The Language of Food during the Pandemic: 
Hints about the Dietary Effects of Covid-19

Hoang Van, Ahmad Musa, Mihai Surdeanu, Stephen Kobourov
Computer Science Department, University of Arizona
{vnhh,ahmadmusa,msurdeanu,kstephen}@email.arizona.edu

Abstract

We study the language of food on Twitter during the pandemic lockdown in the United States, focusing on the two month period of March 15 to May 15, 2020. Specifically, we analyze over 770,000 tweets published during the lockdown and the equivalent period in the five previous years and highlight several worrying trends. First, we observe that during the lockdown there was a notable shift from mentions of healthy foods to unhealthy foods. Second, we show an increased pointwise mutual information of depression hashtags with food-related tweets posted during the lockdown and an increased association between depression hashtags and unhealthy foods, tobacco, and alcohol during the lockdown.

1 Introduction

The severe acute respiratory syndrome coronavirus (SARS-CoV) emerged in late 2002, and caused an outbreak of severe acute respiratory syndrome (SARS) (Petersen et al., 2020). The SARS-CoV-2 virus, which causes the Coronavirus Disease 2019 (Covid-19), is closely related to SARS-CoV. On March 11, 2020, World Health Organization (WHO) declared Covid-19 as a pandemic (WHO, 2020). Covid-19 has claimed 682,855 lives worldwide as of August 1, 2020¹.

While the death toll of the disease is horrific, there are other indirect public health effects of Covid-19 that are longer term and harder to measure. For example, the disease has already caused many countries to enter a phase of recession (Sułkowski, 2020), which is likely to also affect individual well being. Community anxiety can rise abruptly during a pandemic like Covid-19 (Lima et al., 2020). This paper focuses on the early detection of some of these indirect effects using Twitter.

Like other social media platforms, Twitter² can potentially serve as a valuable information resource for various public health applications. Among others, the paper (Bell et al., 2018) proposed a method to detect individuals who are at risk of getting type-2 diabetes by analyzing their tweets. The predictive power of the language of food on Twitter and analyzed changes in this language over time are shown in (Van et al., 2019). (De Choudhury et al., 2013) showed that social media posts contain useful signals that can predict the rise of depression in individuals. Symptoms associated with mental illness are observable on Twitter, Facebook, and web forums (Guntuku et al., 2017).

Our work compares the language of food on Twitter during the lockdown period in the U.S. (March 15 to May 15, 2020) against the equivalent months in the five previous years. Our analysis highlights worrying trends in dietary patterns and mental health during this interval, which are, thus, likely to be indirect effects of the disease and the associated lockdown. In particular, the contributions of this work are two-fold:

1. Using a dataset of 772,142 tweets and a lexicon of 800 good phrases, we show that during the lockdown period there was a significant change in the language of food on the U.S. Twitter. In particular, we observe a shift from mentions of healthy foods to unhealthy foods on the U.S. Twitter, compared to previous years. This trend is worrying because previous works showed that the consumption of unhealthy food leads to obesity (Karnik and Kanekar, 2012; Chopra and Darnton-Hill, 2004). Further, there is strong evidence that

¹https://en.wikipedia.org/wiki/Template:COVID-19_pandemic_data
²https://twitter.com/
| Term    | # of tweets | # of tweets localized in U.S. |
|---------|-------------|------------------------------|
| #dinner | 6,007,037   | 1,520,573                    |
| #breakfast | 5,736,525   | 1,335,432                    |
| #lunch  | 5,514,638   | 1,213,171                    |
| #brunch | 2,160,015   | 768,577                      |
| #snack  | 898,178     | 246,842                      |
| #meal   | 553,214     | 115,101                      |
| #supper | 138,658     | 24,860                       |
| Total   | 28,854,023  | 5,198,387                    |

Table 1: Seven meal related hashtags and corresponding number of tweets in the complete tweet dataset, from October 2013 to May 2020. The right-most column indicates the number of tweets we could localize to a U.S. state or Washington D.C.

Obesity is a significant comorbidity factor for complications from the corona virus (Stefan et al., 2020). Our result contradicts the general belief that the lockdown led to better diets due to an increase in home cooking, and evidence from other countries such as Italy (Di Renzo et al., 2020).

2. We analyze the incidence of depression during the lockdown interval using a set depression hashtags such as #depression and #suicidalthoughts. Our analysis shows a considerable increase in pointwise mutual information (PMI) of these depression hashtags and tweets coming from the lockdown interval vs. equivalent months in the previous years. Further, we observe an increased association between depression hashtags and unhealthy foods during lockdown compared to previous years. This analysis suggests that not only was there an increase in depression during the lockdown, but also that this increase led to a larger negative impact on dietary patterns than before the Covid-19 pandemic.

