Applications in Sentiment Analysis and Machine Learning for Identifying Public Health Variables Across Social Media

A Dissertation Presented

by

Eric M. Clark

to

The Faculty of the Graduate College

of

The University of Vermont

In Partial Fulfillment of the Requirements
for the Degree of Doctor of Philosophy
Specializing in Complex Systems & Data Science

January, 2019

Abstract

Twitter, a popular social media outlet, has evolved into a vast source of linguistic data, rich with opinion, sentiment, and discussion. We mined data from several public Twitter endpoints to identify content relevant to healthcare providers and public health regulatory professionals. We began by compiling content related to electronic nicotine delivery systems (or e-cigarettes) as these had become popular alternatives to tobacco products. There was an apparent need to remove high frequency tweeting entities, called bots, that would spam messages, advertisements, and fabricate testimonials. Algorithms were constructed using natural language processing and machine learning to sift human responses from automated accounts with high degrees of accuracy. We found the average hyperlink per tweet, the average character dissimilarity between each individual's content, as well as the rate of introduction of unique words were valuable attributes in identifying automated accounts. We performed a 10-fold Cross Validation and measured performance of each set of tweet features, at various bin sizes, the best of which performed with 97% accuracy. These methods were used to isolate automated content related to the advertising of electronic cigarettes. A rich taxonomy of automated entities, including robots, cyborgs, and spammers, each with different measurable linguistic features were categorized.

Electronic cigarette related posts were classified as automated or organic and content was investigated with a hedonometric sentiment analysis. The overwhelming majority ($\approx$ 80%) were automated, many of which were commercial in nature. Others used false testimonials that were sent directly to individuals as a personalized form of targeted marketing. Many tweets advertised nicotine vaporizer fluid (or e-liquid) in various "kid-friendly" flavors including 'Fudge Brownie', 'Hot Chocolate', 'Circus Cotton Candy' along with every imaginable flavor of fruit, which were long ago banned for traditional tobacco products. Others offered free trials, as well as incentives to retweet and spread the post among their own network. Free prize giveaways were also hosted whose raffle tickets were issued for sharing their tweet. Due to the large youth presence on the public social media platform, this was evidence that the marketing of electronic cigarettes needed considerable regulation. Twitter has since officially banned all electronic cigarette advertising on their platform.

Social media has the capacity to afford the healthcare industry with valuable feedback from patients who reveal and express their medical decision-making process, as well as self-reported quality of life indicators both during and post treatment. We have studied several active cancer patient populations, discussing their experiences with the disease as well as survivor-ship. We experimented with a Convolutional Neural Network (CNN) as well as logistic regression to classify tweets as patient related. This led to a sample of 845 breast cancer survivor accounts to study, over 16 months. We found positive sentiments regarding patient treatment, raising support, and spreading awareness. A large portion of negative sentiments were shared regarding political legislation that could result in loss of coverage of their healthcare. We refer to these online public testimonies as "Invisible Patient Reported Outcomes" (iPROs), because they carry relevant indicators, yet are difficult to capture by conventional means of self-reporting. Our methods can be readily applied interdisciplinary to obtain insights into a particular group of public opinions. Capturing iPROs and public sentiments from online communication can help inform healthcare professionals and regulators, leading to more connected and personalized treatment regimens. Social listening can provide valuable insights into public health surveillance strategies.

Citations

Material from this dissertation has been published in the following form:

Clark, E. M., Williams, J. R., Jones, C. A., Galbraith, R. A., Danforth, C. M., & Dodds, P. S.. (2016). Sifting robotic from organic text: a natural language approach for detecting automation on Twitter. *Journal of Computational Science, 16, 1-7.*

Clark, E. M., Jones, C. A., Williams, J. R., Kurti, A. N., Norotsky, M. C., Danforth, C. M., & Dodds, P. S. (2016). Vaporous marketing: uncovering pervasive electronic cigarette advertisements on Twitter. PLoS One, 11(7), e0157304.

AND

Material from this dissertation has been submitted for publication in the journal of Computers in Biology and Medicine on 08/24/2018 in the following form:

Clark, E. M., James, T., Jones, C. A., Alapati, A., Ukandu, P., Danforth, C. M., & Dodds, P. S. (2018). A Sentiment Analysis of Breast Cancer Treatment Experiences and Healthcare Perceptions Across Twitter. Computers in Biology and Medicine.

in loving memory of

Richard ("Ricky") Bakalian (December 24, 1952 - May 21, 2017)

"Now's the time to be alive - to see it all happen, to be a part of it. That makes the blood race, and each breath is an adventure."

-David Eddings

Acknowledgements

The authors wish to acknowledge the Vermont Advanced Computing Core, which is supported by NASA (NNX-08AO96G) at the University of Vermont which provided High Performance Computing resources that contributed to the research results reported within this poster. EMC was supported by the Vermont Complex Systems Center. CMD and PSD were supported by an NSF BIGDATA grant IIS-1447634. CJ is supported in part by the Center of Biomedical Research Excellence Award P20GM103644 from the National Institute of General Medical Sciences.

Table of Contents

Citations . . . . **ii**

Dedication . . . . **iii**

Acknowledgements . . . . **iv**

List of Figures . . . . **xi**

List of Tables . . . . **xii**

1 Introduction and Literature Review **1**

1.1 Introduction **1**

1.2 Literature Review **4**

1.3 References **7**

2 Sifting Robotic from Organic Text: A Natural Language Approach for Detecting Automation on Twitter **10**

2.1 Introduction **11**

2.2 Data Handling **12**
- 2.2.1 Data-Collection 12
- 2.2.2 Social HoneyPots 13
- 2.2.3 Human Classification of Geo-Tweets 13
- 2.2.4 Types of Users 13

2.3 Methods **16**
- 2.3.1 Classification Algorithm 16
- 2.3.2 Average Pairwise Tweet Dissimilarity 16
- 2.3.3 Word Introduction Decay Rate 17
- 2.3.4 Average URLs per Tweet 18
- 2.3.5 Cross Validation Experiment 19

2.4 Results and Discussion **20**
- 2.4.1 Geo-Tweet Classification Validation 20
- 2.4.2 HoneyPot External Validation 22
- 2.4.3 Calibrated Classifier Performance 23

2.5 Conclusion **26**

2.6 References **27**

3 Vaporous Marketing: Uncovering Pervasive Electronic Cigarette Advertisements on Twitter **29**

3.1 Introduction **30**

3.2 Materials and Methods **32**
3.2.1 Data Collection 32
3.2.2 Automation Classification 33
3.2.3 Categorization by Topics 34

3.3 Results and Discussion **36**
3.3.1 Tweet Sentiment Analysis 37
3.3.2 Sub-Categorical Tweet Topics 41

3.4 Conclusion **43**

3.5 References **44**

4 A Sentiment Analysis of Breast Cancer Treatment Experiences and Healthcare Perceptions Across Twitter **48**

4.1 Introduction **49**

4.2 Materials and methods **50**
4.2.1 Data Description 50
4.2.2 Sentiment Analysis and Hedonometrics 52
4.2.3 Relevance Classification: Logistic Model and CNN Architecture 53

4.3 Results **54**

4.4 Discussion **60**

4.5 Conclusion **62**

4.6 References **63**

5 Measuring Sentiments and Public Perceptions of Surgery Across Twitter **66**

5.1 Introductory Remarks **67**

5.2 Introduction **67**

5.3 Methods **68**

5.4 Results **70**

5.5 Discussion **77**

5.6 Conclusion **78**

5.7 Remarks on Extending Our Framework **78**

5.8 References **80**

Appendices **95**

A Sifting Robotic from Organic Text **96**
A.0.1 Cross Sectional Classifier Performance 96
A.0.2 Model Comparisons 98
A.0.3 Word Introduction Decay Parameter 100
A.0.4 Human Annotation of Twitter Accounts 102
A.0.5 Mixed User Sample Tweets 105

B Vaporous marketing **107**

C Breast Cancer Treatment Experiences and Healthcare Perceptions **112**
C.0.1 Raw 'Cancer' Twitter Data Overview 112
C.0.2 Calculating the Tweet Sampling Proportion 114
C.0.3 Interpreting Word Shift Graphs 116
C.0.4 Sentence Classification Methodology 118
C.0.5 Hashtag Table Sorted by Average Word Happiness 119

D Sentiments and Public Perceptions of Surgery **121**
D.0.1 Surgical Term List Twitter Counts and Sentiments 121

List of Figures

1.1 Twitter policy banning electronic cigarette marketing 3
1.2 LabMT survey example. The smiling emojis are converted to a 9 point scale. 5

2.1 The feature distribution of the 1000 hand coded users are summarized with histograms and violin plots. These show the wide variation in automated features versus Organics. Violin plots show the kernel density estimation of each distribution. Using the Organic features, automated entities are identified by exclusion. 14
2.2 The rank-unique word gap distribution is plotted on a logscale for each user class. 18
2.3 The receiver operator characteristic curve from the 10-fold Cross Validation Experiment performed on the Geo Tweets collected from April through July 2014. The True Positive (TP), False Positive (FP), and thresholds, N, are averaged across the 10 trials. The accuracies are approximated by the AUCs, which we compute using the trapezoid rule. The points depict the best experimental model thresholding window (N). 20
2.4 Accuracy, computed as the AUC is plotted as a function of number of tweets, ranging from 25 to 500. The average True Positive and False Positive Rates over 10 trials is given on twin axes with error bars drawn using the standard error. 21
2.5 Honey Pot Data Set, 10-fold Cross Validation Performance for users with 200 tweets. The black curve represents the 1 000 hand coded HoneyPot users, while the blue curve is the entire English Honeypot dataset. The accuracy increases from 84% to 96%. 23
2.6 Calibrated Classifier Performance on 1 000 User Geo-Tweet Dataset. Correctly classified humans (True Negative), are coded in Green, while correctly identified automatons (True Positives) are coded in red. The black lines demonstrates each feature cutoff. 25

3.1 Tweets from a random sample of 500 organic classified and 500 automated classified accounts were hand coded to gauge the accuracy of the detection algorithm. The feature set of each sampled individual is plotted in three dimensions. The traced box indicate the organic feature cutoff. True Positives (red) are correctly identified automatons, True Negatives (green) are correctly identified Humans, False Negatives (blue) are automatons classified as humans and False Positives (orange) are humans classified as automatons. 35
3.2 Left: Binned User E-cigarette Keyword Tweet Distribution (2012-2014). Right: 2013 Automated Tweet Rank-Frequency Word Cloud. High frequency stop words ('of','the', etc.) are removed from the rank-frequency word distribution. 37
3.3 Categorical Tweet Word-shift Graphs: On the left, Organic Tweets from 2013 are the reference distribution to compare sentiments of Organic Tweets made in 2014 where we see a negative shift in the calculated average word happiness. Due to tweets tagged #EUEcig Ban, January 2014 and December 2013 are omitted. The computed average happiness (h_{avg}) decreases from 5.82 to 5.77 due to both an increase in the negative words 'tobacco', 'drug', 'ban', 'poison', and a decrease in the positive words 'love', 'like', 'haha', 'cool' among others. On the right, Organic Tweets from 2013 are the reference distribution to compare Automated Tweets from 2013. The words 'free' and 'trial' are excluded from the graph, since their high frequency and happiness scores distorts the image. With these key words included the the automated tweet h_{avg} increases from 6.17 to 6.59. 40

4.1 (left) The distribution of tweets per given user is plotted on a log axis. The tail tends to be high frequency automated accounts, some of which provide daily updates or news related to cancer. (right) A frequency time-series of the tweets collected, binned by day. 51

4.2 (left) The distribution of tweets per given patient/survivor is plotted on a log axis along with a statistical summary of patient tweeting behavior. (right) A frequency time-series of patient tweets collected, binned by day. 55
4.3 Computed average word happiness as a function of day (small markers) and month (large markers) for both the 'breast','cancer' and patient distributions. The patient monthly average was less positive than the reference distribution ($h_{\text{avg}} = 5.78$ v. 5.93). 56
4.4 (Left) Word shift graph comparing collected Breast Cancer Patient Tweets, T_{comp}, to all Breast Cancer Tweets, T_{ref}. Patient Tweets were less positive ($h_{\text{avg}} = 5.78$ v. 5.97), due to a decrease in positive words 'mom', 'raise', 'awareness', 'women', 'daughter', 'pink', and 'life' as well as an increase in the negative words 'no(t)', 'patients, 'dying', 'killing', 'surgery' 'sick', 'sucks', and 'bill'. (Right) Word shift graph comparing tweets mentioning the American Healthcare Act (AHCA, 10.5k tweets) to the Affordable Care Act (ACA, 16.9k tweets). AHCA tweets were more negative ($h_{\text{avg}} = 5.48$ v. 6.05) due to a relative increase in the negative words 'scared', 'lose', 'zombie', 'defects', 'depression', 'harm', 'killing', and 'worse'. 58

5.1 (Left) The distribution of tweets per account is plotted on a log axis along with a statistical summary of user tweeting behavior. (Right) A frequency time-series of tweets collected, binned by day. 69
5.2 A time series of the average computed word happiness as a function of both month (large markers) and day (small makers) for the Surgery Tweets (red crosses) and Gardenhose reference tweets (blue circles) for comparison. 72
5.3 (Left) A word shift graph comparing tweets collected mentioning surgery (T_{comp}) to a random unfiltered reference sample of tweets from the same time frame. Surgery tweets were slightly less positive ($h_{\text{avg}} = 5.88$ v. 6.01) due to an increase in negative words including 'cancer', 'hospital' , 'fight', 'pain', 'risk'. This set had a relative increase in positive words 'heart', 'successful', 'well', 'hope', 'luck,'prayers','brain','recovery'. (Right) This graph compares surgery tweets mentioning 'Plastic' (T_{comp}) to 'Cosmetic' surgery tweets (T_{ref}). Plastic surgery tweets were less positive ($h_{\text{avg}} = 5.73$ v. 6.02) due to a relative increase in negative words 'ugly', 'bad', 'never', 'deny', 'fails', 'worst', 'drugs', 'addicted', 'worried', 'die', 'wrong', 'fake', and 'warning' as well as a relative increase of the negative words 'ugly', 'never', 'bad', 'deny', 'fails', 'worst', 'addicted', and 'warning'. 75

A.1 1,000 User Geo Tweet Performance 97
A.2 Feature Receiver Operator Curves 99
A.3 Unique Word Introduction Gaps per Word Introduction Number (rank) 100
A.4 Word Decay Deviations 101

B.1 **European Union E-cigarette Ban Political Debate (#EUecigBan)** (Left) Word shift graph comparing tweets tagged #EUecigBan against 2013 English Organic User Tweets (untagged). (top-right) The automated and Organic tagged tweet distributions are plotted. A histogram displays the counts per language and user class. (bottom-right) Word clouds compare ranked-word frequencies across language and user type. Each categorical time-series exhibits a severe negative trend occurring between December 2013 and January 2014. There is an inverse relationship with the average happiness scores during this time period. This was during the time that the EU was debating strict regulation and a possible ban on specific e-cigarette products. Hashtags (#) allow users to categorize the content of their tweets. During this period, 13,227 sampled tweets were tagged with #EUecigBan. In S1 Fig, a word shift graph (left) visualizes the sentiments from English Organic users using #EUecigBan versus the remaining Organic tweets from 2013. English Tweets tagged #EuEcigBan are the comparison distribution in reference to all other tweets from 2013. Tweets containing #EuEcigBan are on average much more negative (h_{avg} 5.81 versus 5.37) due to an increase in the negative words 'ban', 'stop', 'no', 'not', 'fight', 'against', 'disaster', 'death', 'corruption', 'tobacco', 'kills', etc. The positive words also disfavor the legislation, with the words 'save', 'millions', 'lives', 'support', 'healthy' occurring more frequently. English, French, and German tagged tweets were the most prevalent, and word clouds help visualize themes between language and user class. This shows that Twitter sentiments can be useful in gauging public opinion toward regulation of electronic cigarettes. There is also a heavy automated tweet presence in each language with a similar attitude regarding the legislation, as depicted in the word clouds. Future work should also investigate if and how automated users can impact organic opinion on legislation. 108

C.1 A Frequency time-series of raw 'cancer' tweets collected, binned by day. This sample was compiled from a 10% random sample of Twitter, the 'Gardenhose' feed. (left) The distribution of tweets per given user is plotted on a log axis. The tail tends to be high frequency automated accounts, some of which provide daily updates on horoscope information, or about news related to cancer. The kink in the center is also abnormal and could be representative of other classes of automation. This shows the necessity to sift irrelevant tweets using combinations of keyword removal and content classifiers. 112

C.2 Another frequency time-series of raw 'cancer' tweets collected, binned by day. This sample was compiled from a 1% random sample of Twitter, the 'Spritzer' feed concentrated on keyword 'cancer', during the same time interval as Figure 4.1. We collected over 76 million tweets, which accounted for approximately 65.2% of all tweets mentioning 'cancer' while the data stream was active (i.e., not accounting for power/network outages). The kink that was visible in the previous figure seems to moved outward by almost a factor of 10, since this is a much larger sample of 'cancer' tweets (10% versus $\approx$ 65%). This serves as a comparison to the tweets collected using keywords 'breast' and 'cancer' and to raw 'cancer' tweets collected from the Gardenhose feed. 113

C.3 Overflow limit statistics, plotted per day for both the cancer and breast cancer Twitter feeds with the corresponding approximation of the sampling proportion over the study time frame. 115

C.4 (Left) A word shift graph comparing tweets collected mentioning breast cancer, T_{comp}, to a random unfiltered reference sample of tweets along the same time period. Breast cancer tweets were slightly less positive ($h_{\text{avg}} = 5.97$ v. 6.01) due to an increase in negative words 'fight(ing)', 'stop' , 'battle', 'lost', and 'risk'. This set of tweets featured a relative increase in positive words 'women', 'mom', 'daughter', 'awareness', 'pink', 'save', 'support', and 'survivors', which are referencing aspects of breast cancer awareness, support, and the experiences of survivors and patients. (Right) A word shift graph comparing breast cancer patient tweets to the unfiltered sample. These were more negative (h_{avg} = 5.78 v. 6.01) due to a relative increase in negative words such as 'dying', 'sick', 'killing', 'radiation', and 'hospital' among other terms similar to the figure on the left. 117

C.5 This word cloud displays the most prominent hashtags from all collected "breast cancer" tweets. The hashtag sizes are proportionate to their relative frequencies and colors represent their average ambient happiness scores. Here, light blue terms appear with the most positive LabMT words while purple hashtags appear with relatively more negative terms. 120

List of Tables

3.1 Electronic Cigarette Tweet Category Counts and Twitter Account Classification 36
3.2 Automated Tweet Subcategory Counts 42
3.3 * 42

4.1 **Diagnostic Training Sample Tweet Phrases:** A sample of self-reported diagnostic phrases from tweets used to train the logistic regression content classifier (modified to preserve anonymity). 54
4.2 A table of the 50 most frequently tweeted hashtags (#) from all collected breast cancer tweets (left) and from sampled breast cancer patients (right). The relative computed ambient happiness h_{avg} for each hashtag is colored relative to the group average (blue- negative, orange - positive). 59
4.3 **Sampled Predicted Diagnostic Tweets:** A sample of key phrases from self-reported diagnostic tweets predicted from the CNN classifier with the patient relevant proportional ratio, $\alpha = 1 : 10$. 61

5.1 **Surgery Types** A table key word Twitter stats related to relevant surgical types. The relative computed average happiness h_{avg} for each term is colored relative to the group's average (blue- negative, orange - positive). 73
5.2 **50 Most Frequently Tweeted Hashtags** (#) from all collected surgery tweets (left) and the same list sorted by average happiness score. The relative computed average happiness h_{avg} for each tag is colored relative to the group average (blue- negative, orange - positive). . . . 76

A.1 Annotation Discrepancies of Twitter Accounts 102
A.2 **Robot Sample Tweets** 105
A.3 **Cyborg Sample Tweets** 105
A.4 **Spammer Sample Tweets** 105
A.5 **Human Sample Tweets** 106

B.1 **Electronic Cigarette Table of Key Words** 110

C.1 A table 50 frequently tweeted hashtags (#) sorted by average word happiness from all collected breast cancer tweets (left) and from sampled breast cancer patients (right). 119

D.1 **Broad Symptom Terms** A table key word Twitter stats related to relevant surgical symptoms. The relative computed average happiness h_{avg} for each term is colored relative to the group's average (blue- negative, orange - positive). 122
D.2 **Action Terms :** A table of key word Twitter stats related to relevant to surgical actions. . . 124
D.3 **Basic Symptom Terms :** A table key word Twitter stats related to relevant surgical symptoms. 125
D.4 **Body Region Terms :** A table key word Twitter stats related to relevant body region. 127
D.5 **Body Region Terms :** A table key word Twitter stats related to relevant body region. 128

Chapter 1

Introduction and Literature Review

1.1 Introduction

Social media has become a cornerstone of political discussion as well as an outlet for sharing information and opinions. A wide audience uses Twitter, a prominent social media platform, which allows individuals to publicly interact with the user population, including celebrities, political entities, organizations, and people. Twitter features shorter message lengths originally allowing only 140 characters per post (tweet), which recently increased to 280 characters in 2018. As opposed to the "friend" system implemented by Facebook, a competitor of Twitter, people can choose to "Follow" and communicate with general groups of people whom they are not mutually connected. Hashtags (#) are user designated topics that categorize content within tweets. Using hashtags, people can search for information as well as interact with a targeted public audience. As hashtags proliferate through the social network , some reach a large audience (go viral) and become the trending topic of conversation. Comparing sentiments of tweets mentioning the most prominent hashtags as well as the individuals authoring the content can provide insights into public perception and attitude towards worldly events. We aim to measure these interactions and provide evidence for how each group's perceived attitudes differ.

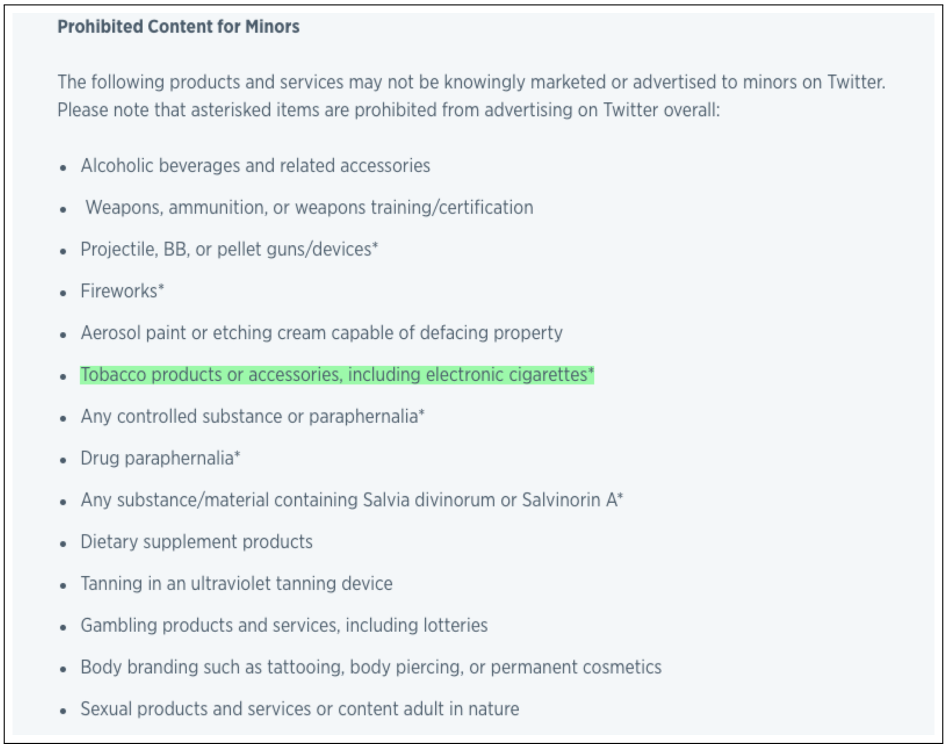
Integrating social media into actionable public health related decision mechanisms is of great interest to healthcare professionals. Social media posts describing an individual's experiences related to symptoms, treatment, survivor-ship are referred to as "Invisible Patient Reported Outcomes" (iPROs), since they are clinically relevant information, yet hard to capture by traditional standards. Using content classifiers, these relevant linguistic features of self- diagnostic tweets, can find individuals whom are actively posting public information about their disease state. All collected tweets from these groups of people could be studied to find comparative emotional themes within the data. Working to capture these iPROs and public opinions from online social media can provide useful information to healthcare professionals, regulators, and public knowledge.

Our methodology focuses on mining social media posts and sifting relevant content using a combination of natural language processing and machine learning. We use hedonometrics,(Dodds et al., 2011, 2015), a data-driven word happiness distribution and set of graphical tools used to investigate emotional themes within sets of text. The average word happiness value, h_{avg}, is calculated and is useful for comparing differing sentiments between groups of tweets and how they change over time. Lower scores correspond to an abundance of negative words and higher scores to more positive terms. Wordshift graphs, (Dodds et al., 2011), identify the terms responsible for a calculated shift in average computed word happiness (h_{avg}).

These methods were applied to data collected from Twitter's streaming Application Programming Interface (API). A 10% random sample of Twitter, the *Gardenhose*, has been archived spanning several years, January 2012 through December 2016. We also used 1% targeted streams, *Spritzer Feeds*, to obtain more comprehensive data from key word searches. This allowed us to examine several million tweets relevant to each analyzed topic.

