

Natural vs. Artificially Sweet Tweets: Characterizing Discussions of Non-Nutritive Sweeteners on Twitter

Hande Batan¹, Dianna Radpour¹, Ariane Kehlbacher², Judith
Klein-Seetharaman³, and Michael J. Paul¹

¹ University of Colorado Boulder, Boulder, CO USA
[firstname].[lastname]@colorado.edu

² University of Reading, Reading, Berkshire, UK

³ Colorado School of Mines, Golden, CO, USA

Abstract. This ongoing project aims to use social media data to study consumer behaviors regarding natural and artificial sweeteners. Following the recent shifts to natural sweeteners such as Stevia versus artificial, and traditionally-used ones like aspartame in recent years, there has been discussion around potential negative side effects, including memory loss and other chronic illnesses. These issues are discussed on Twitter, and we hypothesize that Twitter may provide insights into how people make nutritional decisions about the safety of sweeteners given the inconclusive science surrounding the topic, how factors such as risk and consumer attitude are interrelated, and how information and misinformation about food safety is shared on social media. As an initial step, we describe a new dataset containing 308,738 de-duplicated English-language tweets spanning multiple years. We conduct a topic model analysis and characterize tweet volumes over time, showing a diversity of sweetener-related content and discussion. Our findings suggest a variety of research questions that these data may support.

Keywords: Social media · Public health · Nutrition

1 Introduction

As the general public becomes increasingly aware of the possible associations between artificial sweeteners and various health concerns, natural alternatives have seen a surge in popularity and general usage. However, with the science around both artificial and natural sweeteners being largely inconclusive, data from social media platforms such as Twitter presents itself as a valuable resource for identifying the factors and information that influence the decisions consumers make in choosing what sweeteners to use or avoid.

1.1 Background: New Sweeteners

The law mandates that every new substance that gets introduced into the market for direct human consumption must be approved as safe for entering the

human body. Though the consequences for an incorrectly assigned designation could lead to serious health repercussions, the designations are rarely questioned or altered once they are granted, even if that designation is founded in science and research studies from decades ago. This is precisely the phenomenon we are observing today with artificial and low-calorie sweeteners, such as aspartame, sucralose, and saccharin. Though many recent studies suggest that these artificial sweeteners might be linked to a number of negative side effects, including memory loss, cancer, and other chronic illnesses, their regulatory statuses remain unchanged, even with approvals dating as far back as 1981 (aspartame) and 1998 (sucralose). Despite their unchanged regulatory statuses, we have started to observe a shift in widespread consumption of artificial sweeteners to natural alternatives.

1.2 Stevia as a Natural Alternative

Stevia rebaudiana, commonly known as Stevia has, in recent years, become more widely used and drastically grown in market popularity as an everyday sweetener [11]. Stevia is a South American plant native to Paraguay that traditionally has been used as a sugar substitute in tea and other beverages and its safety has been approved by some medical, scientific and regulatory authorities, as well as some countries worldwide, including the World Health Organization, the European Food Safety Authority and the Joint Expert Committee on Food Additives.

While stevia has been approved in over 60 countries and the Food and Drug Administration (FDA) has not questioned the Generally Recognized as Safe (GRAS) status of some specific high-purity steviol glycosides for use in food, stevia leaf and the crude stevia extracts are still not considered GRAS and do not have approval for use in food. While studies, including human studies on safety, metabolism and intake, support their safety, they still await approval from the FDA. In the context of the open marketplace, new substances can be seamlessly integrated into recommended product search results, even when the advertised substances growing in popularity are questionable, with the science around their safety still largely inconclusive. These substances often take the form of supplements, alternative medicines, and questionably derived teas that consumers are taking in large and frequent quantities. While some of these products simply leverage the power of the placebo effect more than any real health effects, the fact that they often lack the basic seal of safety and approval can result in often serious health repercussions.

1.3 The Role of Social Media

With Twitter being one of the world's most prominent social media and microblogging platforms, attracting an approximate 126 million daily users, its potential as a resource in the realm of public health and surveillance continues to prove valuable. However, with its vast influence on millions of users comes the inevitable drawbacks, one being the ability for unreliable health information to spread and lead to the mass consumption of substances that people are exposed

to only on Twitter. When faced with conflicting information, how do people make choices about what to consume? Our goal is to use Twitter data to help answer this question in the domain of sweeteners, which are commonly used but also controversial.

Related research has examined social media platforms like Twitter to understand a variety of issues related to health, most commonly for disease forecasting [10]. In the areas of diet and nutrition, multiple studies have looked at mentions of food consumption in Twitter [5, 1, 8], Instagram [7, 12], and search query logs [13, 2, 6]. [4] analyzed food content on Instagram to study how food consumption is related to the availability of food in different locations. Related work has studied weight loss advice posted on social media [9]. To our knowledge, prior work has not specifically examined online discussion of sweeteners.