2 Data

The dataset we collected focuses on food-related tweets. We used Twitter’s public streaming API to collect tweets and their metadata such as geolocation and temporal information. We filtered the tweets using a list of seven meal related hashtags (Table 1). Tweets have been stored into a Lucene-backed Solr instance, which was used to localize the tweets within the U.S. In total, we have collected 28 million tweets from the period between October 2, 2013 and May 25, 2020. Similar to (Van et al., 2019), we localized 5 million of these tweets to a U.S. location using either the geo tag associated with the tweet, or the user’s self-reported location.

From the above dataset, we extracted tweets that were localized in the U.S., contained a meal-related hashtag, and were posted between March and May from year 2015 to 2020. This smaller dataset is summarized in Table 2. This is the dataset we used for the analyses reported in this paper. Table 2 indicates that the number of meal-related tweets decreased from 185K in March–May 2015 to only 63K in the same period in 2020. This decrease may be explained by multiple overlapping factors: (a) the overall number of tweets is decreasing; (b) the number of tweets that are geo tagged is decreasing due to an increasing user focus on privacy; and (c) lastly, we suspect that the meal-focused social media is migrating to platforms that are more multi-modal friendly such as Instagram. Nevertheless, the overall counts are sufficient to draw reliable conclusions.

---

3See, for example, the Google trend for the word “sourdough”: https://trends.google.com/trends/explore?q=sourdough&geo=US.
4https://developer.twitter.com/en/docs/tweets/filter-realtime/guides/connecting.html
5https://lucene.apache.org/. Solr is the open-source NoSQL search platform from the Apache Lucene project.
6https://www.businessinsider.com/tweets-on-twitter-is-in-serious-decline-2016-2
7https://www.forbes.com/sites/kalevleetaru/2019/03/04/visualizing-seven-years-of-tweeters-evolution-2012-2018/
Table 2: Number of tweets containing meal-related hashtags, which were localized in the U.S. and were posted in the March–May interval during the six years analyzed.

| Year (March-May) | # of tweets localized in U.S. |
|------------------|-------------------------------|
| 2020             | 63,898                        |
| 2019             | 81,036                        |
| 2018             | 131,418                       |
| 2017             | 135,734                       |
| 2016             | 174,792                       |
| 2015             | 185,264                       |

3 Approach

The two main aims of this work are: (a) to explore the change in dietary patterns mentioned in these tweets during the Covid-19 lockdown period, and (b) to identify the change in mental health trends in the same period, and their associations with diet. We discuss the settings of these experiments next.

3.1 Food trends in tweets

We investigate dietary trends (as mentioned in tweets) through three distinct analyses that range from nation-level, to region-level, to city-level. For city-level data, we analyze the changes in food trends in New York City, Chicago, and Los Angeles, which are the top three cities most affected by the pandemic (Dong et al., 2020). We use the tweets from the above dataset as the only input for our analyses. To show the effect of the Covid-19 pandemic lockdown, we compare tweets in the interval March 15 to May 15 between the years 2015 to 2020.

Nation-level food trends: For each tweet localized in the U.S., we count the number of matches of healthy and unhealthy food words from the dictionary included in (Van et al., 2019). This dictionary contains a food lexicon, with foods organized in three categories: healthy, unhealthy, and neutral. Table 4
shows several examples from this dictionary. We normalize the food word counts collected from tweets by the overall number of food words observed per year.

| Healthy          | peach, blueberry, succotash, whey, acorn, cabbage, mush- room, beans, beetroot, banana |
|------------------|-----------------------------------------------------------------------------------|
| Unhealthy        | quesadilla, ham, wine, beer, cake, soda, liquor, milkshake, hamburger, cheese       |
| Neutral          | crunch, ketchup, drink, mussels, fish, meat, soup, salt, sandwich, protein          |

Table 4: Examples for each food category in the food vocabulary in (Van et al., 2019).

**Region-level food trends:** We use the region division by the U.S. Census Bureau, which identifies the following 4 regions: Northeast, Midwest, South, and West. We analyze the data in a way similar to that for the nation-level, but focusing on each of the regions individually. We count mentions of healthy/unhealthy food words using the dictionary of (Van et al., 2019), and normalize these counts by the total number of food words in each region.

**City-level food trends:** We analyze the food trends in New York City, Chicago, and Los Angeles, which are the top three cities most affected by the Covid-19 lockdown (Dong et al., 2020). Again, we use the healthy/unhealthy food word dictionary, and normalize counts by the total count of food words in each city.