Machine learning content classification was used to identify tweets relevant to each studied group. Novel machine learning algorithms for detection of automated entities, known as bots, were built and tested using annotated sets of training data. Bots spam messages , advertisements, or other promotional content that pollute organic (human) content and influence the results of a sentiment analysis. These automated entities were easily distinguishable under our classification framework. We identified a rich taxonomy of automation, including robots, cyborgs, and spammers, each with different measurable linguistic characteristics.

Electronic nicotine delivery systems (or e-cigarettes) tweets were mined and analyzed with the Gardenhose feed. A high portion, ($\approx$ 80%), of messages came from automated promotional entities. Sentiments from the automated accounts were measurably more positive, with words to describe their product, discounts, free samples, among other promotional strategies. Some of these profiles were providing false testimonials regarding quitting smoking combustible tobacco using electronic cigarettes. We compared sentiments from both organic and automated entities as well as their abundance. We categorized (with overlap) types of marketed automated content as ‘Commercial’, ‘Cessation Related’, ‘Discounts’, and ‘Flavor Description’. Since the publication of this study, in order to protect minors on their platform, Twitter has officially banned the marketing of electronic cigarettes, see Figure 1.1.



Prohibited Content for Minors

The following products and services may not be knowingly marketed or advertised to minors on Twitter. Please note that asterisked items are prohibited from advertising on Twitter overall:

- Alcoholic beverages and related accessories
- Weapons, ammunition, or weapons training/certification
- Projectile, BB, or pellet guns/devices*
- Fireworks*
- Aerosol paint or etching cream capable of defacing property
- Tobacco products or accessories, including electronic cigarettes*
- Any controlled substance or paraphernalia*
- Drug paraphernalia*
- Any substance/material containing Salvia divinorum or Salvinorin A*
- Dietary supplement products
- Tanning in an ultraviolet tanning device
- Gambling products and services, including lotteries
- Body branding such as tattooing, body piercing, or permanent cosmetics
- Sexual products and services or content adult in nature

Figure 1.1: Twitter policy banning electronic cigarette marketing

Using our framework, (Crannell et al., 2016), studied a population of cancer patients publicly describing their condition and their experiences with the disease. The study applied hedonometrics to rate sentiments attributed to each cancer type from patients. These individuals were compiled using progressive key word searches read by a clinical professional. The process of identifying relevant patient posts was tedious, so we developed a methodology to find patient reported tweets using a data driven content classifier. We've expanded on this study by using a targeted Spritzer Feed to collect over 5.3 million “breast cancer” related tweets spanning 16 months. Using similarly annotated data, we created a set of tweets from patients and survivors describing their condition. We trained a content classifier using a convolutional neural network, (Britz, 2015a), to find and study a group of 845 patients publicly tweeting about their conditions. During the study, political healthcare regulation was highly politicized and debated. A newly proposed plan to replace the Affordable Care Act, (Kaplan et al., 2017), could have stripped healthcare opportunities from individuals with preexisting conditions including patients and survivors afflicted with cancer. There was a stark negative response to the potential loss in coverage, which was measurable using hedonometrics. Our analysis worked as an interdisciplinary proof of concept for conducting social media experiments to gain insights into public health regulatory perspectives.

1.2 Literature Review

There are an abundance of studies that applied Twitter data mining for monitoring public health trends. (Paul and Dredze, 2011), found over a dozen health related ailment types to track symptom related tweets over time for disease monitoring. Influenza was similarly tracked and monitored with tweets, (Lamb et al., 2013). (Reece et al., 2016), could forecast the onset of mental illness using tweets. (Alajajian et al., 2017) used a caloric metric to compare food trends mentioned on Twitter across America. Public perspectives of less serious ailments like acne, (Shive et al., 2013), were categorized by finding users providing experiences with the condition to study their self described treatment process. Many have similarly incorporated machine learning into extracting useful public health information, (Dredze, 2012). We chose to study patients afflicted with cancer for whom social media provides a rich online forum, (Sugawara et al., 2012).

Spam detection, identifying the perpetrating robotic social media accounts, has been a topic of interest for researchers and industry professionals. These accounts produce vastly more content than organic (human) profiles for targeted influence manipulation, (Subrahmanian et al., 2016; Harris, 2013), which can skew results of topical and sentiment analyses. Studies have used tweet metadata, including the number of followers, posting frequency, account age, number of user mentions/replies, username length, and number of retweets, provided by the streaming API to build classification algorithms, (Ferrara et al., 2014; Benevenuto et al., 2010; Chu et al., 2010; Zhang and Paxson, 2011). Others using the daily cycle of human activity, (Ferrara et al., 2014), focused on identifying entities that appear to be organic but are actually financially motivated to spread promotional content. Some of these human accounts rent their influence to advertisers who send short bursts of spam and elude detection using their organic appearing meta data, (Thomas et al., 2011). (Chu et al., 2012) used a content classifier approach to measure entropy over tweet time for classification using a comparable tweet size as our algorithm. SentiBot, created with (Dickerson et al., 2014), is another content automation classification method that utilizes latent Dirichlet allocation (LDA) for topical categorization combined along with sentiment analysis techniques. Our method serves as a proof of concept in solely using linguistic attributes of the tweets, which can be easily integrated into other classification strategies.

Sentiment analysis uses natural language processing to identify positive or negative emotional themes within texts. Hedonometrics is a data-driven metric for calculating the average happiness score of a text, using a lexicon with over the 10,000+ most frequently appearing terms across literature, news articles, and across the web. (Dodds et al., 2011, 2015) used an online survey method to recruit 50 participants to rate each term on a happiness scale, but with varying emojis spanning frown to smile, an example is shown in 1.2

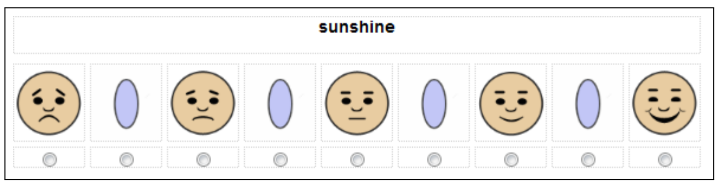


Figure 1.2: LabMT survey example. The smiling emojis are converted to a 9 point scale.

The happiness score of each term was defined as the average value among the 50 ratings, where each word was rated on a 9 point scale ranging from extremely negative (e.g., 'die' 1.74', 'hate' 2.34,, emergency' 3.06) to positive (e.g.,'healthy' 8.02, 'love' 8.42, 'laughter' 8.50). Using these values, the average happiness score of a text can be calculated with the mean frequency of each LabMT term in the corpus, and weighting each term by its corresponding average happiness score within LabMT. This tool has been applied to isolate emotional themes from large sets of text and has been previously applied to monitor public opinion, (Cody et al., 2015). Another approach used the latitude and longitude coordinates of each tweet to make geographical sentiment comparative analyses, (Mitchell et al., 2013).

The hedonometer is a useful relative metric for comparing emotional context between word frequency distributions. We often make this comparison using the average happiness score (h_{avg}) from subsets of text and show the terms that are influencing a positive or negative shift in the frequency of emotionally charged terms. This is calculated by tallying the frequency of each LabMT term per corpus, weighted by each word's happiness score. When applied to social media, the hedonometer, (see `www.hedonometer.org`), helps to visualize the average happiness score of a random 10% sample of Twitter, encompassing millions of posts per day. The daily mean happiness score normally sits around 6.00 and generally fluctuates within 0.05 units by day. With larger samples (i.e., millions of tweets), the average happiness score is more stable, so a noticeable emotional signal is detectable when h_{avg} deviates by 0.5 to 0.1 from the daily average. For example, Christmas day (in 2018 $h_{\text{avg}} = 6.28$) is among the highest daily happiness scores, followed by other holidays and events: 2015 New Years Eve (6.16) , 2015 Valentine's Day (6.18), 2018 International Woman's Day (6.09). Negative deviations from the daily tweet average happiness score usually correspond to worldly events, including natural disasters: 2011 tsunami in Japan (5.97), 2012 Earthquake in Indonesia (5.94); celebrity news: death of Michael Jackson (5.92); as well as states of emergency, like the bombing of the Boston Marathon in 2013 (5.88) or the September 2016 Shooting in Dallas (5.87). When applying hedonometrics to key-word subsets of tweets with much smaller sample sizes (i.e., 1000's of posts), a notable comparative emotional signal is generally associated with a shift of 0.15 to 0.2 from the sample's mean happiness score, which was further explored in (Reagan et al., 2015). These shifts are then quantified by identifying the most influential emotionally charged terms responsible for the computed shift in average word happiness.

Many lexicon (i.e. dictionary) based sentiment analysis methods exist. A similar, but smaller, lexicon sentiment analysis tool, the Affective Norms of English Words (ANEW (Nielsen, 2011), has 2,477 words and phrases which were constructed by a group of individuals. In (Reagan et al., 2015), six lexicon based methods were correlated including: ANEW, Multi-Perspective Question Answering (MPQA) Subjectivity Dictionary (Wilson et al., 2005), Warriner and Kuperman (WK) rated words from SUBTLEX by Mechanical Turk (Warriner and Kuperman, 2015) , Linguistic Inquiry and Word Count (LIWC) (Pennebaker et al., 2001), and Opinion Lexicon (Liu, 2010). Each lexicon based approach showed favorable correlation to results using hedonometrics. Some troubling inconsistencies regarding dictionaries with a binary ranged happiness rating scale, (LIWC and MPQA) in comparison to scores given by LabMT were uncovered. For example, from (Reagan et al., 2015), the following words which are appropriately scored very positively in LabMT all corresponded to negative words (-1) in MPQA: moonlight (7.50), cutest (7.62), finest (7.66), funniest (7.76), comedy (7.98), laughs (8.18), laughing (8.20), laugh (8.22), laughed (8.26), laughter (8.50). We chose the hedonometer for all sentiment calculations due to its performance and numerous applications validating its credibility for analyzing tweet content.

Sentence classification is a difficult computational task that has seen vast improvement with recent advances in machine learning. Statistical methods, including logistic regression (Hosmer Jr et al., 2013), have been applied with varying levels of success. The goal is to assign a binary classification of a text's relevance to a given topic. For our purposes, these were tweets that had been shared by patients whom described their experience with breast cancer. Our study used these classification schemes to collect invisible patient reported outcomes and then analyze all patient tweets separately from the general public.

Maximum Entropy Logistic regression content classifiers, (Genkin et al., 2007), convert sentences from a text to word vectors - called the vocabulary of the classifier. Within the vocabulary, words are weighted by a frequency statistic. A popular metric, (Salton et al., 1975), uses the term frequency crossed with the inverse document frequency (tf-idf), which dampens non-relevant words (like 'of', 'the', 'and', etc.) and bolsters the weight of more rare informative terms. Using this information, the content classifier assigns a binary score to a set of text.

Machine learning algorithms, like Convolutional Neural Networks (CNNs), (Britz, 2015a), have greatly improved upon the accuracy of sentence classification. The CNN loosely works by implementing a filter, called convolution functions, across various subregions of the feature landscape, (Johnson and Zhang, 2015; Britz, 2015b). For text classification, the model tests the robustness of different word embeddings (e.g., phrases) by randomly removing filtered pieces during optimization to find the best predictive terms over the course of training. Then, the input labeled data is divided into training and evaluation to successively test for the best word embedding predictors. The trained model then assigns a binary classification to the relevance of text. These were implemented to find tweets relevant to patient experiences.

1.3 References

Alajajian, S. E., J. R. Williams, A. J. Reagan, S. C. Alajajian, M. R. Frank, L. Mitchell, J. Lahne, C. M. Danforth, and P. S. Dodds (2017). The Lexicocalorimeter: Gauging public health through caloric input and output on social media. *PloS ONE 12*(2), e0168893.

Benevenuto, F., G. Magno, T. Rodrigues, and V. Almeida (2010). Detecting spammers on twitter. In *In Collaboration, Electronic messaging, Anti-Abuse and Spam Conference (CEAS*.

Britz, D. (2015a). Implementing a CNN for text classification in tensorflow.

Britz, D. (2015b, November). Understanding convolutional neural networks for NLP.

Chu, Z., S. Gianvecchio, H. Wang, and S. Jajodia (2010). Who is tweeting on twitter: Human, bot, or cyborg? In *Proceedings of the 26th Annual Computer Security Applications Conference*, ACSAC '10, New York, NY, USA, pp. 21–30. ACM.

Chu, Z., S. Gianvecchio, H. Wang, and S. Jajodia (2012). Detecting automation of twitter accounts: Are you a human, bot, or cyborg? *Dependable and Secure Computing, IEEE Transactions on 9*(6), 811–824.

Cody, E. M., A. J. Reagan, L. Mitchell, P. S. Dodds, and C. M. Danforth (2015). Climate change sentiment on Twitter: An unsolicited public opinion poll. *PloS ONE 10*(8), e0136092.

Crannell, W. C., E. Clark, C. Jones, T. A. James, and J. Moore (2016). A pattern-matched Twitter analysis of U.S. cancer-patient sentiments. *Journal of Surgical Research 206*(2), 536–542.

Dickerson, J. P., V. Kagan, and V. Subrahmanian (2014). Using sentiment to detect bots on twitter: Are humans more opinionated than bots? In *Advances in Social Networks Analysis and Mining (ASONAM), 2014 IEEE/ACM International Conference on*, pp. 620–627. IEEE.

Dodds, P. S., E. M. Clark, S. Desu, M. R. Frank, A. J. Reagan, J. R. Williams, L. Mitchell, K. D. Harris, I. M. Kloumann, J. P. Bagrow, K. Megerdoomian, M. T. McMahon, B. F. Tivnan, and C. M. Danforth (2015). Human language reveals a universal positivity bias. *Proceedings of the National Academy of Sciences 112*(8), 2389–2394.

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

Dredze, M. (2012). How social media will change public health. *IEEE Intelligent Systems 27*(4), 81–84.

Ferrara, E., O. Varol, C. Davis, F. Menczer, and A. Flammini (2014). The rise of social bots. *CoRR abs/1407.5225*.

Genkin, A., D. D. Lewis, and D. Madigan (2007). Large-scale Bayesian logistic regression for text categorization. *Technometrics 49*(3), 291–304.

Harris, D. (2013, Feb). Can evil data scientists fool us all with the world's best spam? `goo.gl/psEguf`.

Hosmer Jr, D. W., S. Lemeshow, and R. X. Sturdivant (2013). *Applied logistic regression*, Volume 398. John Wiley & Sons.

Johnson, R. and T. Zhang (2015). Semi-supervised convolutional neural networks for text categorization via region embedding. In *Advances in Neural Information Processing Systems*, pp. 919–927.

Kaplan, T., R. Pear, and E. Huetteman (2017, March). Lacking the votes for passage, house calls off Obamacare repeal vote.

Lamb, A., M. J. Paul, and M. Dredze (2013). Separating fact from fear: Tracking flu infections on Twitter. In *HLT-NAACL*, pp. 789–795.

Liu, B. (2010). Sentiment analysis and subjectivity. *Handbook of natural language processing 2*, 627–666.

Mitchell, L., M. R. Frank, K. D. Harris, P. S. Dodds, and C. M. Danforth (2013). The geography of happiness: Connecting twitter sentiment and expression, demographics, and objective characteristics of place. *PloS one 8*(5), e64417.

Nielsen, F. Å. (2011). A new anew: Evaluation of a word list for sentiment analysis in microblogs. *arXiv preprint arXiv:1103.2903*.

Paul, M. J. and M. Dredze (2011). You are what you tweet: Analyzing Twitter for public health. *ICWSM 20*, 265–272.

Pennebaker, J. W., M. E. Francis, and R. J. Booth (2001). Linguistic inquiry and word count: Liwc 2001. *Mahway: Lawrence Erlbaum Associates 71*(2001), 2001.

Reagan, A. J., B. Tivnan, J. R. Williams, C. M. Danforth, and P. S. Dodds (2015). Benchmarking sentiment analysis methods for large-scale texts: A case for using continuum-scored words and word shift graphs. *arXiv preprint arXiv:1512.00531*.

Reece, A. G., A. J. Reagan, K. L. Lix, P. S. Dodds, C. M. Danforth, and E. J. Langer (2016). Forecasting the onset and course of mental illness with Twitter data. *arXiv preprint arXiv:1608.07740*.

Salton, G., A. Wong, and C.-S. Yang (1975). A vector space model for automatic indexing. *Communications of the ACM 18*(11), 613–620.

Shive, M., M. Bhatt, A. Cantino, J. Kvedar, and K. Jethwani (2013). Perspectives on acne: what twitter can teach health care providers. *JAMA Dermatology 149*(5), 621–622.

Subrahmanian, V., A. Azaria, S. Durst, V. Kagan, A. Galstyan, K. Lerman, L. Zhu, E. Ferrara, A. Flammini, F. Menczer, et al. (2016). The darpa twitter bot challenge. *arXiv preprint arXiv:1601.05140*.

Sugawara, Y., H. Narimatsu, A. Hozawa, L. Shao, K. Otani, and A. Fukao (2012). Cancer patients on twitter: a novel patient community on social media. *BMC Research Notes 5*(1), 699.

Thomas, K., C. Grier, D. Song, and V. Paxson (2011). Suspended accounts in retrospect: An analysis of twitter spam. In *Proceedings of the 2011 ACM SIGCOMM Conference on Internet Measurement Conference*, IMC '11, New York, NY, USA, pp. 243–258. ACM.

Warriner, A. B. and V. Kuperman (2015). Affective biases in english are bi-dimensional. *Cognition and Emotion 29*(7), 1147–1167.

Wilson, T., J. Wiebe, and P. Hoffmann (2005). Recognizing contextual polarity in phrase-level sentiment analysis. In *Proceedings of the conference on human language technology and empirical methods in natural language processing*, pp. 347–354. Association for Computational Linguistics.

Zhang, C. M. and V. Paxson (2011). Detecting and analyzing automated activity on twitter. In *Proceedings of the 12th International Conference on Passive and Active Measurement*, PAM'11, Berlin, Heidelberg, pp. 102–111. Springer-Verlag.

Chapter 2

Sifting Robotic from Organic Text: A Natural Language Approach for Detecting Automation on Twitter

Twitter, a popular social media outlet, has evolved into a vast source of linguistic data, rich with opinion, sentiment, and discussion. Due to the increasing popularity of Twitter, its perceived potential for exerting social influence has led to the rise of a diverse community of automatons, commonly referred to as bots. These inorganic and semi-organic Twitter entities can range from the benevolent (e.g., weather-update bots, help-wanted-alert bots) to the malevolent (e.g., spamming messages, advertisements, or radical opinions). Existing detection algorithms typically leverage metadata (time between tweets, number of followers, etc.) to identify robotic accounts. Here, we present a powerful classification scheme that exclusively uses the natural language text from organic users to provide a criterion for identifying accounts posting automated messages. Since the classifier operates on text alone, it is flexible and may be applied to any textual data beyond the Twittersphere.

2.1 Introduction

Twitter has become a mainstream social outlet for the discussion of a myriad of topics through microblogging interactions. Members chiefly communicate via short text-based public messages restricted to 140 characters, called tweets. As Twitter has evolved from a simple microblogging social media interface into a mainstream source of communication for the discussion of current events, politics, consumer goods/services, it has become increasingly enticing for parties to gameify the system by creating automated software to send messages to organic (human) accounts as a means for personal gain and for influence manipulation (Subrahmanian et al., 2016; Harris, 2013). The results of sentiment and topical analyses can be skewed by robotic accounts that dilute legitimate public opinion by algorithmically generating vast amounts of inorganic content. Nevertheless, data from Twitter is becoming a source of interest in public health and economic research in monitoring the spread of disease (Sadilek et al., 2012; Wagstaff and Culyer, 2012) and gaining insight into public health trends (Mitchell et al., 2013).

In related work (Ferrara et al., 2014; Benevenuto et al., 2010; Chu et al., 2010; Zhang and Paxson, 2011), researchers have built classification algorithms using metadata idiosyncratic to Twitter, including the number of followers, posting frequency, account age, number of user mentions/replies, username length, and number of retweets. However, relying on metadata can be problematic: sophisticated spam algorithms now emulate the daily cycle of human activity and author borrowed content to appear human (Ferrara et al., 2014). Another problematic spam tactic is the renting of accounts of legitimate users (called sponsored accounts), to introduce short bursts of spam and hide under the user's organic metadata to mask the attack (Thomas et al., 2011).

A content based classifier proposed by (Chu et al., 2012) measures the entropy between Twitter time intervals along with user meta data to classify Twitter accounts, and requires a comparable number of tweets ($\geq$ 60) for adequate classification accuracy as our proposed method. SentiBot, another content based classifier (Dickerson et al., 2014), utilizes latent Dirichlet allocation (LDA) for topical categorization combined with sentiment analysis techniques to classify individuals as either bots or humans. We note that as these automated entities evolve their strategies, combinations of our proposed methods and studies previously mentioned may be required to achieve reasonable standards for classification accuracy. Our method classifies accounts solely based upon their linguistic attributes and hence can easily be integrated into these other proposed strategies.

We introduce a classification algorithm that operates using three linguistic attributes of a user's text. The algorithm analyzes:

1. the average URL count per tweet
2. the average pairwise lexical dissimilarity between a user's tweets, and
3. the word introduction rate decay parameter of the user for various proportions of time-ordered tweets

We provide detailed descriptions of each attribute in the next section. We then test and validate our algorithm on 1 000 accounts which were hand coded as automated or human.

We find that for organic users, these three attributes are densely clustered, but can vary greatly for automatons. We compute the average and standard deviation of each of these dimensions for various numbers of tweets from the human coded organic users in the dataset. We classify accounts by their distance from the averages from each of these attributes. The accuracy of the classifier increases with the number of tweets collected per user. Since this algorithm operates independently from user metadata, robotic accounts do not have the ability to adaptively conceal their identities by manipulating their user attributes algorithmically. Also, since the classifier is built from time ordered tweets, it can determine if a once legitimate user begins demonstrating dubious behavior and spam tactics. This allows for social media data-miners to dampen a noisy dataset by weeding out suspicious accounts and focus on purely organic tweets.

2.2 Data Handling

2.2.1 Data-Collection

We filtered a 1% sample of Twitter's streaming API (the spritzer feed) for tweets containing geo-spatial metadata spanning the months of April through July in 2014. Since roughly 1% of tweets provided GPS located spatial coordinates, our sample represents nearly all of the tweets from users who enable geotagging. This allows for much more complete coverage of each user's account. From this sample, we collected all of the geo-tweets from the most active 1 000 users for classification as human or robot and call this the Geo-Tweet dataset.

2.2.2 Social HoneyPots

To place our classifier in the context of recent work, we applied our algorithm to another set of accounts collected from the Social HoneyPot Experiment (Lee et al., 2011). This work exacted a more elaborate approach to find automated accounts on Twitter by creating a network of fake accounts, called 'Devils' (Lee et al., 2010), that would tweet about trending topics amongst themselves in order to tempt robotic interactions. The experiment was analyzed and compiled into a dataset containing the tweets of "legitimate users" and those classified as "content polluters". We note that the users in this dataset were not hand coded. Accounts that followed the Devil honeypot accounts were deemed robots. Their organic users were compiled from a random sample of Twitter, and were only deemed organic because these accounts were not suspended by Twitter at the time. Hence the full HoneyPot dataset can only serve as an estimate of the capability of this classification scheme.

2.2.3 Human Classification of Geo-Tweets

Each of the 1 000 users were hand classified separately by two evaluators. All collected tweets from each user were reviewed until the evaluator noticed the presence of automation. If no subsample of tweets appeared to be algorithmically generated, the user was classified as human. The results were merged, and conflicting entries were resolved to produce a final list of user ids and codings. See Figure 1 for histograms and violin plots summarizing the distributions of each user class. We note that any form of perceived automation was sufficient to deem the account as automated. See SI for samples of each of these types of tweets from each user class and a more thorough description of the annotation process.