2 Data

From June 2017 to September 2018, approximately 851k tweets were collected from the Twitter search API matching four relevant terms: stevia, sucralose, aspartame, sweetener. We removed duplicates and non-English tweets. After this filtering step, approximately 309k sweetener-related tweets remained.

We were additionally interested in specifically analyzing tweets related to certain diseases: cancer, diabetes, and Lyme disease. These diseases have been discussed in the context of natural and artificial sweeteners, with cancer and diabetes being linked to the use of artificial sweeteners and claims being made for stevia as a potential cure for Lyme disease.

Data	# Tweets	# Users
Raw with Duplicates	851319	368503
Filtered without duplicates	308738	154585
Cancer	607	516
Diabetes	1443	833
Lyme	96	69

Table 1. Dataset statistics. The bottom three rows are subsampled from the ‘filtered without duplicates’ set, keyword-filtered for the corresponding disease.

3 Topic Analysis

Latent Dirichlet Allocation (LDA) [3], a probabilistic topic model, was used to infer topics relevant to the main issues surrounding sweeteners. The output of the model contained 100 “topics,” which are clusters of words, with some words in other languages. While the data was filtered for English only, Twitter’s language identifier is not always accurate and some tweets contained a mix of multiple languages which is the probable cause behind the multilingual topics.

ID	Relevant Topics
7	benefits, sugar, health, #stevia, natural, safe, effects, plant, stevia,
12	diet, coke, taste, aspartame, life, sucralose, make, pepsi, salt,
40	artificial, sweeteners, sugar, sucralose, aspartame, drinks, diet, people, think,
82	sugar, artificial, daddy, sweeteners, weight, sucralose, health, blood, diabetes,
32	lyme, disease, study, antibiotics, better, kills, pathogen, confirms,
90	sugar , natural, sweeteners, artificial, helaty, free, drink,honey, drinks,
96	artificial, sweeteners, sucralose, sugar, gum, insulin, aspartame, study, popular

Table 2. The top words of the relevant topics that were produced by the LDA model.

ID	Irrelevant Topics
37	@arianagrande , love, sweetener , sweetener , album, wait, excited,
48	deal, trade, think,throm, going need, maybe, money, sugar, thats, tax,
64	leche, en, la el ,1 avena, que, caf,una ,2 ,por, sin, es, le, para, canela, lo, mi
92	ad, try, @intheraw, giveaway, cake, stevia, pumpkin , sweepstake, pecan

Table 3. The top words of the *irrelevant* topics that were produced by the LDA model.

We used this approach as a way to automatically extract rough representations (lists of related words) of the major themes in the text, to characterize the topics of discussion on Twitter. Domain experts examined the output and identified salient topics. While many topics were hard to understand, some interesting topics were identified.

Examples of relevant topics are provided in Table 2. Topic 7 is that of general health efficacy around natural and plant-based sweeteners. Topic 12 revolves around specific artificial sweeteners’ taste in the context of sodas. Topic 40 is similar to Topic 12, but focuses more on the general thoughts around artificial sweeteners. Topic 82 is about sweeteners, artificial as well as sugar, as they relate to health issues of the gut, diabetes, and weight. Topic 32 embodies the discussions around a recent study showing that Stevia is more effective in treating Lyme disease than antibiotics. Topic 90 is about artificial and natural sweeteners in food and drinks, including honey. Topic 96 is about artificial sweeteners, sucralose and aspartame, with reference to their common usage in gums and their effects on insulin levels.

Table 3 shows examples of topics that are arguably irrelevant to discussion of sweeteners. Topic 37 is related to the Ariana Grande album, “Sweetener.” Many of the LDA topics were related to this album. Topic 48 contains conversational words related to money, but it is not clear if this is connected to sweeteners. Topic 64 is an example of a topic with non-English words, which we tried to filter out for this analysis. Topic 92 seems to describe advertisements rather than organic discussion of sweeteners.

When manually reviewing samples of tweets, we have observed a very broad array of content, including people describing their usage or abstinence of sweeteners, sharing information/research, and expressing concerns or other opinions.

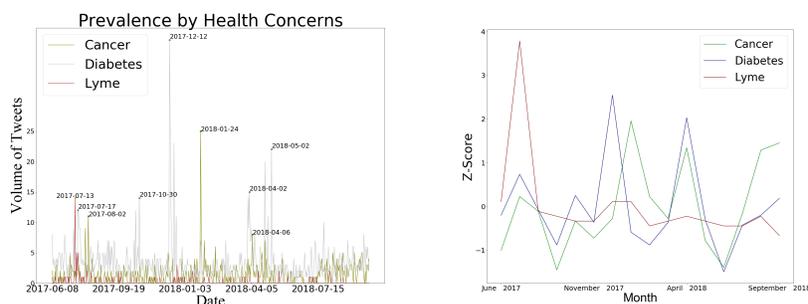


Fig. 1. Volume of tweets (raw volume, left; standardized volume, right) mentioning each of the three diseases we considered.