### 3.2 Mental health effects of the lockdown

Previous work has shown that the Covid-19 pandemic lockdown can worsen mental health and cause mental problems such as anxiety and depression (Kumar and Nayar, 2020; Torales et al., 2020; Gualano et al., 2020). To complement this work, we analyze mentions of hashtags related to mental health problems in tweets during the Covid-19 lockdown, compared against the same period of the previous 5 years. We provide three analyses for this goal: trends for depression-related hashtags, associations between these depression-related hashtags and food words, and the top depression-related food choices over time.

**Trends of depression-related hashtags:** To analyze these trends over time, we measure the pointwise mutual information (PMI) between depression hashtags (see Table 3) and tweets in the time period corresponding to the lockdown vs. the same period in previous years. In particular, we divide our corpus into six parts, each containing the complete set of tweets between March 15 and May 15 of the corresponding year (2015–2020), and then calculate the PMI for pairs (depression hashtags, year $y$) using the formula:

$$PMI(t, y) = \frac{C(t, y)}{C(t) \cdot C(y)}$$

(1)

where $C(t, y)$ is the number of tweets with depression hashtags in the period corresponding to year $y$, $C(t)$ is the total number of tweets with depression hashtags across all years, and $C(y)$ is the total number
| # | Year | U.S.  | Northeast | Midwest | South | West  |
|---|------|-------|-----------|---------|-------|-------|
| 1 | 2020 | 70.23 | 70.67     | 71.38   | 67.50 | 70.17 |
| 2 | 2019 | 60.66 | 59.65     | 65.35   | 60.52 | 58.99 |
| 3 | 2018 | 62.04 | 61.71     | 66.99   | 62.53 | 61.13 |
| 4 | 2017 | 65.31 | 64.32     | 68.55   | 65.34 | 64.91 |
| 5 | 2016 | 62.41 | 62.00     | 63.86   | 62.77 | 60.83 |
| 6 | 2015 | 61.48 | 60.17     | 62.93   | 61.01 | 59.36 |

Table 5: Percentage of unhealthy foods mentioned in tweets during the interval March 15 – May 15 in each year analyzed. The U.S. column lists the percentage of mentions of unhealthy foods nation wide. The Northwest, Midwest, South, and West columns list the percentages of mentions of unhealthy foods for the four U.S. regions.

of tweets in the period corresponding to year $y$. Intuitively, the higher the PMI value of a term in a given year, $PMI(t, y)$ the more that term is associated with tweets from that year in particular.

**Depression-related foods trends:** For this analysis, we investigate only the U.S. localized tweets that contain depression hashtags. We compute PMI values between food words and tweets that contain depression hashtags (again, considering just the March 15 – May 15 interval) to analyze the association between healthy/unhealthy foods and depression over time. The PMI for all pairs (food $t$, year $y$) is calculated using formula 1, where $C(t, y)$ is the number of depression-related tweets from year $y$ containing food $t$, $C(t)$ is the total number of depression-related tweets with food $t$ across all years, and $C(y)$ is the total number of depression-related tweets in year $y$.

**Depression-related food choices:** We investigate only the U.S. localized tweets that contain depression hashtags, and rely on the PMI values computed in the previous step. Here, we identify the list of foods with largest changes in PMI from previous years. We divide our corpus into two parts, the first part contains tweets in the March–May period in 2020 and the second part contains tweets in the same period in all previous years in our dataset (2015–2019). We then calculate the differences in PMI for each food words between two partitions, and report foods with the largest differences.

## 4 Results and Discussion

Next we discuss the results of the analyses introduced in the previous section.

### 4.1 An unhealthy shift in dietary trends

**Nation-level food trend:** Table 5 lists the percentage of unhealthy foods mentioned in tweets (normalized by the total number of food words in the corresponding interval). The table shows that the largest increase at the nation level in unhealthy food mentions from the previous year occurred in 2020, during the Covid-19 pandemic lockdown (a relative increase of 15.77%). The lockdown period in 2020 had the largest total percentage of unhealthy food mentions, at 70.2%.

Interestingly, the second highest percentage of mentions of unhealthy foods in the March–May period occurred in 2017 (a total of 65.3% with a relative increase of 4.65% from 2016) when the U.S. experienced a seismic shift in national politics. The next two years after 2017 saw a downward trend in mentions of unhealthy foods, but this trend was abruptly reversed in 2020.

This analysis suggests that the lockdown period in 2020 had a notably negative effect on American diets. This is a concerning observation, as dietary patterns are in important indicators of public health (Gorski and Roberto, 2015).