2.2.4 Types of Users

We consider organic content, i.e. from human accounts, as those that have not tweeted in an algorithmic fashion. We focused on three distinct classes of automated tweeting:

Robots: Tweets from these accounts draw on a strictly limited vocabulary. The messages follow a very structured pattern, many of which are in the form of automated updates. Examples include Weather

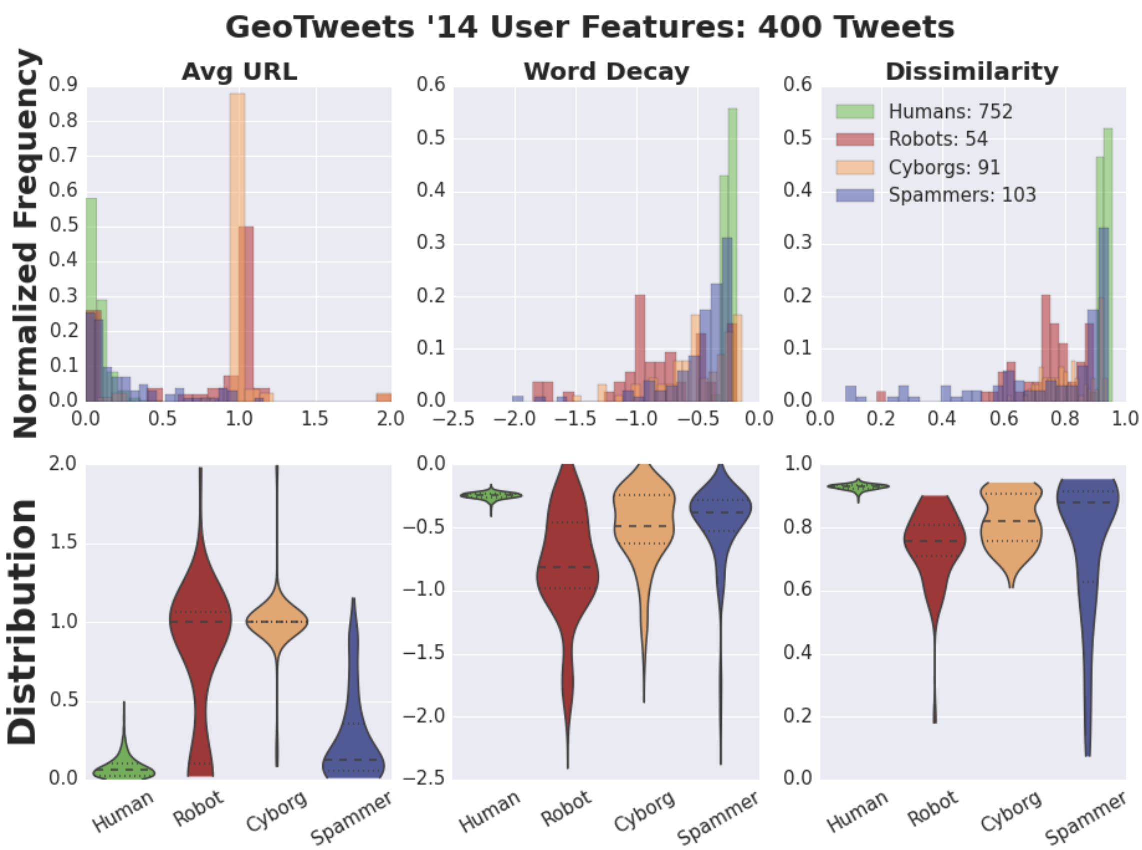


Figure 2.1: The feature distribution of the 1000 hand coded users are summarized with histograms and violin plots. These show the wide variation in automated features versus Organics. Violin plots show the kernel density estimation of each distribution. Using the Organic features, automated entities are identified by exclusion.

Condition Update Accounts, Police Scanner Update Accounts, Help Wanted Update Accounts, etc.

Cyborgs: The most covert of the three, these automatons exhibit human-like behavior and messages through loosely structured, generic, automated messages and from borrowed content copied from other sources. Since many malicious cyborgs on Twitter try to market an idea or product, a high proportion of their tweets contain URLs, analogous to spam campaigns studied on Facebook (Gao et al., 2010). Messages range from the backdoor advertising of goods and services (Huang et al., 2014a) to those trying to influence social opinion or even censor political conversations (Thomas et al., 2012). These accounts act like puppets from a central algorithmic puppeteer to push their product on organic users while trying to appear like an organic user (Wu et al., 2013). Since these accounts tend to borrow content, they have a much larger vocabulary in comparison to ordinary robots. Due to Twitter's 140 character-per-tweet restriction, some of the borrowed content being posted must be truncated. A notable attribute of many cyborgs is the presence of incomplete messages followed by an ellipsis and a URL. Included in this category are 'malicious promoter' accounts (Lee et al., 2011) that are radically promoting a business or an idea systematically.

Human Spammers: These are legitimate accounts that abuse an algorithm to post a burst of almost indistinguishable tweets that may differ by a character in order to fool Twitter's spam detection protocols. These messages are directed at a particular user, commonly for a follow request to attempt to increase their social reach and influence.

Although we restrict our focus to the aforementioned classes, we did notice the presence of other sub-classes, which we have named "listers", and "quoters", that have both organic and automaton features. Listers are accounts that send their messages to large groups of individuals at once. Quoters are dedicated accounts that are referencing distant passages from literature or song lyrics. Most of the tweets from these accounts are all encased in quotations. These accounts also separately tweet organic content. We classified these accounts as human because there was not sufficient evidence suggesting these behaviors were indeed automated.

2.3 Methods

2.3.1 Classification Algorithm

The classifier, $\mathcal{C}$, takes ordinal samples of tweets from each user, μ, of varying number, s, to determine if the user is a human posting strictly organic content or is algorithmically automating tweets:

$$\mathcal{C} : \mu^s \to \{0, 1\} = \{\text{Organic, Automaton}\}.$$

Although we have classified each automaton into three distinct classes, the classifier is built more simply to detect and separate organic content from automated. To classify the tweets from a user, we measure three distinct linguistic attributes:

1. Average Pairwise Tweet Dissimilarity,
2. Word Introduction Rate Decay Parameter,
3. Average number of URLs (hyperlinks) per tweet.

2.3.2 Average Pairwise Tweet Dissimilarity

Many algorithmically generated tweets contain similar structures with minor character replacements and long chains of common substrings. Purely organic accounts have tweets that are very dissimilar on average. The length of a tweet, t, is defined as the number of characters in the tweet and is denoted $|t|$. Each tweet is cleaned by truncating multiple whitespace characters and the metric is performed case insensitively. A sample of s tweets from a particular user is denoted T^s_μ. Given a pair of tweets from a particular user, $t_i, t_j \in T^s_\mu$, the pairwise tweet dissimilarity, $D(t_i, t_j)$, is given by subtracting the length of the longest common subsequence of both tweets, $|LCS(t_i, t_j)|$ and then weighting by the sum of the lengths of both tweets:

$$\text{D}(t_i, t_j) = \frac{|t_i| + |t_j| - 2 \cdot |LCS(t_i, t_j)|}{|t_i| + |t_j|}.$$

The average tweet dissimilarity of user μ for sample size of s tweets is calculated as:

$$\mu^s_{lcs} = \frac{1}{\binom{s}{2}} \cdot \sum_{t_i, t_j \in T^s_\mu} D(t_i, t_j).$$

For example, given the two tweets:
(t_1, t_2) = (I love Twitter, I love to spam). Then $|t_1|$ = $|t_2|$ = 14, $LCS(t_1, t_2) = |\text{I love t}| = 8$ (including whitespaces) and we calculate the pairwise tweet dissimilarity as:

$$D(t_1, t_2) = \frac{14 + 14 - 2 \cdot 8}{14 + 14} = \frac{12}{28} = \frac{3}{7}.$$

2.3.3 Word Introduction Decay Rate

Since social robots *automate* messages, they have a limited and crystalline vocabulary in comparison to organic accounts. Even cyborgs that mask their automations with borrowed content cannot fully mimic the rate at which organic users introduce unique words into their text over time. The word introduction rate is a measure of the number of unique word types introduced over time from a given sample of text (Williams et al., 2015). The rate at which unique words are introduced naturally decays over time, and is observably different between automated and organic text. By testing many random word shufflings of a text, we define $\overline{m_n}$ as the average number of words between the n^{th} and $n + 1^{st}$ initial unique word type appearances. From (Williams et al., 2015), the word introduction decay rate, $\alpha(n)$, is given as

$$\alpha(n) = 1/\overline{m_n} \propto n^{-\gamma} \qquad \text{for } \gamma > 0.$$

For each user, the scaling exponent of the word introduction decay rate, α, is approximated by performing standard linear regression on the last third of the log-transformed tail of the average gap size distribution as a function of word introduction number, n (Williams et al., 2015). In figure 2.2 below, the log transformed rank-unique word gap distribution is given for each individual in the data set. Here the human population (green) is distinctly distributed in comparison to the automatons.

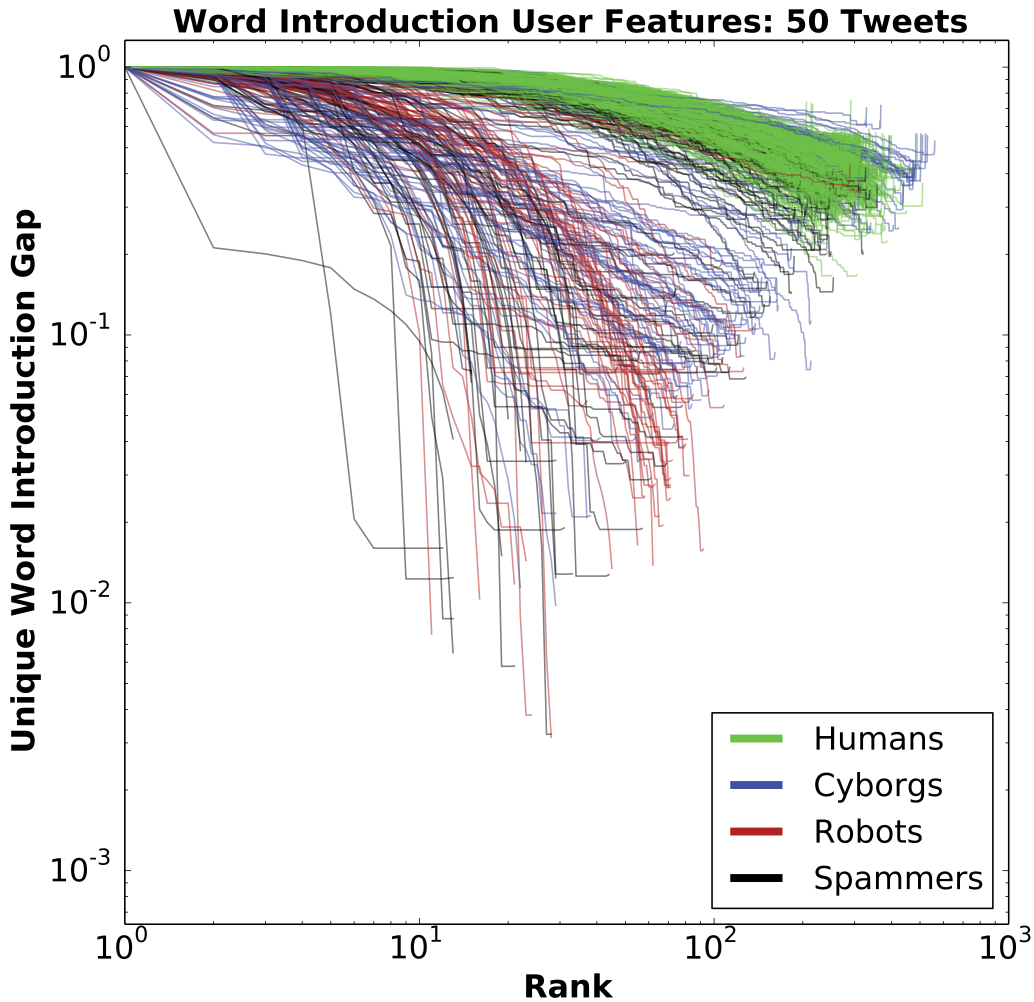


Figure 2.2: The rank-unique word gap distribution is plotted on a logscale for each user class.

2.3.4 Average URLs per Tweet

Hyperlinks (URLs) help automatons spread spam and malware (Thomas et al., 2011; Brown et al., 2008; Wagner, Mitter, Strohmaier, and Korner, Wagner et al.). A high fraction of tweets from spammers tend to contain some type of URL in comparison to organic individuals, making the average URLs per tweet a valuable attribute for bot classification algorithms (Chu et al., 2010; Lee et al., 2010a,b). For each user, the average URL rate is measured by the total number of occurrences of the substring 'http:' within tweets, and then divided by the total number of tweets authored by the user in the sample of size s:

$$\mu^s_{url} = \frac{\#\text{Occurrences of 'http:'}}{\#\text{Sampled Tweets}}.$$

2.3.5 Cross Validation Experiment

We perform a standard 10-fold Cross Validation procedure on the 2014 Geo-Tweet data set to measure the accuracy of using each linguistic feature for classifying Organic accounts. We divided individuals into 10 equally sized groups. Then 10 trials are performed where 9 of the 10 groups are used to train the algorithm to classify the final group.

During the Calibration phase, we measure each of the three features for every human coded account in the training set. We sequentially collect tweets from each user from a random starting position in time. We record the arithmetic mean and standard deviation of the Organic attributes to classify the remaining group. The classifier distinguishes Human from Automaton by using a varying threshold, n, from the average attribute value computed from the training set. For each attribute, we classify each user as an automaton if their feature falls further than n standard deviations away from the organic mean, for varying n.

For each trial, the False Positives and True Positives for a varying window size, n, are recorded. To compare to other bot-detection strategies, we rate True Positives as the success at which the classifier identifies automatons by exclusion, and False Positives as humans that are incorrectly classified as automatons. The results of the trials for varying tweet sizes are averaged and visualized with a Receiver Operator Characteristic curve (ROC) (see Figure 2.3). The accuracy of each experiment is measured as the area under the ROC, or AUC. To benchmark the classifier, a 10-fold cross validation was also performed on the HoneyPot tweet-set which we describe in the following section.

2.4 Results and Discussion

2.4.1 Geo-Tweet Classification Validation

The ROC curves for the Geo-Tweet 10-fold Cross Validation Experiment for varying tweet bins in Figure 2.3 show that the accuracy increases as a function of number of tweets.

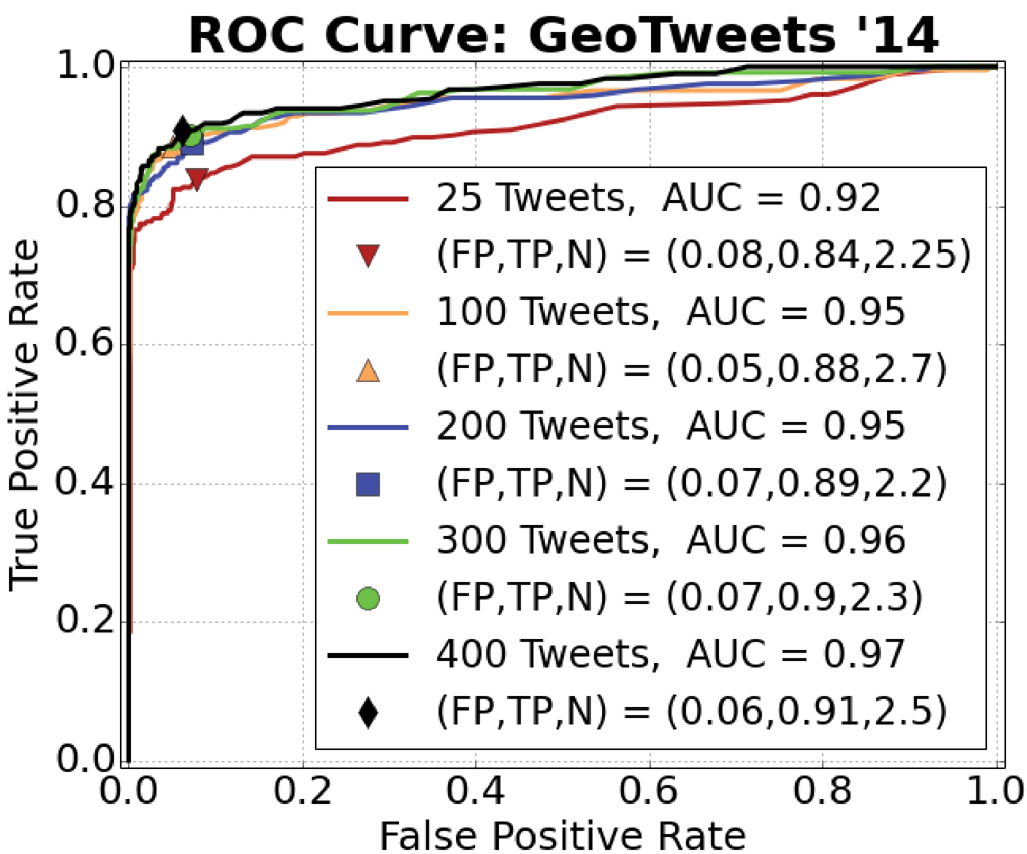


Figure 2.3: The receiver operator characteristic curve from the 10-fold Cross Validation Experiment performed on the Geo Tweets collected from April through July 2014. The True Positive (TP), False Positive (FP), and thresholds, N, are averaged across the 10 trials. The accuracies are approximated by the AUCs, which we compute using the trapezoid rule. The points depict the best experimental model thresholding window (N).

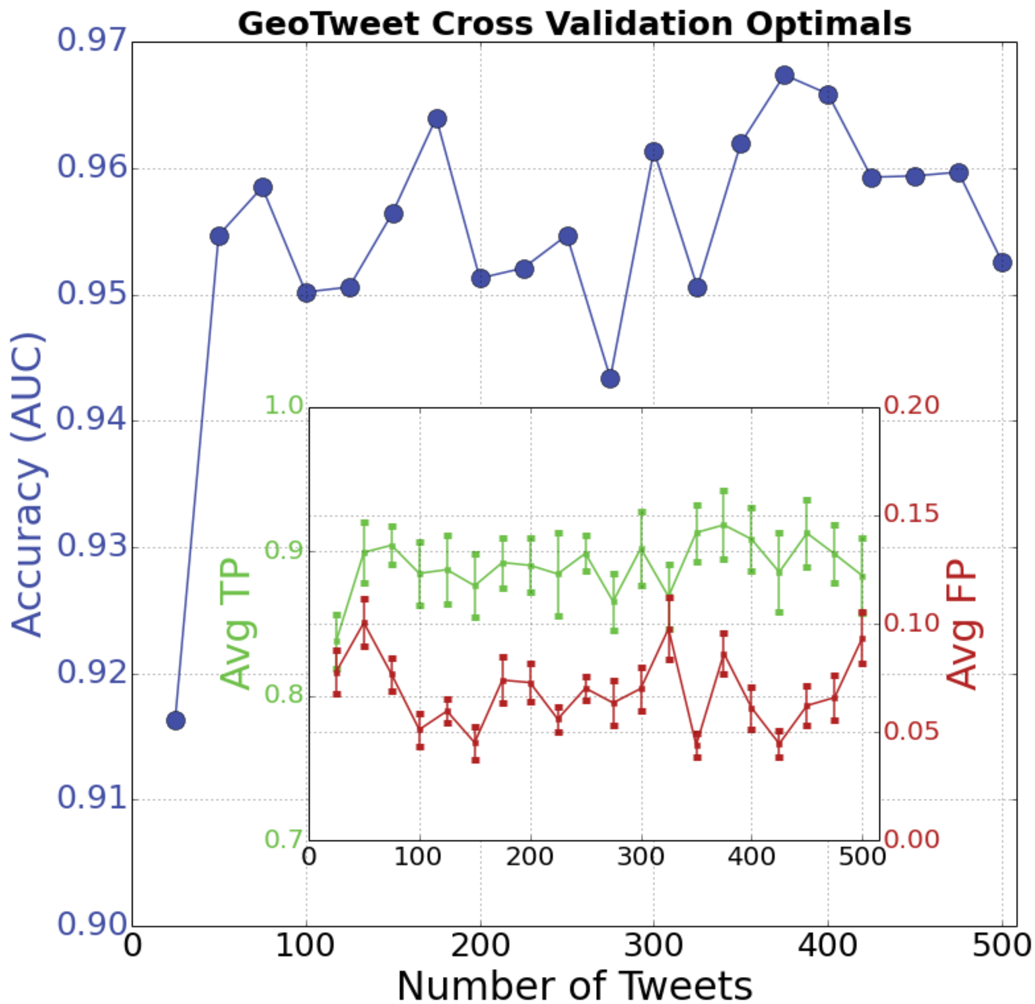


Figure 2.4: Accuracy, computed as the AUC is plotted as a function of number of tweets, ranging from 25 to 500. The average True Positive and False Positive Rates over 10 trials is given on twin axes with error bars drawn using the standard error.

Although the accuracy of the classifier increases with the number of collected tweets, we see in Figure 2.4 that within 50 tweets the accuracy of the average of 10 random trials is only slightly higher than a 500 tweet user sample. While this is very beneficial to our task (isolating humans), we note that larger samples see greater returns when one instead wants to isolate spammers, that tweet random bursts of automation.

2.4.2 HoneyPot External Validation

The classifier was tested on the Social Honeypot Twitter-bot dataset provided by (Lee et al., 2011). Results are visualized with a ROC curve in Figure 2.5. The averaged optimal threshold for the full English user dataset (blue curve) had a high true positive rate (correctly classified automatons: 86%), but also had a large false positive rate (misclassified humans: 22%).

The Honeypot Dataset relied on Twitter's spam detection protocols to label their randomly collected "legitimate users". Some forms of automation (weather-bots, help-wanted bots) are permitted by Twitter. Other cyborgs that are posting borrowed organic content can fool Twitter's automation criterion. This ill formation of the training set greatly reduces the ability of the classifier to distinguish humans from automatons, since the classifier gets the wrong information about what constitutes a human. To see this, a random sample of 1 000 English Honeypot users was hand-coded to mirror the previous experiment. On this smaller sample (black curve in Figure 4), the averaged optimal threshold accuracy increased to 96%.

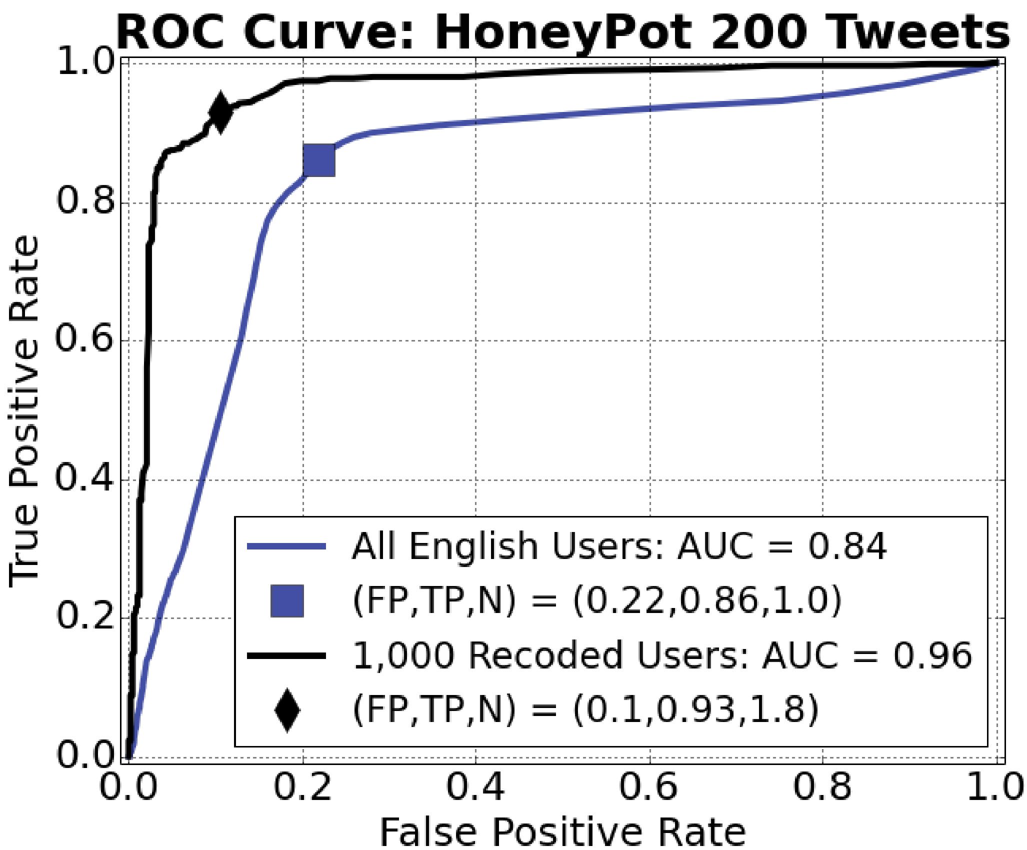


Figure 2.5: Honey Pot Data Set, 10-fold Cross Validation Performance for users with 200 tweets. The black curve represents the 1 000 hand coded HoneyPot users, while the blue curve is the entire English Honeypot dataset. The accuracy increases from 84% to 96%.

2.4.3 Calibrated Classifier Performance

We created the thresholding window of final calibrated classifier using the results from the calibration experiment. We average the optimal parameters from the 10-fold cross validation on the Geo-Tweet dataset from each of the 10 calibration trials for tweet bins ranging from 25 to 500 in increments of 25 tweets. We also average and record the optimal parameter windows, n_{opt} and their standard deviations, σ_{opt}. The standard deviations serve as a tuning parameter to increase the sensitivity of the classifier, by increasing the feature cutoff window (n). The results from applying the calibrated classifier to the full set of 1 000 users, using 400 tweet bags is given in Figure 2.6. The feature cutoff window (black lines) estimates if the user's content is organic or automated. Human feature sets (True Negatives: 716) are densely distributed with a 4.79% False Positive Rate (i.e., humans classified as robots). The classifier accurately classified 90.32% of the automated accounts and 95.21% of the Organic accounts. See Figure S1 for a cross sectional comparison of each feature set. We note that future work may apply different methods in statistical classification to optimize these feature sets, and that using these simple cutoffs already leads to a high level of accuracy.

