	0.0107	0.0105
	London	0.0104	0.0102
	Mississauga	0.0107	0.0105
	Montreal	0.0108	0.0105
	NorthYork	0.0106	0.0105
	Okanagan	0.0107	0.0109
	Ottawa	0.0103	0.0103
	Quebec	0.0108	0.0106
	Scarborough	0.0107	0.0106
	Surrey	0.0107	0.0104
	Toronto	0.0107	0.0105
	Vancouver	0.0105	0.0103
	Victoria	0.0105	0.0103
	Windsor	0.0112	0.0108
	Winnipeg	0.0105	0.0103
USA	Austin	0.0108	0.0105
	Boston	0.0109	0.0104
	Charlotte	0.0108	0.0105
	Chicago	0.0109	0.0107
	Columbus	0.0107	0.0105
	Dallas	0.0110	0.0107
	Denver	0.0109	0.0105
	Detroit	0.0110	0.0108
	ElPaso	0.0110	0.0108
	FortWorth	0.0110	0.0108
	Houston	0.0112	0.0110
	Indianapolis	0.0107	0.0104
	Jacksonville	0.0109	0.0106
	LosAngeles	0.0111	0.0108
	Memphis	0.0110	0.0107
	Nashville	0.0108	0.0104
	NewYork	0.0111	0.0107
	Philadelphia	0.0110	0.0107
	Phoenix	0.0111	0.0107
	SanAntonio	0.0110	0.0108
	SanDiego	0.0110	0.0106
	SanFrancisco	0.0108	0.0104
	SanJose	0.0107	0.0103
	Seattle	0.0108	0.0106
	Washington	0.0108	0.0106

Rise Rate

Country	City	valence 2020	valence 2021
Canada	Brampton	0.0107	0.0104
	Calgary	0.0108	0.0104
	Edmonton	0.0105	0.0103
	Etobicoke	0.0105	0.0104
	Halifax	0.0104	0.0102
	Hamilton	0.0105	0.0104
	Laval	0.0107	0.0105
	London	0.0104	0.0101
	Mississauga	0.0107	0.0104
	Montreal	0.0107	0.0104
	NorthYork	0.0106	0.0105
	Okanagan	0.0107	0.0104
	Ottawa	0.0103	0.0102
	Quebec	0.0108	0.0106
	Scarborough	0.0107	0.0107
	Surrey	0.0105	0.0104
	Toronto	0.0107	0.0104
	Vancouver	0.0106	0.0104
	Victoria	0.0103	0.0103
	Windsor	0.0110	0.0108
	Winnipeg	0.0105	0.0103
USA	Austin	0.0108	0.0105
	Boston	0.0107	0.0104
	Charlotte	0.0107	0.0105
	Chicago	0.0109	0.0107
	Columbus	0.0108	0.0105
	Dallas	0.0110	0.0107
	Denver	0.0108	0.0105
	Detroit	0.0110	0.0108
	ElPaso	0.0110	0.0107
	FortWorth	0.0109	0.0107
	Houston	0.0111	0.0109
	Indianapolis	0.0107	0.0104
	Jacksonville	0.0109	0.0106
	LosAngeles	0.0111	0.0108
	Memphis	0.0109	0.0107
	Nashville	0.0107	0.0104
	NewYork	0.0110	0.0107
	Philadelphia	0.0110	0.0107
	Phoenix	0.0110	0.0107
	SanAntonio	0.0109	0.0107
	SanDiego	0.0110	0.0106
	SanFrancisco	0.0107	0.0104
	SanJose	0.0106	0.0104
	Seattle	0.0109	0.0106
	Washington	0.0107	0.0106

Recovery Rt.

Figure 9: TED: Tweet Valence Variability (left), Rise Rate (centre), and Recovery Rate (right) across American and Canadian cities in 2020 and 2021 (using tweets from TUSC-City).