To further highlight this negative trend, we also investigated the number of mentions for delivery service hashtags over time. This analysis is motivated by the strong association between online food delivery services and unhealthy eating habits (Cetateanu and Jones, 2014). Figure 1 shows an upward trend for
Figure 2: A series of U.S. regional maps, where the (a)–(f) maps plot the years 2015 to 2020, respectively. Regions with higher percentage of unhealthy food tweets (posted in the March–May period each year) have darker red color. The year 2020 shows a clear peak over the previous five years.

Figure 3: PMI values for depression hashtags over the years.

delivery service hashtags over the six years investigated. (Intuitively, the higher the PMI value of a term in a given time period the more that term is associated with tweets from that period.) However, the PMI score for delivery service hashtags during the Covid-19 pandemic lockdown increased drastically more than the trend observed in previous years, more than doubling its value compared to the same period in 2019.

Region-level food trends: The Northeast and West regions of the U.S. were the most affected by the Covid-19 pandemic (Dong et al., 2020). The two regions were the first to experience state-wide lockdowns. Table 5 reflects these impacts by the Covid-19 lockdown on the regional level. The Northeast and West
Figure 4: PMI values for each food category (healthy and unhealthy) over the six years, for depression-related tweets.

|    | New York | Chicago | Los Angeles |
|----|----------|---------|-------------|
| 1  | 2020     | 72.01   | 73.27       | 71.68       |
| 2  | 2019     | 61.61   | 64.39       | 61.02       |
| 3  | 2018     | 63.56   | 66.74       | 62.37       |
| 4  | 2017     | 65.74   | 67.22       | 63.45       |
| 5  | 2016     | 63.51   | 66.69       | 63.86       |
| 6  | 2015     | 61.08   | 64.17       | 63.22       |

Table 6: Percentage of unhealthy foods mentioned in tweets in New York City, Chicago, and Los Angeles over the six years analyzed.

regions have the largest relative increases in unhealthy foods, with 18.47% and Hoang 18.95%, respectively. These increases are well above the increase in unhealthy food mentions at country level. The Midwest and South regions also experience large relative increases in mentions of unhealthy foods (9.23% and 11.53%) from previous years, but these increases are below the national average. Figure 2 summarizes these trends visually, highlighting the worrying increases in 2020. These sharp increases are worrying because the US already has very high obesity rates (39.8% of adults aged 20 and over are obese and that another 31.8% were overweight (Fryar et al., 2018)). Further, there is strong evidence that obesity is a significant comorbidity factor for complications from the corona virus (Stefan et al., 2020).

City-level food trends: To conclude the impacts of the Covid-19 pandemic lockdown on dietary habits, we also investigate the unhealthy food trends in tweets in the three most affected cities by the Covid-19 (based on both number of cases and deaths) in the U.S.: New York City, Chicago, and Los Angeles (Dong et al., 2020). Table 6 shows the prevalence of unhealthy foods in the three cities. It is notable that the number of mentions of unhealthy foods in tweets in these three cities is well above both the national average as well as the average increase in the regions to which they belong.
4.2 Mental health effects of the Covid-19

Depression trend: Figure 3 plots the depression PMI values for the 2015–2020 period. In general, the higher the PMI value, the stronger the association between depression hashtags and tweets from that period is. The figure shows that there is an upward trend for depression PMI values from 2015 to 2020, with a sharp increase during the Covid-19 pandemic lockdown in 2020. Further, the PMI score for depression hashtags in 2020 is the highest across the six periods investigated. This analysis suggests that the Covid-19 lockdown had a negative impact on the mental health of people in the U.S. Our observation aligns with other measurements of depression during the lockdown, which indicated that depression increased during the lockdown (Gualano et al., 2020; Fullana et al., 2020; Smith et al., 2020).

Table 7: Top food words associated with depression hashtags for the March 15–May 15 period over six years, according to their PMI values. The words are listed in descending order of PMI values from left to right.

| Year | Food Words |
|------|------------|
| 2020 | wine, beer, cigarette, quesadilla, liquor, cocktail, succotash, milkshake, blueberry, citron |
| 2019 | quesadilla, peach, prune, blueberry, succotash, whey, acorn, ham, burrito, mushroom |
| 2018 | quesadilla, tomatillo, oyster, yolks, seafood, liquor, lychee, pate, broccoli, rum |
| 2017 | scallion, quesadilla, nectarine, liquor, beer, pancake, brandy, sirloin, ham, dragonfruit |
| 2016 | quesadilla, tomatillo, nectarine, maize, pumpkin, ham, soda, coriander, tequila, guacamole |
| 2015 | lychee, dragonfruit, quesadilla, tequila, coriander, endive, lard, daikon, guacamole, seaweed |