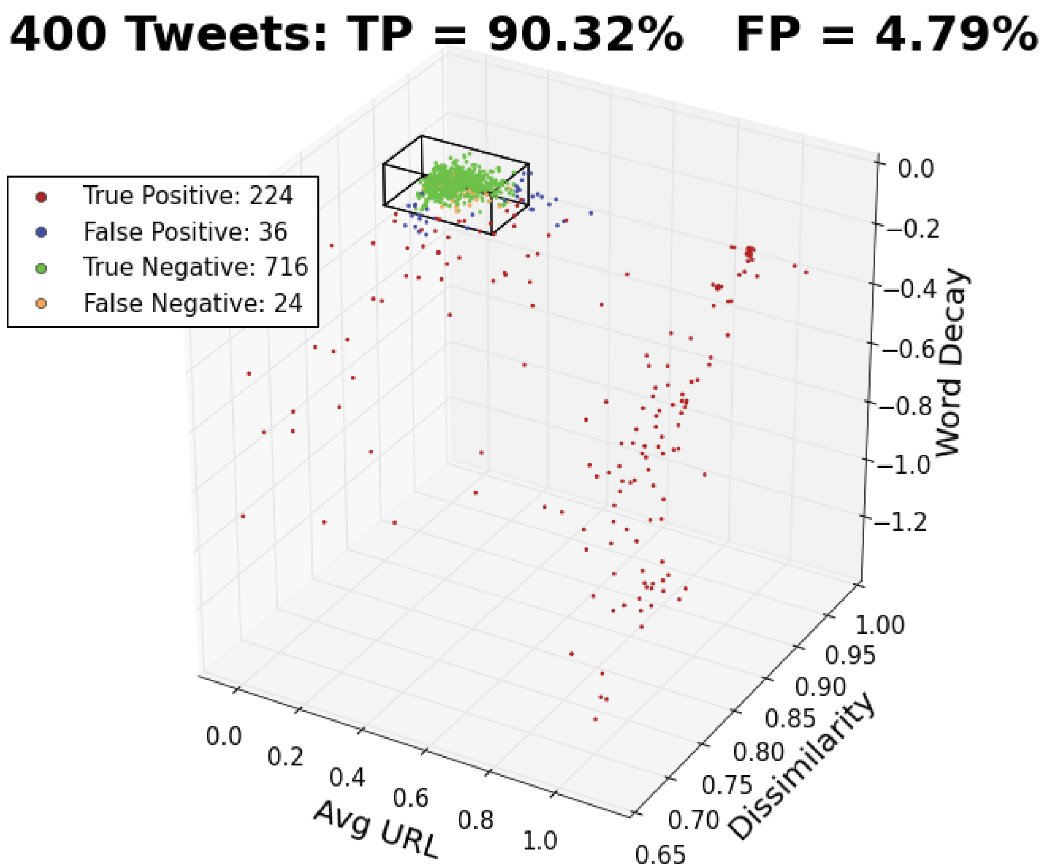


Figure 2.6: Calibrated Classifier Performance on 1 000 User Geo-Tweet Dataset. Correctly classified humans (True Negative), are coded in Green, while correctly identified automatons (True Positives) are coded in red. The black lines demonstrates each feature cutoff.

2.5 Conclusion

Using a flexible and transparent classification scheme, we have demonstrated the potential of using linguistic features as a means of classifying automated activity on Twitter. Since these features do not use the metadata provided by Twitter, our classification scheme may be applicable outside of the Twittersphere. Future work can extend this analysis multilingually and incorporate additional feature sets with an analogous classification scheme. URL content can also be more deeply analyzed to identify organic versus SPAM related hyperlinks.

We note the potential for future research to investigate and to distinguish between each sub-class of automaton. We formed our taxonomy according to the different modes of text production. Our efforts were primarily focused in separating any form of automation from organic,human content. In doing so we recognized three distinct classes of these types of automated accounts. However, boundary cases (e.g. cyborg-spammers, robot-spammers, robotic-cyborgs, etc.) along with other potential aforementioned subclasses (e.g. listers, quoters, etc.) can limit the prowess of our current classification scheme tailored towards these subclasses. We have shown that human content is distinctly different from these forms of automation, and that for a binary classification of automated or human, these features have a very reasonable performance with our proposed algorithm.

Our study distinguishes itself by focusing on automated behavior that is tolerated by Twitter, since both types of inorganic content can skew the results of sociolinguistic analyses. This is particularly important, since Twitter has become a possible outlet for health economics (Wagstaff and Culyer, 2012) research including monitoring patient satisfaction and modeling disease spread (Broniatowski et al., 2013; Sadilek et al., 2012). Monitoring excessive social media marketing of electronic nicotine delivery systems (also known as e-cigarettes), discussed in (Clark et al., 2015; Huang et al., 2014b), makes classifying organic and automated activity relevant for research that can benefit policy-makers regarding public health agendas. Isolating organic content on Twitter can help dampen noisy data-sets and is pertinent for research involving social media data and other linguistic data sources where a mixture of humans and automatons exist.

In health care, a cardinal problem with the use of electronic medical records is their lack of interoperability. This is compounded by a lack of standardization and use of data dictionaries which results in a lack of precision concerning our ability to collate signs, symptoms, and diagnoses. The use of millions or billions of tweets concerning a given symptom or diagnosis might help to improve that precision. But it would be a major setback if the insertion of data tweeted from automatons would obscure useful interpretation of such data. We hope that the approaches we have outlined in the present manuscript will help alleviate such problems.

2.6 References

Benevenuto, F., G. Magno, T. Rodrigues, and V. Almeida (2010). Detecting spammers on twitter. In *In Collaboration, Electronic messaging, Anti-Abuse and Spam Conference (CEAS*.

Broniatowski, D. A., M. J. Paul, and M. Dredze (2013). National and local influenza surveillance through twitter: An analysis of the 2012-2013 influenza epidemic. *PloS one 8*(12), e83672.

Brown, G., T. Howe, M. Ihbe, A. Prakash, and K. Borders (2008). Social networks and context-aware spam. In *Proceedings of the 2008 ACM Conference on Computer Supported Cooperative Work*, CSCW '08, New York, NY, USA, pp. 403–412. ACM.

Chu, Z., S. Gianvecchio, H. Wang, and S. Jajodia (2010). Who is tweeting on twitter: Human, bot, or cyborg? In *Proceedings of the 26th Annual Computer Security Applications Conference*, ACSAC '10, New York, NY, USA, pp. 21–30. ACM.

Chu, Z., S. Gianvecchio, H. Wang, and S. Jajodia (2012). Detecting automation of twitter accounts: Are you a human, bot, or cyborg? *Dependable and Secure Computing, IEEE Transactions on 9*(6), 811–824.

Clark, E. M., C. A. Jones, J. R. Williams, A. N. Kurti, M. C. Nortotsky, C. M. Danforth, and P. S. Dodds (2015). Vaporous marketing: Uncovering pervasive electronic cigarette advertisements on twitter. *arXiv preprint arXiv:1508.01843*.

Dickerson, J. P., V. Kagan, and V. Subrahmanian (2014). Using sentiment to detect bots on twitter: Are humans more opinionated than bots? In *Advances in Social Networks Analysis and Mining (ASONAM), 2014 IEEE/ACM International Conference on*, pp. 620–627. IEEE.

Ferrara, E., O. Varol, C. Davis, F. Menczer, and A. Flammini (2014). The rise of social bots. *CoRR abs/1407.5225*.

Gao, H., J. Hu, C. Wilson, Z. Li, Y. Chen, and B. Y. Zhao (2010). Detecting and characterizing social spam campaigns. In *Proceedings of the 10th ACM SIGCOMM Conference on Internet Measurement*, IMC '10, New York, NY, USA, pp. 35–47. ACM.

Harris, D. (2013, Feb). Can evil data scientists fool us all with the world's best spam? `goo.gl/psEguf`.

Huang, J., R. Kornfield, G. Szczypka, and S. L. Emery (2014a). A cross-sectional examination of marketing of electronic cigarettes on twitter. *Tobacco control 23*(suppl 3), iii26–iii30.

Huang, J., R. Kornfield, G. Szczypka, and S. L. Emery (2014b). A cross-sectional examination of marketing of electronic cigarettes on twitter. *Tobacco control 23*(suppl 3), iii26–iii30.

Lee, K., J. Caverlee, and S. Webb (2010a). The social honeypot project: Protecting online communities from spammers. In *Proceedings of the 19th International Conference on World Wide Web*, WWW '10, New York, NY, USA, pp. 1139–1140. ACM.

Lee, K., J. Caverlee, and S. Webb (2010b). Uncovering social spammers: Social honeypots + machine learning. In *Proceedings of the 33rd International ACM SIGIR Conference on Research and Development in Information Retrieval*, SIGIR '10, New York, NY, USA, pp. 435–442. ACM.

Lee, K., B. D. Eoff, and J. Caverlee (2010). Devils, angels, and robots: Tempting destructive users in media. In W. W. Cohen and S. Gosling (Eds.), *ICWSM*. The AAAI Press.

Lee, K., B. D. Eoff, and J. Caverlee (2011). Seven months with the devils: a long-term study of content polluters on twitter. In *In AAAI Int?l Conference on Weblogs and Social Media (ICWSM*.

Mitchell, L., M. R. Frank, K. D. Harris, P. S. Dodds, and C. M. Danforth (2013). The geography of happiness: Connecting twitter sentiment and expression, demographics, and objective characteristics of place. *PloS one 8*(5), e64417.

Sadilek, A., H. A. Kautz, and V. Silenzio (2012). Modeling spread of disease from social interactions. In *ICWSM*.

Subrahmanian, V., A. Azaria, S. Durst, V. Kagan, A. Galstyan, K. Lerman, L. Zhu, E. Ferrara, A. Flammini, F. Menczer, et al. (2016). The darpa twitter bot challenge. *arXiv preprint arXiv:1601.05140*.

Thomas, K., C. Grier, and V. Paxson (2012). Adapting social spam infrastructure for political censorship. In *Presented as part of the 5th USENIX Workshop on Large-Scale Exploits and Emergent Threats*, Berkeley, CA. USENIX.

Thomas, K., C. Grier, D. Song, and V. Paxson (2011). Suspended accounts in retrospect: An analysis of twitter spam. In *Proceedings of the 2011 ACM SIGCOMM Conference on Internet Measurement Conference*, IMC '11, New York, NY, USA, pp. 243–258. ACM.

Wagner, C., S. Mitter, M. Strohmaier, and C. Korner. When social bots attack: Modeling susceptibility of users in online social networks.

Wagstaff, A. and A. J. Culyer (2012). Four decades of health economics through a bibliometric lens. *Journal of health economics 31*(2), 406–439.

Williams, J. R., J. P. Bagrow, C. M. Danforth, and P. S. Dodds (Accepted 17 March 2015). Text mixing shapes the anatomy of rank-frequency distributions. *Physical Review E (in press)*.

Wu, X., Z. Feng, W. Fan, J. Gao, and Y. Yu (2013). Detecting marionette microblog users for improved information credibility. In H. Blockeel, K. Kersting, S. Nijssen, and F. Yelezn (Eds.), *Machine Learning and Knowledge Discovery in Databases*, Volume 8190 of *Lecture Notes in Computer Science*, pp. 483–498. Springer Berlin Heidelberg.

Zhang, C. M. and V. Paxson (2011). Detecting and analyzing automated activity on twitter. In *Proceedings of the 12th International Conference on Passive and Active Measurement*, PAM'11, Berlin, Heidelberg, pp. 102–111. Springer-Verlag.

Chapter 3

Vaporous Marketing: Uncovering Pervasive Electronic Cigarette Advertisements on Twitter

Background: Twitter has become the "wild-west" of marketing and promotional strategies for advertisement agencies. Electronic cigarettes have been heavily marketed across Twitter feeds, offering discounts, "kid-friendly" flavors, algorithmically generated false testimonials, and free samples.

Methods: All electronic cigarette keyword related tweets from a 10% sample of Twitter spanning January 2012 through December 2014 (approximately 850,000 total tweets) were identified and categorized as Automated or Organic by combining a keyword classification and a machine trained Human Detection algorithm. A sentiment analysis using Hedonometrics was performed on Organic tweets to quantify the change in consumer sentiments over time. Commercialized tweets were topically categorized with key phrasal pattern matching.

Results: The overwhelming majority (80%) of tweets were classified as automated or promotional in nature. The majority of these tweets were coded as commercialized (83.65% in 2013), up to 33% of which offered discounts or free samples and appeared on over a billion twitter feeds as impressions. The positivity of Organic (human) classified tweets has decreased over time (5.84 in 2013 to 5.77 in 2014) due to a relative increase in the negative words ‘ban’, ‘tobacco’, ‘doesn’t’, ‘drug’, ‘against’, ‘poison’, ‘tax’ and a relative decrease in the positive words like ‘haha’, ‘good’, ‘cool’. Automated tweets are more positive than organic (6.17 versus 5.84) due to a relative increase in the marketing words like ‘best’, ‘win’, ‘buy’, ‘sale’, ‘health’, ‘discount’ and a relative decrease in negative words like ‘bad’, ‘hate’, ‘stupid’, ‘don’t’.

Conclusions: Due to the youth presence on Twitter and the clinical uncertainty of the long term health complications of electronic cigarette consumption, the protection of public health warrants scrutiny and potential regulation of social media marketing.

3.1 Introduction

Electronic Nicotine Delivery Systems, or e-cigs, have become a popular alternative to traditional tobacco products. The vaporization technology present in e-cigarettes allows consumers to simulate tobacco smoking without igniting the carcinogens found in tobacco (Cobb et al., 2010). Survey methods have revealed widespread awareness of e-cigarette products (Zhu et al., 2013; Pearson et al., 2012). The health risks (Vansickel et al., 2010; Goniewicz et al., 2014; Callahan-Lyon, 2014; Kosmider et al., 2014), marketing regulations (Trtchounian and Talbot, 2011), and the potential of these devices as a form of nicotine replacement therapy (Kandra et al., 2014; Grana et al., 2014; Eissenberg, 2010) are hotly debated politically (**?**) and investigated clinically (Palazzolo, 2013; Avdalovic and Murin, 2012). The CDC reports that more people in the US are addicted to nicotine than any other drug and that nicotine may be as addictive as heroin, cocaine, and alcohol (CONTROL et al., 2014; National Institute on Drug Abuse, 2012; American Society of Addiction Medicine., 2008; US Department of Health and Human Services and others, 2010). Nicotine addiction is extremely difficult to quit, often requiring more than one attempt (US Department of Health and Human Services and others, 2010, 2000), however nearly 70% of smokers in the US want to quit (Centers for Disease Control and Prevention (CDC and others, 2011). Data mining can provide valuable insight into marketing strategies, varieties of e-cigarette brands, and their use by consumers (Kim et al., 2015; Yip and Talbot, 2013; Grana and Ling, 2014; Zhu et al., 2014; Aphinyanaphongs et al., 2016).

Twitter, a mainstream social media outlet comprising over 230 million active accounts, provides a means to survey the popularity and sentiment of consumer opinions regarding e-cigarettes over time. Individuals post tweets which are short text based messages restricted to 140 characters (at the time of this study). Using data mining techniques, roughly $850,000$ tweets containing mentions of e-cigarettes were collected from a 10% sample of Twitter's garden hose feed spanning from January 2012 though December 2014. This analysis extends a preliminary study (Huang et al., 2014a) which analyzed all e-cigarette related tweets spanning May through June 2012.

As Twitter has become a mainstream social media outlet, it has become increasingly enticing for third parties to gamify the system by creating self-tweeting automated software to send messages to organic (human) accounts as a means for personal gain and for influence manipulation (Harris, 2013). We recently introduced a classification algorithm that is based upon three linguistic attributes of an individual's tweets (Clark et al., 2015). The algorithm analyzes the average hyperlink (URL) count per tweet, the average pairwise dissimilarity between an individual's tweets, and the unique word introduction decay rate of an individual's tweets.

All tweets mentioning e-cigarettes were categorized using a two-tier classification process. Tweets containing an abundance of marketing slang ('free trial', 'starter kit', 'coupon') are immediately categorized as automated. All of the tweets from individuals that have mentioned an e-cigarette keyword are collected in order to classify the remaining tweets per individual as either organic or automated. The machine learning classifier was trained on the natural linguistic cues from human accounts to identify promotional and SPAM entities by exclusion.

The manipulative effects, agendas, and ecosystem of generalized social media marketing campaigns have been identified and extensively studied (Lee et al., 2013; Ranganath, Hu, Tang, and Liu, Ranganath et al.; Wang et al., 2012). Other work, (Chu et al., 2012), has distinguished between purely automated accounts, or "robots", and human assisted automated accounts referred to as "cyborgs". On Twitter, these campaigns have also been characterized using Markov Random Fields to classify accounts as either promotional or organic (Li et al., 2014). This study was able to achieve very high classification accuracy, but was working under a much shorter time frame (1 month) and was trained on all relevant tweets authored within this time window. Our study compiled a 10% sample of tweets over a three-year period, so we relied on a classifier that was trained on smaller samples of tweets per individual.

The emotionally charged words that contribute to the positivity of various subsets of tweets from each category were quantitatively measured using hedonometrics (Dodds et al., 2011, 2015). Outliers in both the positivity and frequency time-series distributions correspond to political debates regarding the regulation of e-cigarettes. Recent studies(Dutra and Glantz, 2014; Cho et al., 2011; Pepper et al., 2013; Goniewicz and Zielinska-Danch, 2012; Wills et al., 2015) report an alarmingly rapid increase in the youth awareness and consumption of electronic cigarettes; a Michigan study found that the use of e-cigarettes surpass tobacco cigarettes among teens (Johnston et al., 2014). The CDC reports that "the number of never-smoking youth increased three-fold from approximately 79,000 in 2011 to 263,000 in 2013" (Bunnell et al., 2014). During this time-period there has also been a substantial (256%) increase in youth exposure to electronic cigarette television marketing campaigns (Duke et al., 2014). Due to the high youth presence on Twitter (Brenner and Smith, 2013) as well as the clinical uncertainty regarding the risks associated with e-cigarettes, understanding the effect of promotionally marketing vaporization products across social media should be immediately relevant to public health and policy makers.

3.2 Materials and Methods

3.2.1 Data Collection

An exhaustive search from the 10% "garden hose" random sample from Twitter's streaming API spanning 2012 through 2014 yielded approximately 850,000 tweets mentioning a keyword related to electronic cigarettes including: e(-)cig, e(-)cigarette, electronic cigarette, etc. All tweets were tokenized by removing punctuation and performing a case insensitive pattern match for keywords. Using time zone meta-data the tweets were converted into their local post time, in order for a more accurate ordinal sentiment analysis. The language, reported by Twitter, and user features were also collected and analyzed. The data from our study was collected via a program developed by Dodds et al, that pings Twitter's streaming API and complies with Twitter's Terms of Service. Our study collected each account's unique twitter user identification number in order to classify them as either Automated or Organic, however our published data has been anonymized by replacing Twitter's UserIDs with placeholder values.

3.2.2 Automation Classification

As reported in (Huang et al., 2014a) there is a high prevalence of automation among e-cigarette related tweets. Many of these messages were promotional in nature, offering discounted or free samples or advertising specific electronic cigarette paraphernalia. A human detection algorithm defined and tested in (Clark et al., 2015) was implemented to classify accounts as either automated or organic (human in nature). The original classifier was trained on 1000 accounts - 752 were verified as humans and 248 as automated accounts. The classifier operates by isolating organic linguistic characteristics and identifies automated accounts by exclusion. All tweets from each individual appearing in our dataset were collected for the classifier. For each individual, the average URL count, average tweet dissimilarity, and word introduction decay rate were calculated for the individuals with at least 25 sampled tweets.

The majority (94%) of commercial e-cigarette tweets collected by (Huang et al., 2014a) contain a hyperlink (URL). The average URL count per tweet has been demonstrated to be a strong feature for detecting robotic accounts (Chu et al., 2010; Lee et al., 2010a,b). Many algorithmically generated tweets contain similar structures with minor character replacements and long chains of common substrings, as opposed to Organic content. The Pairwise Tweet Dissimilarity of tweets t_i, t_j from a particular individual was estimated by subtracting the length (number of characters) of the longest common subsequence, $|LCS(t_i, t_j)|$ from the length of both tweets, $|t_i| + |t_j|$ and normalizing by the total length of both tweets:

$$\mathrm{D}(t_i, t_j) = \frac{|t_i| + |t_j| - 2 \cdot |LCS(t_i, t_j)|}{|t_i| + |t_j|}.$$

For example, given the two tweets:
(t_1, t_2) = (I love tweeting, I love spamming). Then $|t_1|$ = 16, $|t_2|$ = 15, $LCS(t_1, t_2) =$ |I love | $=$ 7 (including whitespace) and we calculate the pairwise tweet dissimilarity as:

$$D(t_1, t_2) = \frac{16 + 15 - 2 \cdot 7}{16 + 15} = \frac{17}{31}.$$

The average tweet dissimilarity of the individual was then estimated by finding the arithmetic mean of each individual's calculated pairwise tweet dissimilarity. Since automated and promotional accounts have a structured and limited vocabulary, the unique word introduction decay rate introduced in (Williams et al.,

2015) serves as another useful attribute to detect automated accounts. Using these attributes, the calibrated human detection algorithm, tested in (Clark et al., 2015), detected over 90% of automated accounts from a mixed 1000 user sample with less than a 5% false positive rate.

The Human Detection Algorithm was calibrated for a range of tweet sample sizes from hand classified Organic accounts. Ordinal samples of collected tweets from each account were binned into partitions of 25 ranging from 25 to a maximum of 500 tweets. Table 1 below lists the number of automated and organic classified accounts per year. Individuals with less than 25 sampled tweets were not classified with the detection algorithm.

To benchmark the accuracy of the detection algorithm on this sample of tweets, a random sample of 500 accounts algorithmically classified as automatons and 500 classified as Organic were hand classified. All collected tweets were hand coded by two evaluators. Tweets were reviewed until the evaluator noticed the presence of automation. If no subset of tweets appeared to be algorithmically generated, the individual was coded as human. Both evaluators had prior experience distinguishing algorithmic versus organic tweets. Refer to the supplementary materials in (Clark et al., 2015) for a detailed explanation of this annotation process.

In Figure 1, features of each of these 1000 sampled individuals are plotted in three dimensions. Organic features (green) are densely distributed, while the automated features (red points) are more dispersed. The black lines illustrates the organic feature cutoff for the classifier; individuals with features falling outside of the box are classified as automatons. On this sampled set of accounts, the classification algorithm exhibited a 94.6% True Positive rate with a 12.9% False Positive Rate.

3.2.3 Categorization by Topics

Tweets with at least 3 advertising jargon references (e.g. coupon, starter kit, free trial) were immediately classified as automated. All posts from users with at least 10 marketing classified tweets were also flagged as automated. As noted in (Huang et al., 2014a), some Organic users could retweet promotional content for rewards (e.g. winning free samples or discounts). All of these tweets were still classified as automated, but the user was not flagged as such. The remaining tweets were classified as either automated or organic by the human detection algorithm. Posts from users who had an insufficient number of sampled tweets (< 25) to algorithmically classify and who hadn't posted commercial content were classified as Organic. Due to the

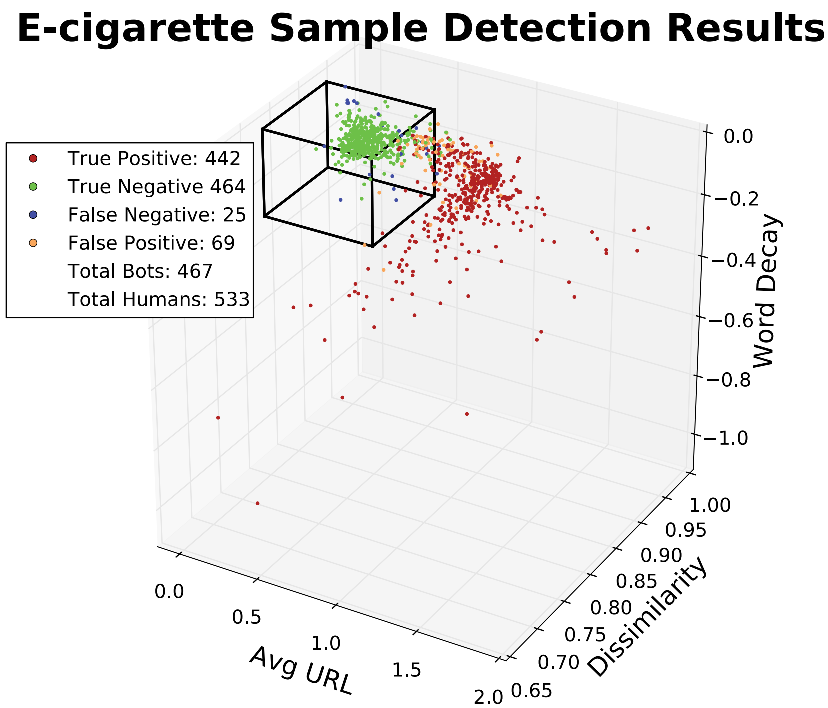


Figure 3.1: Tweets from a random sample of 500 organic classified and 500 automated classified accounts were hand coded to gauge the accuracy of the detection algorithm. The feature set of each sampled individual is plotted in three dimensions. The traced box indicate the organic feature cutoff. True Positives (red) are correctly identified automatons, True Negatives (green) are correctly identified Humans, False Negatives (blue) are automatons classified as humans and False Positives (orange) are humans classified as automatons.