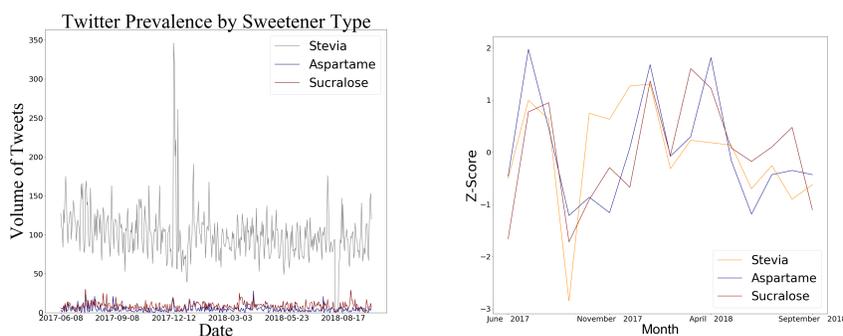


Fig. 2. Volume of tweets (raw, left; standardized, right) mentioning each of the three sweeteners we considered.

4 Temporal Patterns

The volume of tweets was plotted over the time span of data collection to allow for peaks in the chatter to be observed. Figure 1 shows the volumes broken down by different diseases that are mentioned, while Figure 2 shows the volumes broken down by different sweeteners. We see that there is high temporal variability, and upon inspection we find that spikes in volume are usually aligned with something happening in the news.

A thorough search was done for the dates for each of the peaks occurred for the individual health concerns on the news that was being shared. The highest peak for cancer, on January 24th, coincided with a news article that was shared on how artificial sweeteners could someday cause cancer. The peak in July 2017 for Lyme disease coincided with a news article titled, “Stevia the best treatment for Lyme disease, researchers say.” Tweets promoting this article were heavily circulated on that day, with people recommending Stevia as a potential cure.

References

1. Abbar, S., Mejova, Y., Weber, I.: You tweet what you eat: Studying food consumption through Twitter. In: Conference on Human Factors in Computing Systems (CHI) (2015)
2. Ayers, J.W., Althouse, B.M., Johnson, M., Dredze, M., Cohen, J.E.: What's the healthiest day? circaseptan (weekly) rhythms in healthy considerations. *American Journal of Preventive Medicine*, (2014)
3. Blei, D.M., Ng, A., Jordan, M.: Latent Dirichlet allocation. *Journal of Machine Learning Research* **3** (2003)
4. De Choudhury, M., Sharma, S., Kiciman, E.: Characterizing dietary choices, nutrition, and language in food deserts via social media. In: Conference on Computer Supported Cooperative Work and Social Computing (CSCW). pp. 1157–1170. ACM, New York, NY, USA (2016). <https://doi.org/10.1145/2818048.2819956>, <http://doi.acm.org/10.1145/2818048.2819956>
5. Fried, D., Surdeanu, M., Kobourov, S., Hingle, M., Bell, D.: Analyzing the language of food on social media. In: IEEE International Conference on Big Data (2014)
6. Kusmierczyk, T., Trattner, C., Norvag, K.: Temporality in online food recipe consumption and production. In: International Conference on World Wide Web (WWW) (2015)
7. Mejova, Y., Haddadi, H., Noulas, A., Weber, I.: #foodporn: Obesity patterns in culinary interactions. In: International Conference on Digital Health. pp. 51–58 (2015)
8. Nguyen, C.Q., Li, D., Meng, H.W., Kath, S., Nsoesie, E., Li, F., Wen, M.: Building a national neighborhood dataset from geotagged Twitter data for indicators of happiness, diet, and physical activity. *JMIR Public Health Surveillance* **2**(2), e158 (Oct 2016). <https://doi.org/10.2196/publichealth.5869>, <http://www.ncbi.nlm.nih.gov/pubmed/27751984>
9. Pagoto, S., Schneider, K.L., Evans, M., Waring, M.E., Appelhans, B., Busch, A.M., Whited, M.C., Thind, H., Ziedonis, M.: Tweeting it off: characteristics of adults who tweet about a weight loss attempt. *J Am Med Inform Assoc* **21**(6), 1032–1037 (2014)
10. Paul, M.J., Dredze, M.: Social monitoring for public health. In: Synthesis Lectures on Information Concepts, Retrieval, and Services, pp. 1–185. Morgan & Claypool (2017)
11. Pawar, R., Krynitsky, A., Rader, J.: Sweeteners from plants—with emphasis on *Stevia rebaudiana* (bertoni) and *Siraitia grosvenorii* (swingle). *journal = Anal Bioanal Chem*, number = 405, pages = 4397–4407, year = 2013
12. Sharma, S., De Choudhury, M.: Detecting and characterizing nutritional information of food and ingestion content in Instagram. In: International Conference on World Wide Web (WWW) (2015)
13. West, R., White, R.W., Horvitz, E.: From cookies to cooks: Insights on dietary patterns via analysis of web usage logs. In: International Conference on World Wide Web (WWW). Republic and Canton of Geneva, Switzerland (2013)