Depression-related foods trends: We next investigate whether depression is associated with mentions of healthy or unhealthy foods in our data. Figure 4 plots the PMI values of healthy and unhealthy foods over the six time periods analyzed, with respect to tweets that contain at least one depression hashtag. Intuitively, these values indicate the association strength between depression hashtags and healthy/unhealthy food words. Between 2015 and 2019 there are no clear differences between the PMIs of healthy and unhealthy foods. However, during the Covid-19 lockdown, the PMI of unhealthy food words increases sharply, reaching the highest value measured in our dataset. Further, this value is 78% higher than the PMI of healthy foods in the same period. This suggests that during the pandemic lockdown not only did the incidence of depression increase, but also that it is associated with an increased consumption of unhealthy foods.

Depression-related food choices: To further analyze the association between depression and food, we extracted the food words with highest PMI values in depression-related tweets over the six periods investigated. Table 7 shows the top 10 food words per year. Unsurprisingly, the majority of these top 10 food words are unhealthy. In years when the depression PMI peaks (2018 and 2020), unhealthy foods (i.e., liquor, beer, cigarette, rum, wine) are among the top food choices. This is especially clear during the Covid-19 pandemic, when five out of the top 10 food words are related to alcohol and cigarettes. Several examples of such tweets, which contain mentions of depression and dietary items, are shown in Table 8.

Lastly, Table 9 shows the top 20 food words that had the largest increase in PMI relative to depression-related tweets in 2020, compared to all previous years. The top foods in the table are cigarette, wine, vodka, and whiskey, which further confirms a concerning increase in the association between depression hashtags and unhealthy dietary patterns in social media.

5 Conclusion

In this work we compared the language of food during the pandemic lockdown period in the United States against the same period in five previous years. Our analysis indicates that during the lockdown period there was a considerable shift towards mentions of unhealthy foods compared to previous years. Further, we showed that there was a considerable increase in PMI between depression hashtags and tweets posted during the lockdown. Lastly, we highlighted an increased association between depression hashtags and unhealthy foods, alcohol and cigarettes.
During these stressful times, my breakfast is NOW: coffee, cigarettes, AND beer. I work from home.

cigarettes depression breakfast https://t.co/3mJVHXD
cigarettes is a new dinner Covid19 mentalhealthmatters
mentalhealthsupport weed or cigarettes no dinner or breakfast
RT @GPasadena: so pretty! healing depression Paadena Madeline_Garden winetasting champagnebrunch brunch sundaybrunch breakfast

cigarettes is a new dinner Covid19 mentalhealthmatters
mentalhealthsupport weed or cigarettes no dinner or breakfast

Table 8: Examples of tweets containing depression hashtags in two periods: the March–May interval in 2020 and the same interval in the five previous years, highlighting a worrying attitude change in depression-related tweets towards unhealthy foods, tobacco, and alcohol.

cigarette, wine, vodka, whiskey, tarragon, citron, beer, batter, cocktail, brisket, grapefruit, liquor, fava, jicama, shallots, lamb, lentils, flan, squid, tripe

Table 9: Top 20 food words with largest changes in PMI relative to depression-related tweets in 2020 compared to the previous five years. The words are listed in descending order of this value, from left to right.

While all these results are worrying, we did not prove yet that this social media analysis correlates with real life information. We leave this parallel analysis to future work. However, given the previous work that showed that real life public health information can be forecast from social media (Van et al., 2019; Bell et al., 2018; Guntuku et al., 2017; De Choudhury et al., 2013), we envision that our analysis can be used as a near real-time monitoring tool for the rapid identification of important public health factors such as diet and mental health.

References

Dane Bell, Mihai Surdeanu, and Stephen Kobourov. 2018. Detecting diabetes risk from social media activity. In Ninth International Workshop on Health Text Mining and Information Analysis (LOUHI).

Andreea Cetateanu and Andy Jones. 2014. Understanding the relationship between food environments, deprivation and childhood overweight and obesity: evidence from a cross sectional england-wide study. Health & place, 27:68–76.

Mickey Chopra and Ian Darnton-Hill. 2004. Tobacco and obesity epidemics: not so different after all? Bmj, 328(7455):1558–1560.

Munmun De Choudhury, Michael Gamon, Scott Counts, and Eric Horvitz. 2013. Predicting depression via social media. Icwsm, 13:1–10.

L. Di Renzo, Gualtieri P., and Pivari F. et al. 2020. Eating habits and lifestyle changes during covid-19 lockdown: an italian survey. Journal of Translational Medicine, 18(1).