Table 3.1: Electronic Cigarette Tweet Category Counts and Twitter Account Classification

Year	Tweet Categorization				Account Classification		
	Total	Automated	Organic	Discarded	Automated	Organic	N/A*
2012	107,918	85,546	13,492	8,880	12,715	12,052	19,512
2013	426,306	339,111	76,037	11,158	64,874	59,376	120,142
2014	316,424	234,972	68,698	12,754	54,033	63,289	48,528

*Accounts with less than 25 tweets were not classified.

high prevalence of hyperlinks included in tweets from promotional accounts, Tweets with URLs whose user had insufficient tweets to classify algorithmically were discarded (3.85% total tweets). A final list with each tweet classification coding is created by merging the commercial keyword classification with the results from the Human Detection Algorithm.

3.3 Results and Discussion

The number of automated, and in particular promotional, tweets vastly overwhelm (80.7%) the organic (see Figure 2). The identified automated accounts tweet e-cigarette content with much higher frequency than the Organic users. The average number of automated tweets per user was 1.96 with a standard deviation of 35.06 and a max of 14,310. Average organic posts per user were 1.44 with a standard deviation of 4.01 and max of 356 tweets. A total of 607,446 Automated Tweets provided a URL (92.09%).

Frequency WordClouds (see Figure 2) illustrate the most frequently used words by the Automated category. The size of the text reflects the ranked word frequencies. Marketing key words (Free Trial, Brand, Starter Kit, win, Sale) and brand names (V2, Apollo) are prevalent, illustrating commercial intent. Many automated tweets also refer to the health benefits of switching to electronic cigarettes (#EcigsSaveLives), even though they have not been officially approved as such by the Food and Drug Administration, (Zezima, 2009; Ashley et al., 2007). See Table 2 for sub categorical counts of the automated tweets.

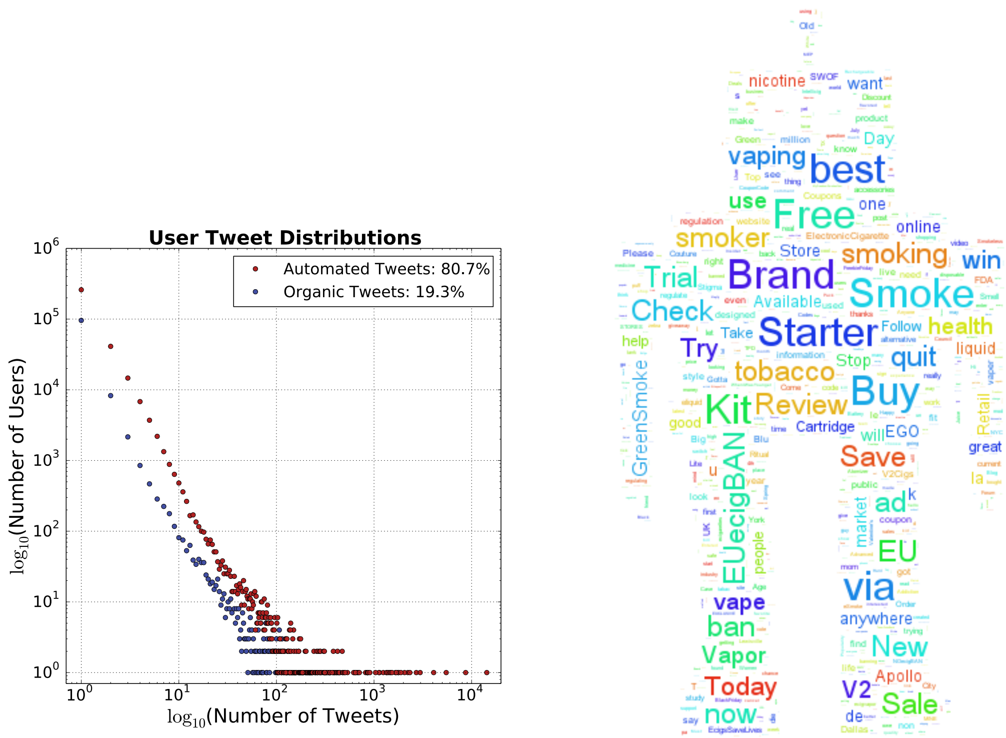


Figure 3.2: Left: Binned User E-cigarette Keyword Tweet Distribution (2012-2014). Right: 2013 Automated Tweet Rank-Frequency Word Cloud. High frequency stop words ('of','the', etc.) are removed from the rank-frequency word distribution.

3.3.1 Tweet Sentiment Analysis

Hedonometrics are performed on the organic subset of electronic cigarette tweets to quantify the change in user sentiments over time. Using the happiness scores of English words from LabMT (Dodds et al., 2011), along with its multi-language companion (Dodds et al., 2015) the average emotional rating of a corpus is calculated by tallying the appearance of words found in the intersection of the word-happiness distribution and a given corpus, in this case subsets of tweets. A weighted arithmetic mean of each word's frequency, f_{word}, and corresponding happiness score, h_{word} for each of the N words in a text yields the average happiness score for the corpus, $\bar{h}_{\text{text}}$:

$$\bar{h}_{\text{text}} = \frac{\sum_{w=1}^{N} f_{\text{w}} \cdot h_{\text{w}}}{\sum_{w=1}^{N} f_{\text{w}}}$$

The average happiness of each word, h_{avg} lies on a 9 point scale: 1 is extremely negative and 9 is extremely positive. Neutral words ($4 \leq h_{avg} \leq 6$), aka 'stop words', were removed from the analysis to bolster the emotional signal of each set of tweets.

Figure 3 shows that automated electronic cigarette tweets are using very positive language to promote their products. The average happiness of the Organic tweets are much more stable, and are becoming slightly more negative over time. Both distributions have a sudden drop in positivity during December 2013, around a debate regarding new e-cigarette legislation by the European Union. These tweets, labeled #EuEcigBan, are investigated separately in the next section. The words that have the largest contributions to changes in sentiments are investigated with Word-shift graphs.

Word-shift graphs, introduced in (Dodds et al., 2011), illustrate the words causing an emotional shift between two word frequency distributions. A reference period (T_{ref}), creates a basis of the emotional words being used to compare with another period, (T_{comp}). The top 50 words responsible for a happiness shift between the two periods are displayed, along with their contribution to shifting the average happiness of the tweet-set. The arrows ($\uparrow, \downarrow$) next to a word indicate an increase or decrease, respectively, of the word's frequency during the comparison period with respect to the reference period. The addition and subtraction signs indicate if the word contributes positively or negatively, respectively, to the average happiness score.

Marketing accounts that delivered personalized advertising by attempting to impersonate organic users were prevalent among these commercial entities. These accounts, along with the traditional marketing robots, were diluting the data with extremely positive sentiments regarding their products. Using hedonometrics, we distinguish the emotionally charged words that influence a shift in computed average word happiness between these types of accounts. The sentiment analysis helps to characterize the thematic differences between Organic and Automated entities.

In Figure 3, below, Word-shift graphs compare the change in Organic sentiments over time, as well as the difference in sentiments between automated and organic tweets. On the left, the 2013 Organic Tweet distribution is used as a reference to compare sentiments from 2014 Organic Tweets. December 2013 and January 2014 are removed to dampen the effect of tweets mentioning the #EUecigBan (see S1 Fig). The average happiness score decreases from 5.84 in 2013 to 5.77 in 2014. This decrease in the average happiness score is due to a relative increase in the negative words ‘ban’, ‘tobacco’, ‘doesn’t’, ‘drug’, ‘against’, ‘poison’, ‘tax’; a relative decrease in the positive words ‘haha’, ‘good’, ‘cool’. Notably, there is also relatively less usage of the words ‘quit’, ‘addicted’, and an increase in ‘health’, ‘kids’, ‘juice’. On the right, Organic tweets from 2013 is the reference distribution to compare Automated tweets from the same year. Automated tweets are more positive (6.17-6.59 versus 5.84) due to a relative increase in the marketing words ‘best’, ‘win’, ‘buy’, ‘sale’, ‘health’, ‘discount’, etc and a relative decrease in the negative words ‘bad’, ‘hate’, ‘stupid’, ‘don’t’, among others.

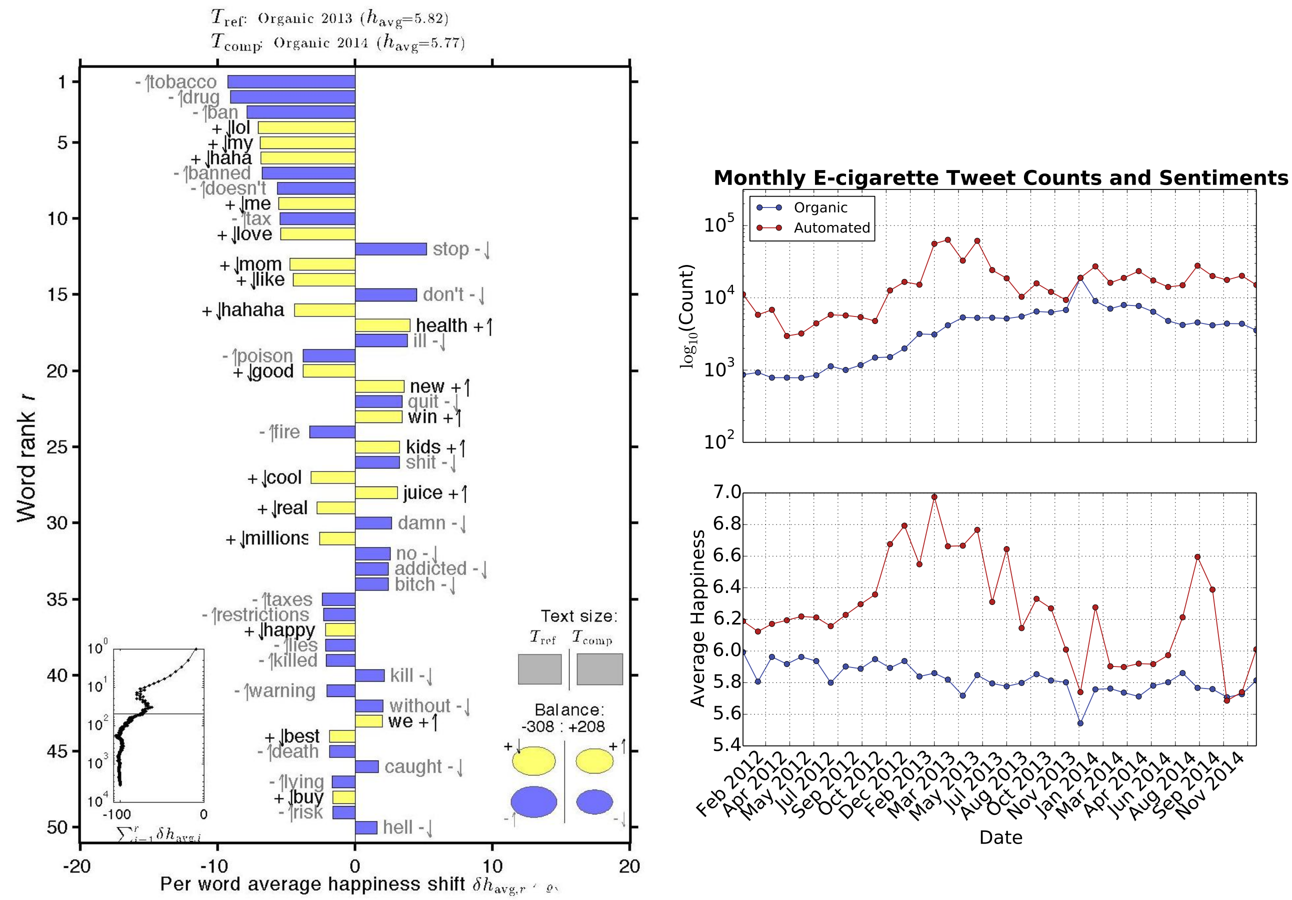


Figure 3.3: Categorical Tweet Word-shift Graphs: On the left, Organic Tweets from 2013 are the reference distribution to compare sentiments of Organic Tweets made in 2014 where we see a negative shift in the calculated average word happiness. Due to tweets tagged #EUEcig Ban, January 2014 and December 2013 are omitted. The computed average happiness (h_{avg}) decreases from 5.82 to 5.77 due to both an increase in the negative words ‘tobacco’, ‘drug’, ‘ban’, ‘poison’, and a decrease in the positive words ‘love’, ‘like’, ‘haha’, ‘cool’ among others. On the right, Organic Tweets from 2013 are the reference distribution to compare Automated Tweets from 2013. The words ‘free’ and ‘trial’ are excluded from the graph, since their high frequency and happiness scores distorts the image. With these key words included the the automated tweet h_{avg} increases from 6.17 to 6.59.

3.3.2 Sub-Categorical Tweet Topics

Pertinent topics related to e-cigarette marketing regulation include kid-friendly flavors, smoking cessation claims, and price reduction (including free trials, and starter kits). The commercialized, smoking cessation claims, and discounts were primary topics in the foundational study (Huang et al., 2014b) that identified these campaigns over a 2 month time window. We included the kid-friendly flavors topic in this list due to recent studies reporting their prevalence (Grana et al., 2014; Zhu et al., 2014) as well as its current spotlight in political controversy.

Keywords from each of these topics are used to sub-classify the automated tweet set per year, see Table 2 below. Purely commercial tweets were those with any marketing keywords including: 'buy', 'save', 'coupon(s)', 'discount', 'price', 'cost', 'deal', 'promo', 'money', 'sale', 'purchase', 'offer', 'review', 'code', 'win(ner)', 'free', 'starter kit(s)', 'premium'. The URL from each tweet was also analyzed for promotional keywords. Any URL with at least three mentions of the above keywords was enough to classify the tweet as commercial.

When an individual on Twitter 'follows' another account, posts from these users appear on the 'timeline' of the individual. We quantify the social reach of each of these sub-categorical tweets by counting the total number of accounts' 'timelines' who could have been exposed to the advertisement. To approximate this, we sum the number of followers from each individual's tweets. The total number of impressions from the commercial category increases from 195.25 million to 951.03 million between 2013 to 2014, even though the total count has dropped from 283k to 149k. This implies that promotional accounts that are successful in deceiving Twitter's SPAM detector may be gaining many more social links to broadcast their commercial context.

In order to gauge the accuracy of these sub-categorical tweet topics, 500 tweets were randomly sampled from each category and were evaluated separately by two people to determine the relevance of the tweet to its categorization. The evaluators had a high level of concordance (84.8%) and the discrepancies were resolved and merged into a final list. Sampled tweets were highly relevant per category, the percentage for each is given in Table 2 below.

Many automated tweets mentioned using electronic cigarettes as a cessation device, or as a safe alternative. Over $20,000$ tweets were classified as cessation related, which potentially appeared on over 76.8

Table 3.2: Automated Tweet Subcategory Counts

Subcategory	Count	Percentage	Impressions	Relevance*	Year
Commercial	53,471	62.51%	59.74M	88.4%	‘12
	283,677	83.65%	195.25M		‘13
	149,333	63.55%	951.03M		‘14
Cessation	6,392	7.47%	8.59M	90.8%	‘12
	6,599	1.95%	25.64M		‘13
	8,386	3.57%	42.72M		‘14
Discount	26,596	31.09%	27.02M	89.8%	‘12
	112,720	33.24%	38.21M		‘13
	37,735	16.06%	160.49M		‘14
Flavor	1,685	1.97%	2.24M	81%	‘12
	2,715	0.80%	4.79M		‘13
	6,133	2.61%	17.51M		‘14

Table 3.3: *
*Relevant percentage of 500 randomly sampled tweets

million individual’s Twitter feed as impressions. Although electronic cigarettes have not been conclusively authorized as an effective cessation device, (Eissenberg, 2010) has demonstrated the infectiveness of electronic cigarettes to suppress nicotine cravings. It is also notable that these affiliate marketing accounts are advertising electronic cigarettes as a completely safe alternative to analog tobacco use, contrary to recent studies (Sussan et al., 2015; CA et al., 2015; Cameron et al., 2014; Williams et al., 2013). Cessation tweets were tallied using the keywords ‘quit’, ‘quitting’, ‘stop smoking’, ‘smoke free’, ‘safe’, ‘safer’, ‘safest’. Many of the purely commercialized tweets mentioned discounts or even free samples. These Discount tweets were categorized with the keywords ‘free trial’, ‘coupon(s)’, ‘discount(s)’, ‘save’, ‘sale’, ‘free (e)lectronic (cig)arette’. Tweets advertising flavors were tallied using the keywords ‘flavor(s)’ and ‘flavour(s)’ along with an extensive list of popular electronic cigarette flavors compiled from a distributor’s website (`https://crazyvapors.com/e-liquid-flavor-list/`).

A noteworthy class of E-cigarette commercial-bots, are those that are masquerading as Organic users to spam pseudo-positive messages towards potential consumers. These “cyborgs”, as defined in (Chu et al., 2010; Clark et al., 2015; Li et al., 2014), spam a positive message regarding a personal experience. One class of these automatons are sending contrived testimonies that e-cigarettes have successfully allowed them to quit smoking cigarettes. These messages are very intentionally structured and tend to swap a few words to appear organic. These messages also target specific individuals as a more personal form of marketing. The general tweet structure from a sample cyborg marketing strategy is given below:

@USER {I,We} {tried,pursued} to {give up, quit} smoking . Discovered BRAND electronic cigarettes and quit in {#} weeks. {Marvelous,Amazing,Terrific}! URL

@USER It's now really easy to {quit,give up} smoking (cigarettes). - these BRAND electronic cigarettes are lots of {fun,pleasure}! URL

@USER electronic cigarettes can assist cigarette smokers to quit, it's well worth the cost URL

@USER It's {incredible,amazing} - the (really) {easy,painless} {answer,method} to quit cigarette smoking through BRAND electronic cigarettes URL

I managed to quit smoking with these e-cigarettes, I highly recommend them: URL @USER

@USER Its {amazing, extraordinary} - I (really) quit smoking after {#} yrs thanks to BRAND electronic cigarettes! URL

Using cyborgs to mimic Organic Users for marketing purposes should be analyzed heavily, to gauge their impact and effectiveness on consumers.

3.4 Conclusion

Our study has identified an abundance of automated, and in particular, promotional tweets, and consequent organic sentiments. The collected categorized tweet data from this analysis is available for follow-up analyses into e-cigarette social media marketing campaigns. Future work can perform a deeper analysis on the URL content, similar to (Grana and Ling, 2014), posted by promotional accounts to get a better sense of the smoking cessation, flavor mentions, and discount prevalence. We take care not to downplay the well recognized health benefits from smoking cessation including: decreased risk of coronary artery disease, cerebrovascular disease, peripheral vascular disease, decreased incidence of respiratory symptoms such as cough, wheezing, shortness of breath, decreased incidence of chronic obstructive pulmonary disease, and decreased risk of infertility in women of childbearing age (CONTROL et al., 2014; US Department of Health and Human Services and others, 2010, 2004). The greatest concern of promotional e-cigarette marketing on Twitter is the risk of enticing younger generations who otherwise may never have commenced consuming nicotine. Due to the unknown but potential long-term adverse health effects of electronic cigarettes and the alarmingly increased youth consumption, monitoring and potentially regulating social media commercialization of these products should be immediately relevant to public health and policy agendas.

3.5 References

American Society of Addiction Medicine. (2008). Public policy statement on nicotine addiction and tobacco. *Chevy Chase (MD): American Society of Addiction Medicine*.

Aphinyanaphongs, Y., A. Lulejian, D. P. Brown, R. Bonneau, and P. Krebs (2016). Text classification for automatic detection of e-cigarette use and use for smoking cessation from twitter: A feasibility pilot. In *Pacific Symposium on Biocomputing. Pacific Symposium on Biocomputing*, Volume 21, pp. 480. NIH Public Access.

Ashley, D., D. Burns, M. Djordjevic, E. Dybing, N. Gray, S. Hammond, J. Henningfield, M. Jarvis, K. Reddy, C. Robertson, et al. (2007). The scientific basis of tobacco product regulation. *World Health Organization technical report series* (951), 1–277.

Avdalovic, M. V. and S. Murin (2012). Electronic cigarettes: no such thing as a free lunch? or puff. *CHEST Journal 141*(6), 1371–1372.

Brenner, J. and A. Smith (2013). 72% of online adults are social networking site users. *Washington, DC: Pew Internet & American Life Project*.

Bunnell, R. E., I. T. Agaku, R. Arrazola, B. J. Apelberg, R. S. Caraballo, C. G. Corey, B. Coleman, S. R. Dube, and B. A. King (2014). Intentions to smoke cigarettes among never-smoking us middle and high school electronic cigarette users, national youth tobacco survey, 2011-2013. *Nicotine & Tobacco Research*, ntu166.

CA, L., S. IK, Y. H, G. J, O. DJ, and et al. (2015). Vapors produced by electronic cigarettes and e-juices with flavorings induce toxicity, oxidative stress, and inflammatory response in lung epithelial cells and in mouse lung. *PloS one 10*(2), e0116732.

Callahan-Lyon, P. (2014). Electronic cigarettes: human health effects. *Tobacco control 23*(suppl 2), ii36–ii40.

Cameron, J. M., D. N. Howell, J. R. White, D. M. Andrenyak, M. E. Layton, and J. M. Roll (2014). Variable and potentially fatal amounts of nicotine in e-cigarette nicotine solutions. *Tobacco control 23*(1), 77–78.

Centers for Disease Control and Prevention (CDC and others (2011). Quitting smoking among adults–united states, 2001-2010. *MMWR. Morbidity and mortality weekly report 60*(44), 1513.

Cho, J. H., E. Shin, and S.-S. Moon (2011). Electronic-cigarette smoking experience among adolescents. *Journal of Adolescent Health 49*(5), 542–546.

Chu, Z., S. Gianvecchio, H. Wang, and S. Jajodia (2010). Who is tweeting on twitter: Human, bot, or cyborg? In *Proceedings of the 26th Annual Computer Security Applications Conference*, ACSAC '10, New York, NY, USA, pp. 21–30. ACM.

Chu, Z., S. Gianvecchio, H. Wang, and S. Jajodia (2012). Detecting automation of twitter accounts: Are you a human, bot, or cyborg? *Dependable and Secure Computing, IEEE Transactions on 9*(6), 811–824.

Clark, E. M., J. R. Williams, R. A. Galbraith, C. M. Danforth, P. S. Dodds, and C. A. Jones (2015). Sifting robotic from organic text: A natural language approach for detecting automation on twitter. *arXiv preprint arXiv:1505.04342*.

Cobb, N. K., M. J. Byron, D. B. Abrams, and P. G. Shields (2010). Novel nicotine delivery systems and public health: the rise of the ?e-cigarette? *American journal of public health 100*(12), 2340–2342.

CONTROL, C. F. D., PREVENTION, et al. (2014). The health consequences of smoking–50 years of progress: A report of the surgeon general. *Rockville, MD: US DEPARTMENT OF HEALTH AND HUMAN SERVICES*, 171.

Dodds, P. S., E. M. Clark, S. Desu, M. R. Frank, A. J. Reagan, J. R. Williams, L. Mitchell, K. D. Harris, I. M. Kloumann, J. P. Bagrow, K. Megerdoomian, M. T. McMahon, B. F. Tivnan, and C. M. Danforth (2015). Human language reveals a universal positivity bias. *Proceedings of the National Academy of Sciences 112*(8), 2389–2394.

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

Duke, J. C., Y. O. Lee, A. E. Kim, K. A. Watson, K. Y. Arnold, J. M. Nonnemaker, and L. Porter (2014). Exposure to electronic cigarette television advertisements among youth and young adults. *Pediatrics 134*(1), e29–e36.

Dutra, L. M. and S. A. Glantz (2014). Electronic cigarettes and conventional cigarette use among us adolescents: a cross-sectional study. *JAMA pediatrics 168*(7), 610–617.

Eissenberg, T. (2010). Electronic nicotine delivery devices: ineffective nicotine delivery and craving suppression after acute administration. *Tobacco control 19*(1), 87–88.

Goniewicz, M. L., J. Knysak, M. Gawron, L. Kosmider, A. Sobczak, J. Kurek, A. Prokopowicz, M. Jablonska-Czapla, C. Rosik-Dulewska, C. Havel, et al. (2014). Levels of selected carcinogens and toxicants in vapour from electronic cigarettes. *Tobacco control 23*(2), 133–139.

Goniewicz, M. L. and W. Zielinska-Danch (2012). Electronic cigarette use among teenagers and young adults in poland. *Pediatrics 130*(4), e879–e885.

Grana, R., N. Benowitz, and S. A. Glantz (2014). E-cigarettes a scientific review. *Circulation 129*(19), 1972–1986.

Grana, R. A. and P. M. Ling (2014). Smoking revolution: a content analysis of electronic cigarette retail websites. *American journal of preventive medicine 46*(4), 395–403.

Harris, D. (2013, Feb). Can evil data scientists fool us all with the world's best spam? `goo.gl/psEguf`.

Huang, J., R. Kornfield, G. Szczypka, and S. L. Emery (2014a). A cross-sectional examination of marketing of electronic cigarettes on twitter. *Tobacco control 23*(suppl 3), iii26–iii30.

Huang, J., R. Kornfield, G. Szczypka, and S. L. Emery (2014b). A cross-sectional examination of marketing of electronic cigarettes on twitter. *Tobacco control 23*(suppl 3), iii26–iii30.

Johnston, L. D., J. G. Bachman, et al. (2014). Monitoring the future: National results on adolescent drug use: Overview of key findings.

Kandra, K. L., L. M. Ranney, J. G. Lee, and A. O. Goldstein (2014). Physicians? attitudes and use of e-cigarettes as cessation devices, north carolina, 2013. *PloS one 9*(7), e103462.

Kim, A. E., T. Hopper, S. Simpson, J. Nonnemaker, A. J. Lieberman, H. Hansen, J. Guillory, and L. Porter (2015). Using twitter data to gain insights into e-cigarette marketing and locations of use: An infoveillance study. *Journal of medical Internet research 17*(11).

Kosmider, L., A. Sobczak, M. Fik, J. Knysak, M. Zaciera, J. Kurek, and M. L. Goniewicz (2014). Carbonyl compounds in electronic cigarette vapors: effects of nicotine solvent and battery output voltage. *Nicotine & Tobacco Research 16*(10), 1319–1326.

Lee, K., J. Caverlee, and S. Webb (2010a). The social honeypot project: Protecting online communities from spammers. In *Proceedings of the 19th International Conference on World Wide Web*, WWW '10, New York, NY, USA, pp. 1139–1140. ACM.

Lee, K., J. Caverlee, and S. Webb (2010b). Uncovering social spammers: Social honeypots + machine learning. In *Proceedings of the 33rd International ACM SIGIR Conference on Research and Development in Information Retrieval*, SIGIR '10, New York, NY, USA, pp. 435–442. ACM.

Lee, K., P. Tamilarasan, and J. Caverlee (2013). Crowdturfers, campaigns, and social media: Tracking and revealing crowdsourced manipulation of social media. In *ICWSM*.

Li, H., A. Mukherjee, B. Liu, R. Kornfield, and S. Emery (2014). Detecting campaign promoters on twitter using markov random fields. In *Data Mining (ICDM), 2014 IEEE International Conference on*, pp. 290–299. IEEE.

National Institute on Drug Abuse (2012). Research report series: Is nicotine addictive? *Bethesda (MD): National Institutes of Health, National Institute on Drug Abuse*.

Palazzolo, D. L. (2013). Electronic cigarettes and vaping: a new challenge in clinical medicine and public health. a literature review. *Frontiers in public health 1*.

Pearson, J. L., A. Richardson, R. S. Niaura, D. M. Vallone, and D. B. Abrams (2012). e-cigarette awareness, use, and harm perceptions in us adults. *American journal of public health 102*(9), 1758–1766.

Pepper, J. K., P. L. Reiter, A.-L. McRee, L. D. Cameron, M. B. Gilkey, and N. T. Brewer (2013). Adolescent males' awareness of and willingness to try electronic cigarettes. *Journal of Adolescent Health 52*(2), 144–150.

Ranganath, S., X. Hu, J. Tang, and H. Liu. Understanding and identifying advocates for political campaigns on social media.

Sussan, T. E., S. Gajghate, R. K. Thimmulappa, J. Ma, J.-H. Kim, K. Sudini, N. Consolini, S. A. Cormier, S. Lomnicki, F. Hasan, et al. (2015). Exposure to electronic cigarettes impairs pulmonary anti-bacterial and anti-viral defenses in a mouse model. *PloS one 10*(2), e0116861.

Trtchounian, A. and P. Talbot (2011). Electronic nicotine delivery systems: is there a need for regulation? *Tobacco control 20*(1), 47–52.

US Department of Health and Human Services and others (2000). *Reducing tobacco use: a report of the Surgeon General*. US Department of Health and Human Services.

US Department of Health and Human Services and others (2004). The health consequences of smoking: a report of the surgeon general. *Atlanta, GA: US Department of Health and Human Services, Centers for Disease Control and Prevention, National Center for Chronic Disease Prevention and Health Promotion, Office on Smoking and Health 62*.

US Department of Health and Human Services and others (2010). How tobacco smoke causes disease: the biology and behavioral basis for smoking-attributable disease: a report of the surgeon general. *Atlanta, GA: US Department of Health and Human Services, Centers for Disease Control and Prevention, National Center for Chronic Disease Prevention and Health Promotion, Office on Smoking and Health 2*.

Vansickel, A. R., C. O. Cobb, M. F. Weaver, and T. E. Eissenberg (2010). A clinical laboratory model for evaluating the acute effects of electronic cigarettes: nicotine delivery profile and cardiovascular and subjective effects. *Cancer Epidemiology Biomarkers & Prevention 19*(8), 1945–1953.

Wang, G., C. Wilson, X. Zhao, Y. Zhu, M. Mohanlal, H. Zheng, and B. Y. Zhao (2012). Serf and turf: crowdturfing for fun and profit. In *Proceedings of the 21st international conference on World Wide Web*, pp. 679–688. ACM.

Williams, J. R., J. P. Bagrow, C. M. Danforth, and P. S. Dodds (Accepted 17 March 2015). Text mixing shapes the anatomy of rank-frequency distributions. *Physical Review E (in press)*.

Williams, M., A. Villarreal, K. Bozhilov, S. Lin, and P. Talbot (2013). Metal and silicate particles including nanoparticles are present in electronic cigarette cartomizer fluid and aerosol. *PloS one 8*(3), e57987.

Wills, T. A., R. Knight, R. J. Williams, I. Pagano, and J. D. Sargent (2015). Risk factors for exclusive e-cigarette use and dual e-cigarette use and tobacco use in adolescents. *Pediatrics 135*(1), e43–e51.

Yip, H. and P. Talbot (2013). Mining data on usage of electronic nicotine delivery systems (ends) from youtube videos. *Tobacco Control 22*(2), 103–106.

Zezima, K. (2009). Cigarettes without smoke, or regulation.

Zhu, S.-H., A. Gamst, M. Lee, S. Cummins, L. Yin, and L. Zoref (2013). The use and perception of electronic cigarettes and snus among the us population. *PloS one 8*(10), e79332.

Zhu, S.-H., J. Y. Sun, E. Bonnevie, S. E. Cummins, A. Gamst, L. Yin, and M. Lee (2014). Four hundred and sixty brands of e-cigarettes and counting: implications for product regulation. *Tobacco control 23*(suppl 3), iii3–iii9.

Chapter 4

A Sentiment Analysis of Breast Cancer Treatment Experiences and Healthcare Perceptions Across Twitter

Background: Social media has the capacity to afford the healthcare industry with valuable feedback from patients who reveal and express their medical decision-making process, as well as self-reported quality of life indicators both during and post treatment. In prior work, (Crannell et al., 2016), we have studied an active cancer patient population on Twitter and compiled a set of tweets describing their experience with this disease. We refer to these online public testimonies as “Invisible Patient Reported Outcomes” (iPROs), because they carry relevant indicators, yet are difficult to capture by conventional means of self-report.
Methods: Our present study aims to identify tweets related to the patient experience as an additional informative tool for monitoring public health. Using Twitter’s public streaming API, we compiled over 5.3 million “breast cancer” related tweets spanning September 2016 until mid December 2017. We combined supervised machine learning methods with natural language processing to sift tweets relevant to breast cancer patient experiences. We analyzed a sample of 845 breast cancer patient and survivor accounts, responsible for over 48,000 posts. We

investigated tweet content with a hedonometric sentiment analysis to quantitatively extract emotionally charged topics.

Results: We found that positive experiences were shared regarding patient treatment, raising support, and spreading awareness. Further discussions related to healthcare were prevalent and largely negative focusing on fear of political legislation that could result in loss of coverage.

Conclusions: Social media can provide a positive outlet for patients to discuss their needs and concerns regarding their healthcare coverage and treatment needs. Capturing iPROs from online communication can help inform healthcare professionals and lead to more connected and personalized treatment regimens.

4.1 Introduction

Twitter has shown potential for monitoring public health trends, (Alajajian et al., 2017; Paul and Dredze, 2011; Shive et al., 2013; Dredze, 2012; Reece et al., 2016), disease surveillance, (Lamb et al., 2013), and providing a rich online forum for cancer patients, (Sugawara et al., 2012). Social media has been validated as an effective educational and support tool for breast cancer patients, (Attai et al., 2015), as well as for generating awareness, (Bender et al., 2011). Successful supportive organizations use social media sites for patient interaction, public education, and donor outreach, (Fussell Sisco and McCorkindale, 2013). The advantages, limitations, and future potential of using social media in healthcare has been thoroughly reviewed, (Moorhead et al., 2013). Our study aims to investigate tweets mentioning “breast” and “cancer” to analyze patient populations and selectively obtain content relevant to patient treatment experiences.
Our previous study, (Crannell et al., 2016), collected tweets mentioning “cancer” over several months to investigate the potential for monitoring self-reported patient treatment experiences. Non-relevant tweets (e.g. astrological and horoscope references) were removed and the study identified a sample of 660 tweets from patients who were describing their condition. These self-reported diagnostic indicators allowed for a sentiment analysis of tweets authored by patients. However, this process was tedious, since the samples were hand verified and sifted through multiple keyword searches. Here, we aim to automate this process with machine learning context classifiers in order to build larger sets of patient self-reported outcomes in order to quantify the patent experience.

Patients with breast cancer represent a majority of people affected by and living with cancer. As such, it becomes increasingly important to learn from their experiences and understand their journey from their own perspective. The collection and analysis of Invisible Patient Reported Outcomes (iPROs) offers a unique opportunity to better understand the patient perspective of care and identify gaps meeting particular patient care needs.

4.2 Materials and methods

4.2.1 Data Description

Twitter provides a free streaming Application Programming Interface (API), (Twitter, 2017), for researchers and developers to mine samples of public tweets. Language processing and data mining, (Roesslein, 2009), was conducted using the Python programming language. The free public API allows targeted keyword mining of up to 1% of Twitter's full volume at any given time, referred to as the 'Spritzer Feed'.

We collected tweets from two distinct Spritzer endpoints from September 15th, 2016 through December 9th, 2017. The primary feed for the analysis collected 5.3 million tweets containing the keywords 'breast' AND 'cancer'. See Fig 4.1 for detailed Twitter frequency statistics along with the user activity distribution. Our secondary feed searched just for the keyword 'cancer' which served as a comparison (76.4 million tweets, see SI 1), and helped us collect additional tweets relevant to cancer from patients. The numeric account ID provided in tweets helps to distinguish high frequency tweeting entities.

Sentence classification combines natural language processing (NLP) with machine learning to identify trends in sentence structure, (Zhang and Wallace, 2015; Blunsom et al., 2014). Each tweet is converted to a numeric word vector in order to identify distinguishing features by training an NLP classifier on a validated set of relevant tweets. The classifier acts as a tool to sift through ads, news, and comments not related to patients. Our scheme combines a logistic regression classifier, (Genkin et al., 2007), with a Convolutional Neural Network (CNN), (Kim, 2014; Britz, 2015), to identify self-reported diagnostic tweets.

It is important to be wary of automated accounts (e.g., bots, spam) whose large output of tweets pollute relevant organic (i.e., human) content, (Clark et al., 2016), and can distort sentiment analyses, (Clark et al.,

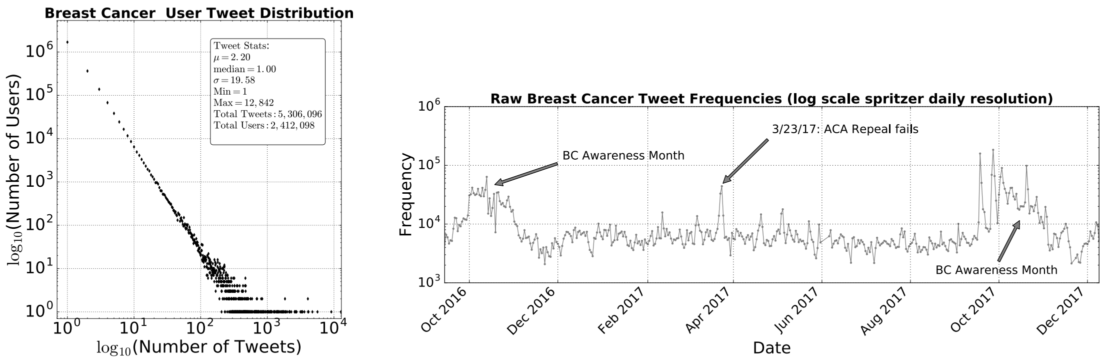


Figure 4.1: (left) The distribution of tweets per given user is plotted on a log axis. The tail tends to be high frequency automated accounts, some of which provide daily updates or news related to cancer. (right) A frequency time-series of the tweets collected, binned by day.

2016). Prior to applying sentence classification, we removed tweets containing hyperlinks to remove automated content (some organic content is necessarily lost with this strict constraint).

The user tweet distribution in Fig 4.1, shows the number of users as a function of the number of their tweets we collected. With an average frequency of 2.2 tweets per user, this is a relatively healthy activity distribution. High frequency tweeting accounts are present in the tail, with a single account producing over 12,000 tweets —an automated account served as a support tool called ‘ClearScan’ for patients in recovery. Approximately 98% of the 2.4 million users shared less than 10 posts, which accounted for 70% of all sampled tweets.

The Twitter API also provided the number of tweets withheld from our sample, due to rate limiting. Using overflow statistics provided by Twitter, we estimated the sampled proportion of collected tweets mentioning these keywords. These targeted feeds were able to collect a large sample of all tweets mentioning these terms; approximately 96% of tweets mentioning “breast,cancer” and 65.2% of all tweets mentioning ‘cancer’ while active. More information regarding the types of Twitter endpoints and calculating the sampling proportion of collected tweets is described in the Appendix.

Our goal was to analyze content authored only by patients. To help ensure this outcome we removed posts containing a URL for classification, (Clark et al., 2016). Twitter allows users to spread content from other users via ‘retweets’. We also removed these posts prior to classification to isolate tweets authored by patients. We also accounted for non-relevant astrological content by removing all tweets containing any of

the following horoscope indicators:
'astrology','zodiac','astronomy','horoscope','aquarius','pisces','aries','taurus','leo','virgo','libra', and 'scorpio'. We preprocessed tweets by lowercasing and removing punctuation. We also only analyzed tweets for which Twitter had identified 'en' for the language English.

4.2.2 Sentiment Analysis and Hedonometrics

We evaluated tweet sentiments with hedonometrics, (Dodds et al., 2011, 2015), using LabMT, a labeled set of 10,000 frequently occurring words rated on a 'happiness' scale by individuals contracted through Amazon Mechanical Turk, a crowd-sourced survey tool. These happiness scores helped quantify the average emotional rating of text by totaling the scores from applicable words and normalizing by their total frequency. Hence, the average happiness score, h_{avg}, of a corpus with N words in common with LabMT was computed with the weighted arithmetic mean of each word's frequency, f_w, and associated happiness score, h_w:

$$h_{\text{avg}} = \frac{\sum\limits_{w=1}^{N} f_w \cdot h_w}{\sum\limits_{w=1}^{N} f_w} \tag{4.1}$$

The average happiness of each word was rated on a 9 point scale ranging from extremely negative (e.g., 'emergency' 3.06, 'hate' 2.34, 'die' 1.74) to positive (e.g., 'laughter' 8.50, 'love' 8.42, 'healthy' 8.02). Neutral 'stop words' ($4 \leq h_{\text{avg}} \leq 6$, e.g., 'of','the', etc.) were removed to enhance the emotional signal of each set of tweets. These high frequency, low sentiment words can dampen a signal, so their removal can help identify hidden trends. One application is to plot h_{avg} as a function of time. The happiness time-series can provide insight driving emotional content in text. In particular, peak and dips (i.e., large deviations from the average) can help identify interesting themes that may be overlooked in the frequency distribution. Calculated scores can give us comparative insight into the context between sets of tweets.
The hedonometer (see `www.hedonometer.org`) is a well-tested relative metric for comparing emotional context between word frequency distributions. This tool can help visualize the average happiness scores from subsets of text as a function of time or topic. The daily average happiness score of a random 10% sample of Twitter generally sits at 6.0 and usually fluctuates by 0.05 per day. Larger daily shifts (≥0.15)

tend to correspond to worldly events. For reference, Christmas day (in 2018 $h_{\text{avg}} = 6.28$) has among the highest happiness scores, while states of emergency tend to have the most negative. For example, the 2016 terrorist attack in Orlando, Florida scored a 5.84 on the hedonometer, dropping by 0.18 from the previous day. These shifts are then quantified by identifying the most influential emotionally charged terms causing the computed shift in average word happiness.

"Word shift graphs" compare the terms contributing to shifts in a computed word happiness from two term frequency distributions. This tool is useful in isolating emotional themes from large sets of text and has been previously validated in monitoring public opinion, (Cody et al., 2015) as well as for geographical sentiment comparative analyses, (Mitchell et al., 2013). See SI 3 for a general description of word shift graphs and how to interpret them.

4.2.3 Relevance Classification: Logistic Model and CNN Architecture

We began by building a validated training set of tweets for our sentence classifier. We compiled the patient tweets verified by, (Crannell et al., 2016), to train a logistic regression content relevance classifier using a similar framework as, (Genkin et al., 2007). To test the classifier, we compiled over 5 million tweets mentioning the word cancer from a 10% 'Gardenhose' random sample of Twitter spanning January through December 2015. See SI 1 for a statistical overview of this corpus.

We tested a maximum entropy logistic regression classifier using a similar scheme as, (Genkin et al., 2007). NLP classifiers operate by converting sentences to word vectors for identifying key characteristics — the vocabulary of the classifier. Within the vocabulary, weights were assigned to each word based upon a frequency statistic. We used the term frequency crossed with the inverse document frequency (tf-idf), as described in , (Genkin et al., 2007). The tf-idf weights helped distinguish each term's relative weight across the entire corpus, instead of relying on raw frequency. This statistic dampens highly frequent non-relevant words (e.g. 'of', 'the', etc.) and enhances relatively rare yet informative terms (e.g. survivor, diagnosed, fighting). This method is commonly implemented in information retrieval for text mining, (Salton et al., 1975). The logistic regression context classifier then performs a binary classification of the tweets we collected from 2015. See SI 4 for an expanded description of the sentence classification methodology.

We validated the logistic model's performance by manually verifying 1,000 tweets that were classified as 'relevant'. We uncovered three categories of immediate interest including: tweets authored by patients

No.	Tweet Key Identifying Phrases
1	Breast cancer fear gone! Tumor removed ...
2	... my tremendously difficult journey through Stage IV Breast Cancer ...
3	... life after breast cancer. I am 11 years Cancer Free ...
4	... IM FIGHTING BREAST CANCER STAGE 3 ...
5	@USER just got diagnosed with breast cancer ...

Table 4.1: **Diagnostic Training Sample Tweet Phrases:** A sample of self-reported diagnostic phrases from tweets used to train the logistic regression content classifier (modified to preserve anonymity).

regarding their condition (21.6%), tweets from friends/family with a direct connection to a patient (21.9%), and survivors in remission (8.8%). We also found users posting diagnostic related inquiries (7.6%) about possible symptoms that could be linked to breast cancer, or were interested in receiving preventative check-ups. The rest (40.2%) were related to 'cancer', but not to patients and include public service updates as well as non-patient authored content (e.g., support groups). We note that the classifier was trained on very limited validated data (N=660), which certainly impacted the results. We used this validated annotated set of tweets to train a more sophisticated classifier to uncover self-diagnostic tweets from users describing their personal breast cancer experiences as current patients or survivors.

We implemented the Convolutional Neural Network (CNN) with Google's Tensorflow interface, (Abadi et al., 2016). We adapted our framework from, (Britz, 2015), but instead trained the CNN on these 1000 labeled cancer related tweets. The trained CNN was applied to predict patient self-diagnostic tweets from our breast cancer dataset. The CNN outputs a binary value: positive for a predicted tweet relevant to patients or survivors and negative for these other described categories (patient connected, unrelated, diagnostic inquiry). The Tensorflow CNN interface reported a 97.6% accuracy when evaluating this set of labels with our trained model. These labels were used to predict self-reported diagnostic tweets relevant to breast cancer patients.

4.3 Results

A set of 845 breast cancer patient self-diagnostic Twitter profiles was compiled by implementing our logistic model followed by prediction with the trained CNN on 9 months of 'breast cancer' tweets. The logistic model sifted 4,836 relevant tweets of which 1,331 were predicted to be self-diagnostic by the CNN. Two independent groups annotated the 1,331 tweets to identify patients and evaluate the classifier's results.

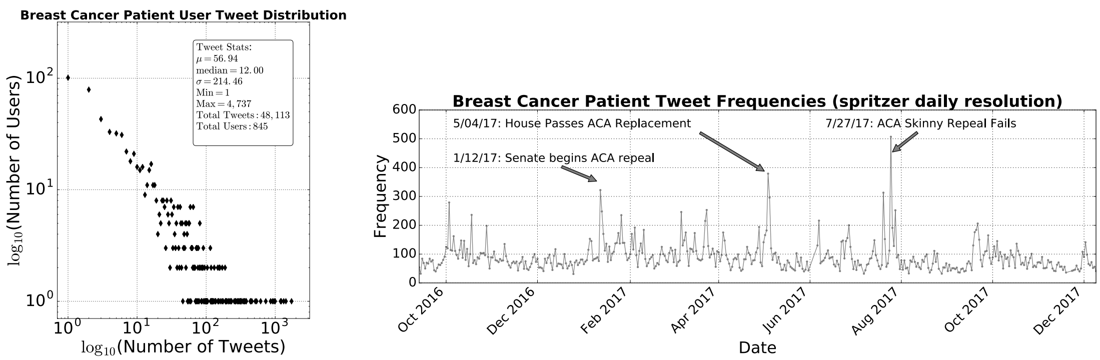


Figure 4.2: (left) The distribution of tweets per given patient/survivor is plotted on a log axis along with a statistical summary of patient tweeting behavior. (right) A frequency time-series of patient tweets collected, binned by day.

The raters, showing high inter-rater reliability, individually evaluated each tweet as self-diagnostic of a breast cancer patient or survivor. The rater's independent annotations had a 96% agreement.

The classifier correctly identified 1,140 tweets (85.6%) from 845 profiles. A total of 48,113 tweets from these accounts were compiled from both the 'cancer' (69%) and 'breast' 'cancer' (31%) feeds. We provided tweet frequency statistics in Fig 4.2. This is an indicator that this population of breast cancer patients and survivors are actively tweeting about topics related to 'cancer' including their experiences and complications.

Next, we applied hedonometrics to compare the patient posts with all collected breast cancer tweets. We found that the surveyed patient tweets were less positive than breast cancer reference tweets. In Fig 4.3, the time series plots computed average word happiness at monthly and daily resolutions. The daily happiness scores (small markers) have a high fluctuation, especially within the smaller patient sample (average 100 tweets/day) compared to the reference distribution (average 10,000 tweets/day). The monthly calculations (larger markers) highlight the negative shift in average word happiness between the patients and reference tweets. Large fluctuations in computed word happiness correspond to noteworthy events, including breast cancer awareness month in October, cancer awareness month in February, as well as political debate regarding healthcare beginning in March May and July 2017.

In Fig 4.4 word shift graphs display the top 50 words responsible for the shift in computed word happiness between distributions. On the left, tweets from patients were compared to all collected breast cancer tweets.

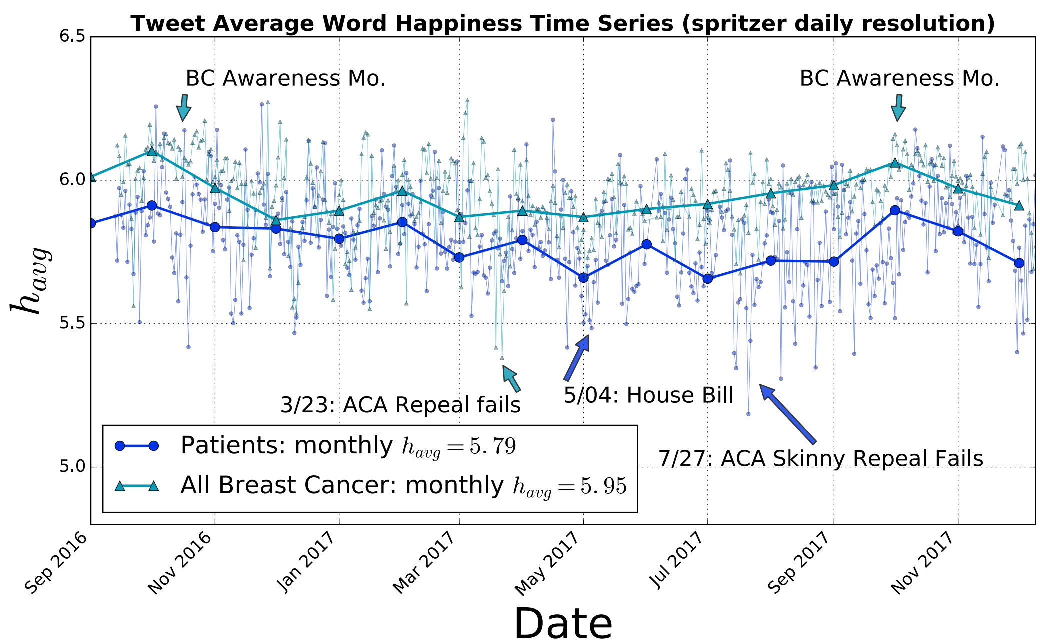


Figure 4.3: Computed average word happiness as a function of day (small markers) and month (large markers) for both the 'breast','cancer' and patient distributions. The patient monthly average was less positive than the reference distribution ($h_{\text{avg}} = 5.78$ v. 5.93).

Patient tweets, T_{comp}, were less positive ($h_{\text{avg}} = 5.78$ v. 5.97) than the reference distribution,T_{ref}. There were relatively less positive words 'mom', 'raise', 'awareness', 'women', 'daughter', 'pink', and 'life' as well as an increase in the negative words 'no(t)', 'patients, 'dying', 'killing', 'surgery' 'sick', 'sucks', and 'bill'. Breast cancer awareness month, occurring in October, tends to be a high frequency period with generally more positive and supportive tweets from the general public which may account for some of the negative shift in the patient sample. Notably, there was a relative increase of the positive words 'me', 'thank', 'you' ,'love', and 'like' which may indicate that many tweet contexts were from the patient's perspective regarding positive experiences. Many tweets regarding treatment were enthusiastic, supportive, and proactive. Other posts were descriptive: over 165 sampled patient tweets mentioned personal chemo therapy experiences and details regarding their treatment schedule, and side effects.

Numerous patients and survivors in our sample had identified their condition in reference to the American healthcare regulation debate. Many sampled views of the proposed legislation were very negative, since repealing the Affordable Care Act without replacement could leave many uninsured. Other tweets mentioned worries regarding insurance premiums and costs for patients and survivors' continued screening. In particular the pre-existing condition mandate was a chief concern of patients/survivors future coverage. This was echoed by 55 of the sampled patients with the hashtag #iamapreexistingcondition (See Table C.1). Hashtags (#) are terms that categorize topics within posts. In Table C.1, the most frequently occurring hashtags from both the sampled patients (right) and full breast cancer corpus (left). Each entry contains the tweet frequency, number of distinct profiles, and the relative happiness score (h_{avg}) for comparisons. Political terms were prevalent in both distributions describing the Affordable Care Act (#aca, #obamacare, #saveaca, #pretectourcare) and the newly introduced American Healthcare Act (#ahca, #trumpcare). A visual representation of these hashtags are displayed using a word-cloud in Figure C.5.

Tweets referencing the AHCA were markedly more negative than those referencing the ACA. This shift was investigated in Fig 4.4 with a word shift graph. We compared American Healthcare Act Tweets, T_{comp}, to posts mentioning the Affordable Care Act, T_{ref}. AHCA were relatively more negative ($h_{\text{avg}} = 5.48$ v. 6.05) due to an increase of negatively charged words 'scared', 'lose', 'tax', 'zombie', 'defects', 'cut', 'depression', 'killing', and 'worse' . These were references to the bill leaving many patients/survivors without insurance and jeopardizing future treatment options. 'Zombie' referenced the bill's potential return for subsequent votes.

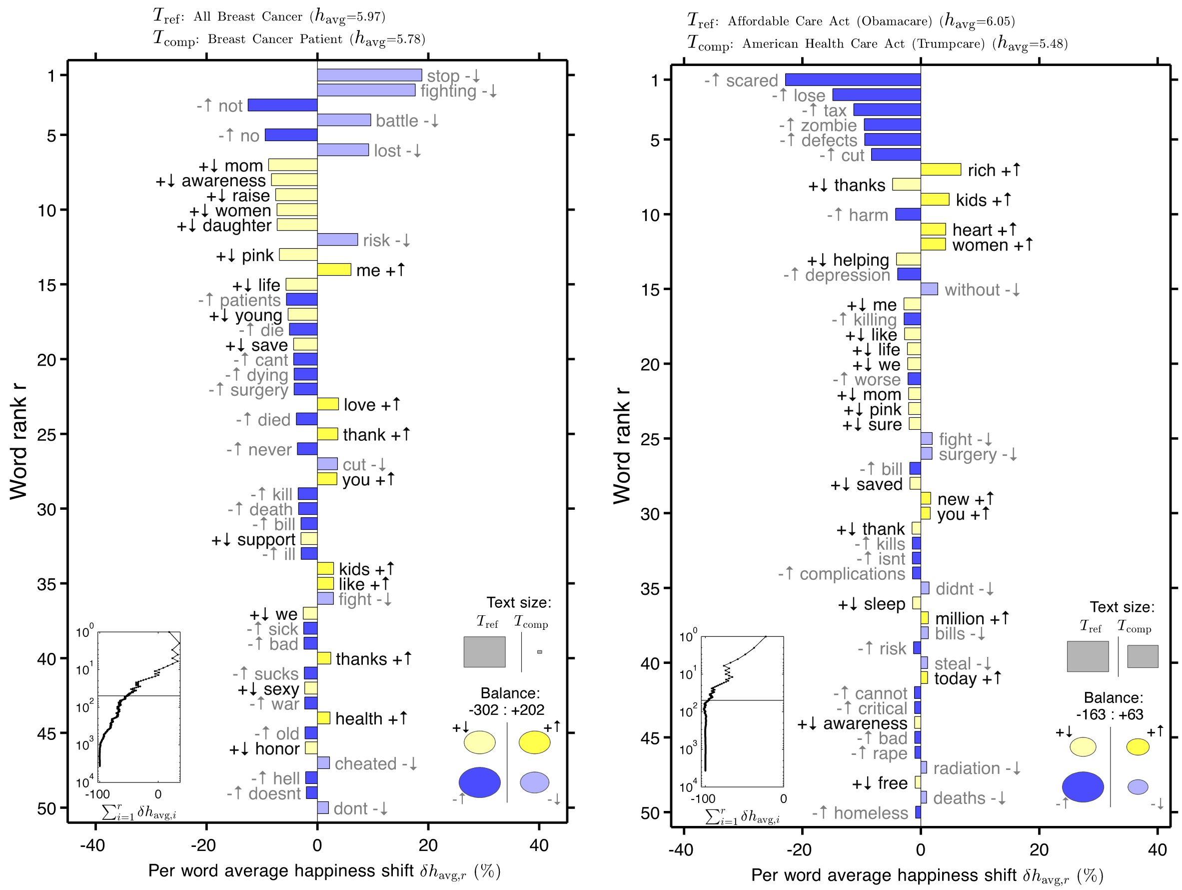


Figure 4.4: (Left) Word shift graph comparing collected Breast Cancer Patient Tweets, T_{comp}, to all Breast Cancer Tweets, T_{ref}. Patient Tweets were less positive ($h_{\text{avg}} = 5.78$ v. 5.97), due to a decrease in positive words ‘mom’, ‘raise’, ‘awareness’, ‘women’, ‘daughter’, ‘pink’, and ‘life’ as well as an increase in the negative words ‘no(t)’, ‘patients, ‘dying’, ‘killing’, ‘surgery’ ‘sick’, ‘sucks’, and ‘bill’. (Right) Word shift graph comparing tweets mentioning the American Healthcare Act (AHCA, 10.5k tweets) to the Affordable Care Act (ACA, 16.9k tweets). AHCA tweets were more negative ($h_{\text{avg}} = 5.48$ v. 6.05) due to a relative increase in the negative words ‘scared’, ‘lose’, ‘zombie’, ‘defects’, ‘depression’, ‘harm’, ‘killing’, and ‘worse’.

Top Hashtags(#): All Breast Cancer				
Rank	Term	Tweets	Users	h_{avg}
1	#cancer	67,111	23,171	5.92
2	#breastcancer	66,400	22,247	5.97
3	#breast	35,544	11,115	6.0
4	#nobraday	23,406	16,785	5.76
5	#breastcancerawarenessmonth	20,961	13,491	6.06
6	#health	17,484	5,696	5.82
7	#twibbon	16,809	14,332	6.18
8	#bcsm	14,955	4,644	5.95
9	#survivor	14,500	1,107	5.98
10	#idrivefor	13,562	8,331	6.06
11	#breastcancerawareness	13,429	8,820	6.13
12	#lymphedema	13,263	2,274	5.88
13	#walk	9,344	246	6.0
14	#aca	8,903	8,105	6.05
15	#ga06	8,266	5,821	5.15
16	#iamapreexistingcondition	7,604	6,215	5.41
17	#himinitiative	7,294	572	6.04
18	#news	6,435	1,680	5.79
19	#malebreastcancer	5,821	1,469	6.0
20	#savethetatas	5,551	5,390	6.11
21	#giveaway	4,861	1,284	6.31
22	#trumpcare	4,778	4,331	5.53
23	#keepkadcyla	3,822	3,064	5.68
24	#awareness	3,697	1,369	6.19
25	#brca	3,652	1,284	5.89
26	#avonrep	3,517	1,620	5.49
27	#pink	3,480	2,763	6.34
28	#ad	3,458	1,383	6.08
29	#nbcf	3,445	1,965	6.49
30	#1savetatas	3,051	1,040	6.51
31	#worldcancerday	2,936	2,430	6.05
32	#exercise	2,740	1,492	5.71
33	#thinkpink	2,707	2,209	6.1
34	#ahca	2,607	2,403	5.67
35	#spas4acause	2,585	1,615	6.48
36	#bcam	2,555	1,961	6.14
37	#thegoodlie	2,314	474	5.63
38	#healthcare	2,261	1,396	5.85
39	#obamacare	2,240	2,059	6.2
40	#pinkribbon	2,201	1,104	5.9
41	#nfl	2,188	647	6.13
42	#oncology	2,188	762	5.85
43	#unitedbyher	2,117	602	6.1
44	#sabcs16	2,104	828	5.67
45	#cnndebatenight	2,097	2,029	5.9
46	#women	2,078	1,231	5.85
47	#nyfw	2,060	1,886	6.11
48	#donate	2,016	1,279	5.76
49	#pinkout	1,946	1,778	6.12
50	#ai	1,937	1,241	6.03
*	**Total**	462,192	155,218	5.96

Top Hashtags(#): Breast Cancer Patient Sample				
Rank	Term	Tweets	Users	h_{avg}
1	#cancer	2,063	239	5.76
2	#bcsm	1,220	61	5.92
3	#lymphedema	680	12	5.93
4	#breastcancer	568	112	5.84
5	#aca	469	88	5.69
6	#trumpcare	168	70	5.39
7	#ahca	165	45	5.4
8	#amsm	165	25	5.61
9	#metastatic	161	17	5.92
10	#malebreastcancer	155	21	5.94
11	#worldcancerday	134	54	5.94
12	#obamacare	132	42	5.77
13	#saveaca	115	47	5.85
14	#bccww	112	24	5.77
15	#lcsm	108	14	5.89
16	#survivor	92	33	5.83
17	#protectourcare	91	37	5.75
18	#iamapreexistingcondition	82	55	5.63
19	#breast	79	24	6.2
20	#breastcancerrealitycheck	64	17	5.66
21	#breastcancerawarenessmonth	62	41	6.04
22	#healthcare	62	32	5.44
23	#kissthis4mbc	61	14	6.13
24	#mbc	59	21	5.69
25	#cancersucks	57	34	5.75
26	#oncology	54	11	5.84
27	#maga	53	34	5.38
28	#trump	53	33	5.09
29	#immunotherapy	52	18	5.81
30	#clinicaltrials	51	12	6.13
31	#acaworks	47	11	5.72
32	#research	46	17	6.02
33	#breastcancerawareness	45	36	5.89
34	#f***cancer	44	21	5.88
35	#nhs	42	16	5.62
36	#brca	42	17	5.81
37	#gop	41	18	5.28
38	#metastaticbc	41	19	5.9
39	#idrivefor	40	19	6.27
40	#grahamcassidy	40	23	5.3
41	#mbcproject	39	11	6.28
42	#health	38	24	6.0
43	#gyncsm	37	10	6.12
44	#sabcs16	36	12	5.75
45	#endcancer	35	13	6.08
46	#wecanican	34	10	5.83
47	#savebeth	34	10	5.88
48	#cancermoonshot	31	18	6.03
49	#moreformbc	30	16	6.17
50	#resist	30	21	5.52
*	**Total**	8,159	398	5.81

Table 4.2: A table of the 50 most frequently tweeted hashtags (#) from all collected breast cancer tweets (left) and from sampled breast cancer patients (right). The relative computed ambient happiness h_{avg} for each hashtag is colored relative to the group average (blue- negative, orange - positive).

4.4 Discussion

We have demonstrated the potential of using sentence classification to isolate content authored by breast cancer patients and survivors. Our novel, multi-step sifting algorithm helped us differentiate topics relevant to patients and compare their sentiments to the global online discussion. The hedonometric comparison of frequent hashtags helped identify prominent topics how their sentiments differed. This shows the ambient happiness scores of terms and topics can provide useful information regarding comparative emotionally charged content. This process can be applied to disciplines across health care and beyond.

Throughout 2017, Healthcare was identified as a pressing issue causing anguish and fear among the breast cancer community; especially among patients and survivors. During this time frame, US legislation was proposed by Congress that could roll back regulations ensuring coverage for individuals with pre-existing conditions. Many individuals identifying as current breast cancer patients/survivors expressed concerns over future treatment and potential loss of their healthcare coverage. Twitter could provide a useful political outlet for patient populations to connect with legislators and sway political decisions.

March 2017 was a relatively negative month due to discussions over American healthcare reform. The American Congress held a vote to repeal the Affordable Care Act (ACA, also referred to as ‘Obamacare’), which could potentially leave many Americans without healthcare insurance, (Kaplan et al., 2017). There was an overwhelming sense of apprehension within the ‘breast cancer’ tweet sample. Many patients/survivors in our diagnostic tweet sample identified their condition and how the ACA ensured coverage throughout their treatment.

This period featured a notable tweet frequency spike, comparable to the peak during breast cancer awareness month. The burst event peaked on March 23rd and 24th (65k, 57k tweets respectively, see Fig 4.1). During the peak, 41,983 (34%) posts contained ‘care’ in reference to healthcare, with a viral retweeted meme accounting for 39,183 of these posts. The tweet read: ”The group proposing to cut breast cancer screening, maternity care, and contraceptive coverage.” with an embedded photo of a group of predominately male legislators, (ME.ME, 2017). The criticism referenced the absence of female representation in a decision that could deprive many of coverage for breast cancer screenings. The online community condemned the decision to repeal and replace the ACA with the proposed legislation with references to people in treatment

No.	Tweet
1	i was diagnosed ... with stage 2 breast cancer ... after 4 years in remission ...
2	obamacare saved me! i have had breast cancer twice ...
3	yesterday i was diagnosed with breast cancer ...
4	... i have breast cancer but i will get through this ...
5	... i've had .. breast cancer and ... i can't get insurance because i can't afford it

Table 4.3: **Sampled Predicted Diagnostic Tweets:** A sample of key phrases from self-reported diagnostic tweets predicted from the CNN classifier with the patient relevant proportional ratio, $\alpha = 1 : 10$.

who could 'die' (n=7,923) without appropriate healthcare insurance coverage. The vote was later postponed and eventually failed, (MJ Lee and Killough, 2017).

Public outcry likely influenced this legal outcome, demonstrating Twitter's innovative potential as a support tool for public lobbying of health benefits. Twitter can further be used to remind, motivate and change individual and population health behavior using messages of encouragement (translated to happiness) or dissatisfaction (translated to diminished happiness), for example, with memes that can have knock-on social consequences when they are re-tweeted. Furthermore, Twitter may someday be used to benchmark treatment decisions to align with expressed patient sentiments, and to make or change clinical recommendations based upon the trend histories that evolve with identifiable sources but are entirely in the public domain.

Analyzing the fluctuation in average word happiness as well as bursts in the frequency distributions can help identify relevant events for further investigation. These tools helped us extract themes relevant to breast cancer patients in comparison to the global conversation.

One area in which Twitter has traditionally fallen short for a communication medium is that of the aural dimension, such as nuances and inflections. However, Twitter now includes pictures, videos and emojis with people revealing or conveying their emotions by use of these communication methods. It is envisaged that the aural and visual dimensions will eventually grow to complement the published text component towards a more refined understanding of feelings, attitudes and health and clinical sentiments.

Lack of widespread patient adoption of social media could be a limiting factor to our analysis. A study of breast cancer patients during 2013–2014, (Wallner et al., 2016), found social media was a less prominent form of online communication (N = 2578, 12.3%), however with the advent of smartphones and the internet of things (iot) movement, social media may influence a larger proportion of future patients. Another finding noted that online posts were more likely to be positive about their healthcare decision experience or about survivorship. Therefore we cannot at this time concretely draw population-based assumptions from social media sampling. Nevertheless, understanding this online patient community could serve as a valuable tool for healthcare providers and future studies should investigate current social media usage statistics across patients.

Because we trained the content classifier with a relatively small corpus, the model likely over-fit on a few particular word embeddings. For example: 'i have stage iv', 'i am * survivor', 'i had * cancer'. However, this is similar to the process of recursive keyword searches to gather related content. Also, the power of the CNN allows for multiple relative lingual syntax as opposed to searching for static phrases ('i have breast cancer', 'i am a survivor'). The CNN shows great promise in sifting relevant context from large sets of data. Other social forums for patient self reporting and discussion should be incorporated into future studies. For example, as of 2017, `https://community.breastcancer.org` has built a population of over 199,000 members spanning 145,000 topics. These tools could help connect healthcare professionals with motivated patients. Labeled posts from patients could also help train future context models and help identify adverse symptoms shared among online social communities.

Our study focused primarily on English tweets, since this was the language of our diagnostic training sample. Future studies could incorporate other languages using our proposed framework. It would be important to also expand the API queries with translations of 'breast' and 'cancer'. This could allow for a cross cultural comparison of how social media influences patients and what patients express on social media.

4.5 Conclusion

We have demonstrated the potential of using context classifiers for identifying diagnostic tweets related to the experience of breast cancer patients. Our framework provides a proof of concept for integrating machine learning with natural language processing as a tool to help connect healthcare providers with patient experiences. These methods can inform the medical community to provide more personalized treatment regimens by evaluating patient satisfaction using social listening. Twitter has also been shown as a useful medium for political support of healthcare policies as well as spreading awareness. Applying these analyses across other social media platforms could provide comparably rich data-sets. For instance, Instagram has been found to contain indicative markers for depression, (Reece and Danforth, 2017).

Integrating these applications into our healthcare system could provide a better means of tracking iPROs across treatment regimens and over time.

One area in which Twitter has traditionally fallen short for a communication medium is that of the aural dimension, such as nuances and inflections. However, Twitter now includes pictures, videos, and emojis with people revealing or conveying their emotions by use of these communication methods. With Siri, augmented reality, virtual reality, and even chatbot interfaces such as trUStr (.us) someday connecting to text-based social media, it is envisaged that the aural and visual dimensions will eventually grow to complement the published text component towards a more refined understanding of feelings, attitudes and health and clinical sentiments.

Follow-on studies to our work could be intended to further develop these models and apply them to larger streams of data. Online crowd sourcing tools, like Amazon's Mechanical Turk, implemented in, (Dodds et al., 2015), can help compile larger sets of human validated labels to improve context classifiers. These methods can also be integrated into delivering online outreach surveys as another tool for validating healthcare providers. Future models, trained on several thousand labeled tweets for various real world applications should be explored. Invisible Patient Reported Outcomes should be further investigated via sentiment and context analyses for a better understanding of how to integrate the internet of things with healthcare.

Twitter has become a powerful platform for amplifying political voices of individuals. The response of the online breast cancer community to the American Healthcare Act as a replacement to the Affordable Care Act was largely negative due to concerns over loss of coverage. A widespread negative public reaction may have helped influence this political result. Social media opinion mining could present as a powerful tool for legislators to connect with and learn from their constituents. This can lead to positive impacts on population health and societal well-being.

4.6 References

Abadi, M., A. Agarwal, P. Barham, E. Brevdo, Z. Chen, C. Citro, G. S. Corrado, A. Davis, J. Dean, M. Devin, et al. (2016). Tensorflow: Large-scale machine learning on heterogeneous distributed systems. *arXiv preprint arXiv:1603.04467*.

Alajajian, S. E., J. R. Williams, A. J. Reagan, S. C. Alajajian, M. R. Frank, L. Mitchell, J. Lahne, C. M. Danforth, and P. S. Dodds (2017). The Lexicocalorimeter: Gauging public health through caloric input and output on social media. *PloS ONE 12*(2), e0168893.

Attai, D. J., M. S. Cowher, M. Al-Hamadani, J. M. Schoger, A. C. Staley, and J. Landercasper (2015). Twitter social media is an effective tool for breast cancer patient education and support: Patient-reported outcomes by survey. *Journal of Medical Internet Research 17*(7).

Bender, J. L., M.-C. Jimenez-Marroquin, and A. R. Jadad (2011). Seeking support on facebook: a content analysis of breast cancer groups. *Journal of Medical Internet Research 13*(1).

Blunsom, P., E. Grefenstette, and N. Kalchbrenner (2014). A convolutional neural network for modelling sentences. In *Proceedings of the 52nd Annual Meeting of the Association for Computational Linguistics*. Proceedings of the 52nd Annual Meeting of the Association for Computational Linguistics.

Britz, D. (2015). Implementing a CNN for text classification in tensorflow.

Clark, E. M., C. A. Jones, J. R. Williams, A. N. Kurti, M. C. Norotsky, C. M. Danforth, and P. S. Dodds (2016). Vaporous marketing: Uncovering pervasive electronic cigarette advertisements on twitter. *PloS ONE 11*(7), e0157304.

Clark, E. M., J. R. Williams, C. A. Jones, R. A. Galbraith, C. M. Danforth, and P. S. Dodds (2016). Sifting robotic from organic text: A natural language approach for detecting automation on Twitter. *Journal of Computational Science 16*, 1–7.

Cody, E. M., A. J. Reagan, L. Mitchell, P. S. Dodds, and C. M. Danforth (2015). Climate change sentiment on Twitter: An unsolicited public opinion poll. *PloS ONE 10*(8), e0136092.

Crannell, W. C., E. Clark, C. Jones, T. A. James, and J. Moore (2016). A pattern-matched Twitter analysis of U.S. cancer-patient sentiments. *Journal of Surgical Research 206*(2), 536–542.

Dodds, P. S., E. M. Clark, S. Desu, M. R. Frank, A. J. Reagan, J. R. Williams, L. Mitchell, K. D. Harris, I. M. Kloumann, J. P. Bagrow, K. Megerdoomian, M. T. McMahon, B. F. Tivnan, and C. M. Danforth (2015). Human language reveals a universal positivity bias. *Proceedings of the National Academy of Sciences 112*(8), 2389–2394.

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

Dredze, M. (2012). How social media will change public health. *IEEE Intelligent Systems 27*(4), 81–84.

Fussell Sisco, H. and T. McCorkindale (2013). Communicating “Pink”: An analysis of the communication strategies, transparency, and credibility of breast cancer social media sites. *International Journal of Nonprofit and Voluntary Sector Marketing 18*(4), 287–301.

Genkin, A., D. D. Lewis, and D. Madigan (2007). Large-scale Bayesian logistic regression for text categorization. *Technometrics 49*(3), 291–304.

Kaplan, T., R. Pear, and E. Huetteman (2017, March). Lacking the votes for passage, house calls off Obamacare repeal vote.

Kim, Y. (2014). Convolutional neural networks for sentence classification. *arXiv preprint arXiv:1408.5882*.

Lamb, A., M. J. Paul, and M. Dredze (2013). Separating fact from fear: Tracking flu infections on Twitter. In *HLT-NAACL*, pp. 789–795.

ME.ME (2017). The group proposing to cut breast cancer screening maternity care and contraceptive coverage. do you see the problem here? meme.

Mitchell, L., M. R. Frank, K. D. Harris, P. S. Dodds, and C. M. Danforth (2013). The geography of happiness: Connecting twitter sentiment and expression, demographics, and objective characteristics of place. *PloS one 8*(5), e64417.

MJ Lee, Lauren Fox, T. B. P. M. and A. Killough (2017, July). GOP Obamacare repeal bill fails in dramatic late-night vote.

Moorhead, S. A., D. E. Hazlett, L. Harrison, J. K. Carroll, A. Irwin, and C. Hoving (2013). A new dimension of health care: Systematic review of the uses, benefits, and limitations of social media for health communication. *Journal of Medical Internet Research 15*(4).

Paul, M. J. and M. Dredze (2011). You are what you tweet: Analyzing Twitter for public health. *ICWSM 20*, 265–272.

Reece, A. G. and C. M. Danforth (2017). Instagram photos reveal predictive markers of depression. *EPJ Data Science 6*(1), 15.

Reece, A. G., A. J. Reagan, K. L. Lix, P. S. Dodds, C. M. Danforth, and E. J. Langer (2016). Forecasting the onset and course of mental illness with Twitter data. *arXiv preprint arXiv:1608.07740*.

Roesslein, J. (2009). Tweepy documentation. *5*.

Salton, G., A. Wong, and C.-S. Yang (1975). A vector space model for automatic indexing. *Communications of the ACM 18*(11), 613–620.

Shive, M., M. Bhatt, A. Cantino, J. Kvedar, and K. Jethwani (2013). Perspectives on acne: what twitter can teach health care providers. *JAMA Dermatology 149*(5), 621–622.

Sugawara, Y., H. Narimatsu, A. Hozawa, L. Shao, K. Otani, and A. Fukao (2012). Cancer patients on twitter: a novel patient community on social media. *BMC Research Notes 5*(1), 699.

Twitter, I. (2017). Twitter streaming API.

Wallner, L. P., K. A. Martinez, Y. Li, R. Jagsi, N. K. Janz, S. J. Katz, and S. T. Hawley (2016). Use of online communication by patients with newly diagnosed breast cancer during the treatment decision process. *JAMA Oncology 2*(12), 1654–1656.

Zhang, Y. and B. Wallace (2015). A sensitivity analysis of (and practitioners' guide to) convolutional neural networks for sentence classification. *arXiv preprint arXiv:1510.03820*.

Chapter 5

Measuring Sentiments and Public Perceptions of Surgery Across Twitter

Background: Twitter, a popular social media platform, can provide valuable public insights for the healthcare industry. Surgeons have established a presence on various social media platforms for interacting with patients and showcasing results. Sentiments surrounding public perceptions of surgical types and methods mined from social media data have yet to be quantified.

Methods: We compiled over 5 million tweets containing the keyword 'surgery' spanning 2012-2016, authored by over 2 million individuals. We applied hedonometrics to identify emotionally charged topics and extract relevant content from the text. Key-word lists of surgical methods, types, and symptoms, compiled by a group of surgeons, were analyzed and compared.

Results: Tweets mentioning surgery were overall less positive and more variable in comparison to a random sample of Twitter. In both the sentiment and frequency distributions, relative extrema corresponded to surgical news related to celebrities and well as world events. A sentiment comparison of broad surgical types revealed plastic surgery tweets were considerably more negative than those mentioning cosmetic surgery. Words associated with plastic relative to cosmetic surgery appear to be oriented towards vanity as opposed to reconstructive which is in stark contrast to the definition of plastic surgery.

Conclusions: Our framework can help provide medical professionals with actionable insights into public perceptions regarding surgical disciplines and procedures. Sentiment analyses and opinion mining using social media can help identify public fallacies and provide a medium for the health care industry to connect with and constructively influence society.

5.1 Introductory Remarks

In this chapter, we present a preliminary analysis of mentions of "Surgery" across Twitter to serve as a guideline for performing hedonometric analyses in measuring public health perceptions with social media surveillance. Further remarks after the conclusion outline this process for the next generation of data scientists and enthusiasts. Results from this analysis directly contributed to (Chopan et al., 2018).

5.2 Introduction

Social media networking sites along with big data analytics are being integrated into health care research and communication, (Alshaikh et al., 2014; Chou et al., 2009; Grajales III et al., 2014; Keller et al., 2014). Twitter can provide health care professionals with relevant public insights into mental health indicators, (Reece et al., 2016), weight and caloric monitoring, (Alajajian et al., 2017), as well discussions about treatment and care for cancer patients (Crannell et al., 2016). Social media is an important information outlet for individuals seeking health related information, especially among young adults, and individuals inflicted with chronic disease, (Thackeray et al., 2013).

Social networking sites have become a mainstream communication outlet for surgeons, (Vardanian et al., 2013; Lui et al., 2017; Eberlin et al., 2018; Vardanian et al., 2013). Facebook profiles, operated by plastic surgeons, are perceived to positively impact their clinical practices and patient outreach, (Chang et al., 2015). An analysis of tweets mentioning #plasticsurgery called for surgeons to participate in social media as a means to connect with patients and the scientific community, rather than for marketing, (Branford et al., 2016). They showed the discussion was dominated by the general public (70% public v. 6% plastic surgeons). We aim to apply Natural Language Processing (NLP) to extract and compare public sentiments regarding surgery posts spanning several years.

We conducted a sentiment analysis using hedonometrics, (Dodds et al., 2011, 2015), to identify charged topics driving the global discussion among tweets mentioning the keyword ‘surgery’. These tools have shown potential for soliciting public opinion polls, (Cody et al., 2015), as well as comparing geographical influences on computed sentiments, (Mitchell et al., 2013). We also compared sentiment scores from several topics relevant to surgical types, symptoms, and body regions. We demonstrate the usefulness of hedonometrics to evaluate and extract emotional context from large scale textual corpora. These methods can be readily applied across health care disciplines and help inform medical professionals.

5.3 Methods

We collected over 5 million publicly available tweets containing the keyword ‘surgery’ spanning January 2012 until December 2016. Data was compiled using Twitter’s Streaming Application Programming Interface (API) from a random 10% sample of public blogs, the ‘Gardenhose Feed’. We only processed tweets for which Twitter had identified ‘en’ for the language English. A statistical summary of user tweets and a frequency time series are given in Figure 5.1. The user tweet distribution (left) plots the number of tweets per user on a log scale. The tail of the distribution is usually composed of automated entities (i.e., bots, cyborgs, or spam), (Clark et al., 2016). Identifying these accounts can be important since their high frequency tweeting behavior can have a profound influence a sentiment analysis, (Clark et al., 2016). In order to account for potential content polluters, we excluded posts from individuals who posted at least 250 tweets mentioning ‘surgery’. This excluded 224 users whom contributed 182,227 posts; the most active of which authored 10,380 tweets mentioning ‘surgery’.

Hedonometrics, a quantitative sentiment analysis procedure, uses LabMT, a word-happiness distribution with over 10,000 words, to calculate the average happiness score among different subsets of written text (i.e., tweets). The metric is relative, to compare various distributions and help identify interesting events with measurable divergence from the average. The average happiness score, h_{avg}, of a set of N words shared with LabMT is computed from the weighted arithmetic mean of each word’s frequency, f_w, and given happiness score, h_w:

$$h_{\text{avg}} = \frac{\sum_{w=1}^{N} f_w \cdot h_w}{\sum_{w=1}^{N} f_w} \tag{5.1}$$

"Word shift graphs" introduced in, (Dodds et al., 2011), are a useful tool for comparing the words that are causing a measured shift in average word happiness relative to a reference distribution. The 50 most influential words are plotted along with their relative weights (up/down arrows) and contribution (+ positive, - negative). These plots help identify emotionally charged topics within the text, (Reagan et al., 2015).

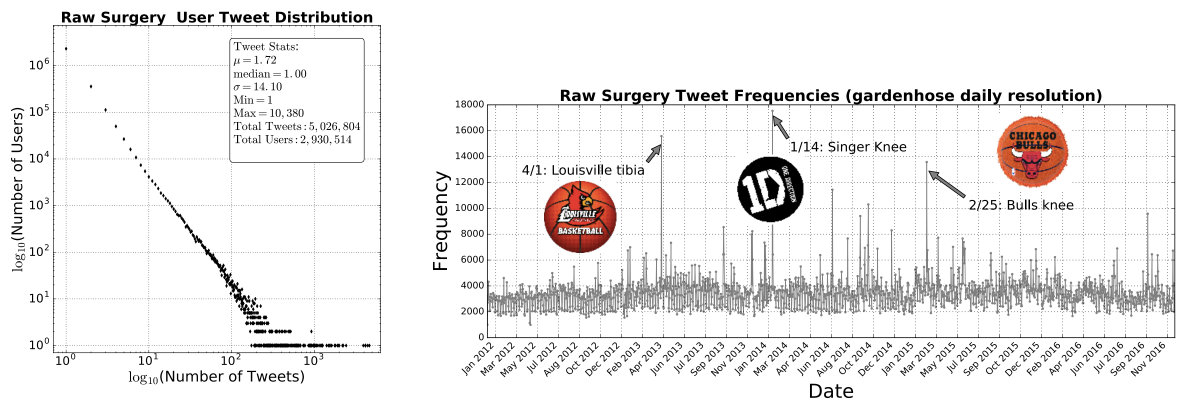


Figure 5.1: (Left) The distribution of tweets per account is plotted on a log axis along with a statistical summary of user tweeting behavior. (Right) A frequency time-series of tweets collected, binned by day.

We also applied the sentiment calculation to several sets of terms related to surgery. This helped compare and contrast their prevalence and relative perception. A broad list of surgical types, relevant terms, symptoms, and body regions was compiled by a group of surgeons. We measured the sentiments of words tweeted with each term to calculate its respective 'ambient happiness.' Hashtags are terms preceded with a # that allow users to self-categorize relevant topics within their tweets. The ambient happiness of the most frequently tweeted hashtags(#) among surgery tweets are also presented and discussed.

5.4 Results

Many frequency spikes in Figure 5.1 correspond to relevant celebrity surgeries of the time. Surgeries involving famous athletes are shared widely among fans and across the news. The peak on April 1st 2013 (15.5k tweets) was related to the successful surgery of a Louisville college basketball player whom fractured his tibia, (Stephen Jones and Himmelsbach, 2013) . The spike on February 25, 2015 (13.5k) referenced professional basketball player, Derrick Rose from the Bulls, needed to undergo knee surgery for a torn meniscus, (Golliver, 2015). On January 17 2014 (17.5k tweets), Niall Horan, a famous singer from the UK successfully underwent knee surgery. He commented that the surgery was a success however he mentioned he had hoped for more privacy, when photos of his recovery had leaked across social media, (Reporter, 2014). Another high frequency event occurred June 23rd, 2014, when potentially false information regarding a World Cup player went viral. Over 6500 tweets mentioned Cristiano Renaldo, many of which referenced him shaving a line in his scalp to stand in solidarity with a young fan suffering from cortical dysplasia. His hair style is believed to be unrelated, however Renaldo did personally finance the surgery of the child, (Sheets, 2014). This stands as a warning for people to be conscious of false information that could quickly spread across social media. This was also an apparent dip in the happiness time series , ($h_{\text{avg}} = 5.09$) as many tweets described the haircut's surgical scar resulting from a brain tumor.

There was a pronounced negative dip in the average happiness time series throughout September 2014. On September 4th 2014 referenced the death of Joan Rivers due to failed elective surgery, (2400 tweets, $h_{\text{avg}} = 5.46$) (Christensen, 2014). Another negative dip during this period described a viral story about a goldfish who underwent surgery for a tumor (September 15-16 2014, 6800 tweets), (Linshi, 2014). Also, pop singer Justin Bieber injured his ear drum while cliff diving on September (September 24-25th 7376 tweets, ($h_{\text{avg}} = 5.19$) (Zurilla, 2014). A severe negative event (2100 tweets, $h_{\text{avg}} = 4.56$) occurred on July 8th, 2016 referencing several Dallas police officers requiring surgery after a deadly shooting, (MANNY FERNANDEZ and BROMWICH, 2016). On January 27th 2013, Celtics basketball player, Rajon Rondo, required surgery due to an ACL tear (Forsberg, 2014), which corresponded to a relatively negative sentiment score ($h_{\text{avg}} = 5.39$). Another basketball player, Russel Westbrook representing the Thunders, required surgery for a serious meniscus tear (April 26, 2013) (von Horn, 2013).

Many positive events were also related to celebrity surgeries or noteworthy surgical outcomes. On February 20th 2013 , singer Lady Gaga underwent hip surgery and thanked her fans for the love and support (2500 tweets, $h_{\text{avg}} = 6.32$), (Derschowitz, 2013). During November 25 2014, #prayfortrevor was widely shared (4647 tweets, $h_{\text{avg}} = 6.47$) in support of a young stroke patient undergoing brain surgery, which was a miraculous success (Boots, 2016). The positive event occuring on June 29th-30 2016 (5776 tweets, $h_{\text{avg}} = 6.38$) corresponded to a viral story about singer Zayn Malik whom helped crowdfund for an abused cat's surgery bills, (Ceron, 2016).

Comparing average tweet word happiness over time can help identify interesting trends within the data that may not be apparent from the raw frequency distribution. In Figure 5.2 average tweet happiness is plotted as a function of both month (large markers) and day (small markers) for surgery tweets (red crosses) and the unfiltered gardenhose reference sample (blue circles). Tweets mentioning 'surgery' were slightly more negative on average ($h_{\text{avg}} = 5.88$ v. 6.01). Extreme events (dips and peaks) tend to correspond to relevant real world events.

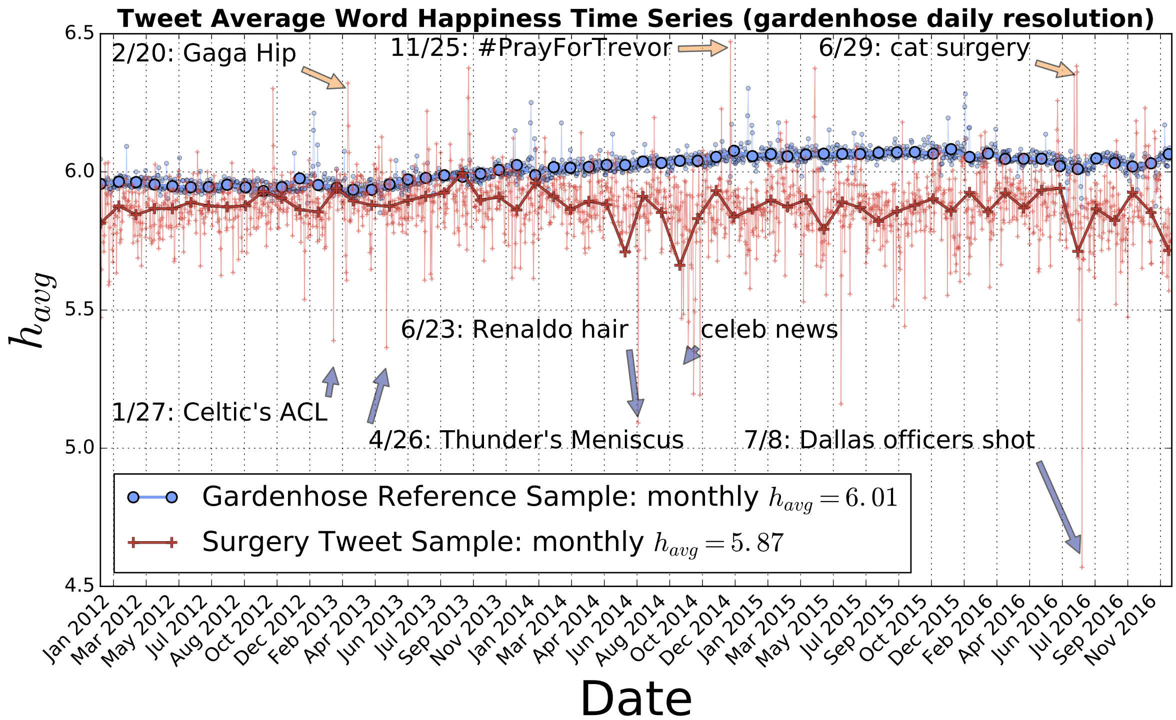


Figure 5.2: A time series of the average computed word happiness as a function of both month (large markers) and day (small makers) for the Surgery Tweets (red crosses) and Gardenhose reference tweets (blue circles) for comparison.

Topics relevant to several surgical types were also investigated. A term list – constructed by medical professionals – of surgical types helped compare and contrast prevalence and sentiments of tweets between disciplines. In Table 5.1 including each terms tweet count, it's relative weight among the sub-list, and it's relative average happiness score - colored relative to the sub-list's average computed word happiness (blue - negative, orange- positive). A comprehensive list additional key terms are provided in Appendix D.

Surgery Types				
Rank	Term	Tweet Count	Percent	h_{avg}
1	plastic	782,505	83.7%	5.72
2	cosmetic	118,540	12.7%	6.0
3	reconstructive	20,155	2.2%	5.92
4	reconstruction	13,453	1.4%	6.08
*	**Total**	934,653	100%	5.93
1	neuro	1,460	64.3%	5.92
2	neurosurgery	811	35.7%	5.96
*	**Total**	2,271	100%	5.94
1	ear	190,382	75.8%	5.67
2	nose	45,029	17.9%	5.72
3	throat	13,537	5.4%	5.68
4	ent	2,099	0.8%	5.92
*	**Total**	251,047	100%	5.75
1	ophthalmology	882	55.2%	5.93
2	ophthalmologist	716	44.8%	6.16
*	**Total**	1,598	100%	6.04
1	trauma	5,533	90.5%	5.75
2	acute care	584	9.5%	5.97
*	**Total**	6,117	100%	5.86
1	colorectal	1,843	100.0%	5.5
1	general	20,785	100.0%	5.95
1	transplant	15,619	95.8%	6.12
2	transplantation	690	4.2%	6.1
*	**Total**	16,309	100%	6.11
1	oncology	950	100.0%	5.58
*	**Total**	950	100%	5.58
1	vascular	4,490	100.0%	5.96
1	thoracic	3,454	47.1%	5.8
2	cardiothoracic	2,084	28.4%	6.16
3	cardiovascular	1,802	24.6%	5.94
*	**Total**	7,340	100%	5.97
1	orthopedic	9,702	64.8%	5.98
2	orthopaedic	3,405	22.7%	5.95
3	ortho	1,865	12.5%	5.91
*	**Total**	14,972	100%	5.95
1	dental	24,706	100.0%	5.91
1	weight loss	56,735	64.4%	5.93
2	bariatric	26,545	30.1%	5.68
3	minimally invasive	4,704	5.3%	5.8
4	minimally-invasive	179	0.2%	5.58
*	**Total**	88,163	100%	5.75
1	urology	1,433	73.5%	6.23
2	urological	517	26.5%	5.81
*	**Total**	1,950	100%	6.02
1	pediatric	4,423	100.0%	5.97

Perioperative				
Rank	Term	Tweet Count	Percent	h_{avg}
1	sleep	19,491	18.1%	5.68
2	procedure	17,227	16.0%	5.87
3	operation	16,344	15.2%	5.65
4	medicine	16,106	14.9%	5.89
5	anesthesia	12,448	11.5%	5.75
6	knife	6,354	5.9%	5.7
7	operating	5,339	5.0%	5.93
8	medication	3,895	3.6%	5.47
9	drug	3,599	3.3%	5.45
10	asleep	2,893	2.7%	5.64
11	scalpel	1,662	1.5%	5.71
12	knocked out	1,022	0.9%	5.8
13	opioid	806	0.7%	5.76
14	rx	623	0.6%	5.75
*	**Total**	107,809	100%	5.72

Types of Tissue				
Rank	Term	Tweet Count	Percent	h_{avg}
1	hair	33,280	100.0%	5.46
*	**Total**	33,280	100%	5.46
1	skin	21,057	98.9%	5.58
2	dermal	224	1.1%	5.7
*	**Total**	21,281	100%	5.64
1	meniscus	34,802	55.9%	5.38
2	ligament	12,335	19.8%	5.32
3	tendon	9,587	15.4%	5.5
4	muscle	5,525	8.9%	5.44
*	**Total**	62,249	100%	5.41
1	bone	22,792	100.0%	5.42
1	nerve	5,926	100.0%	5.45
1	collagen	269	100.0%	5.78
*	**Total**	269	100%	5.78
1	fat	13,518	100.0%	5.65
*	**Total**	13,518	100%	5.65
1	teeth	51,671	79.1%	5.9
2	tooth	9,253	14.2%	5.83
3	gum	3,280	5.0%	5.74
4	gums	1,095	1.7%	5.67
*	**Total**	65,299	100%	5.79

Table 5.1: **Surgery Types** A table key word Twitter stats related to relevant surgical types. The relative computed average happiness h_{avg} for each term is colored relative to the group's average (blue- negative, orange - positive).

In Figure 5.3 word shift graphs illustrate the words responsible for shifts in computed word happiness across distributions. These graphs help us further understand how sentiment calculations (h_{avg}) compare and provide insights into relevant topics causing the negative or positive shift. On the left, all collected 'surgery' tweets are compared to the random reference unfiltered Gardenhose sample. Surgery tweets were less positive ($h_{\text{avg}} = 5.88$ v. 6.01) due to a decrease in the positive words 'love', 'me', 'happy' ,'you', as well as a relative increase in the negative words 'cancer', 'hospital', 'pain', 'loss', 'emergency' ,'injury' ,'ill', 'tumor' ,'patients', 'dying', 'bad' among others. There was also a relative increase in the positive words 'heart', 'successful', 'hope', 'luck', 'prayers', 'brain', 'recovery'.

From the previous term list, we noticed 'plastic' surgery tweets were measurably less positive than 'cosmetic' surgery tweets. This relationship is investigated by the Word Shift Graph on the right. Here, tweets mentioning 'Plastic' surgery (T_{comp}) were compared to 'Cosmetic' surgery posts (T_{ref}). Plastic surgery tweets were less positive ($h_{\text{avg}} = 5.73$ v. 6.00) due to a relative increase in negative words 'ugly', 'bad', 'never', 'deny', 'fails', 'worst', 'drugs', 'addicted', 'worried', 'die', 'wrong', 'fake', and 'warning' as well as a relative increase of the negative words 'ugly', 'never', 'bad', 'deny', 'fails', 'worst', 'addicted', and 'warning'.

In Table 5.2, we present the most frequently appearing hashtags from tweets mentioning 'surgery'. The left side is sorted by frequency of appearance, while the right is sorted with respect to the ambient average happiness. Interestingly, here the terms #cosmeticsurgery ($h_{\text{avg}} = 6.16$) and #plasticsurgery ($h_{\text{avg}} = 6.04$) appear much more positive than #plastic ($h_{\text{avg}} = 5.89$) and #cosmetic ($h_{\text{avg}} = 5.99$) . In general, surgeons and medical professionals use #cosmeticsurgery and #plasticsurgery to promote their practices, (Branford et al., 2016), while #plastic and #cosmetic may be more commonly tweeted by the general public. This may be an indication of a public misconception regarding the scope of each medical discipline.

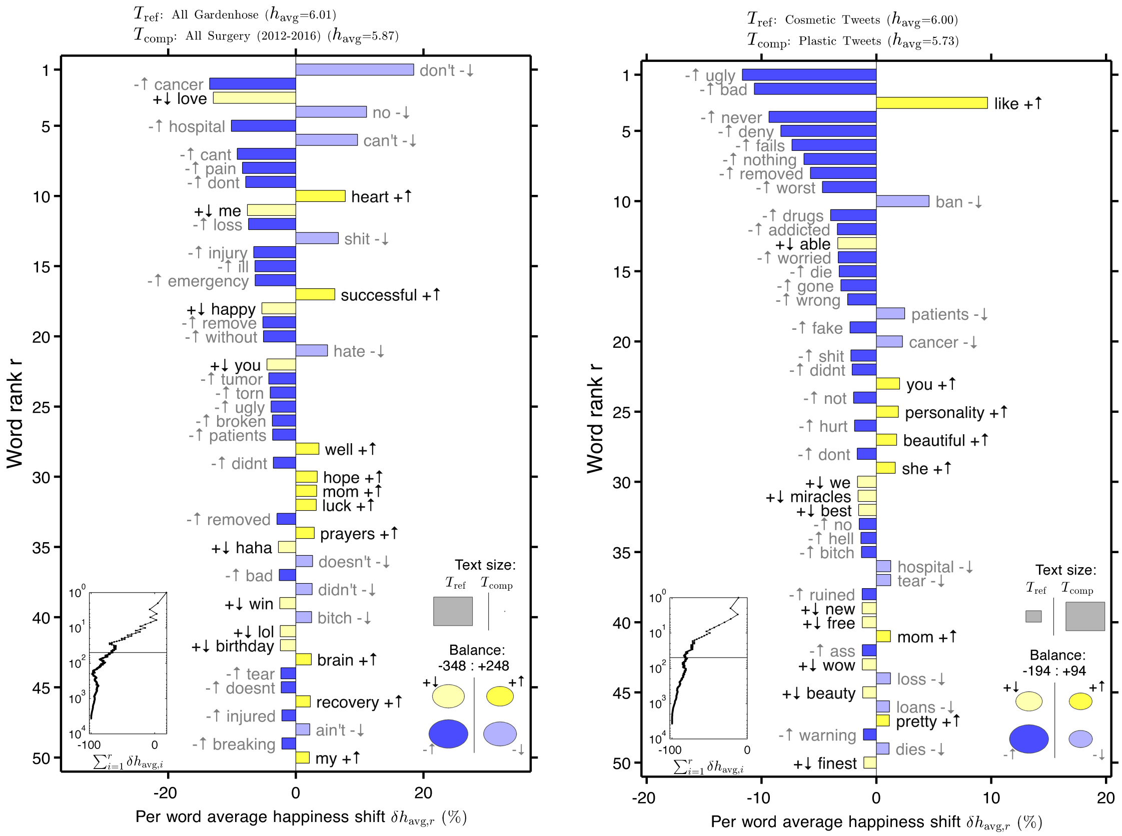


Figure 5.3: (Left) A word shift graph comparing tweets collected mentioning surgery (T_{comp}) to a random unfiltered reference sample of tweets from the same time frame. Surgery tweets were slightly less positive ($h_{\text{avg}} = 5.88$ v. 6.01) due to an increase in negative words including ‘cancer’, ‘hospital’ , ‘fight’, ‘pain’, ‘risk’. This set had a relative increase in positive words ‘heart’, ‘successful’, ‘well’, ‘hope’, ‘luck,’prayers’,’brain’,’recovery’. (Right) This graph compares surgery tweets mentioning ‘Plastic’ (T_{comp}) to ‘Cosmetic’ surgery tweets (T_{ref}). Plastic surgery tweets were less positive ($h_{\text{avg}} = 5.73$ v. 6.02) due to a relative increase in negative words ‘ugly’, ‘bad’, ‘never’, ‘deny’, ‘fails’, ‘worst’, ‘drugs’, ‘addicted’, ‘worried’, ‘die’, ‘wrong’, ‘fake’, and ‘warning’ as well as a relative increase of the negative words ‘ugly’, ‘never’, ‘bad’, ‘deny’, ‘fails’, ‘worst’, ‘addicted’, and ‘warning’.

Frequency Sorted Surgery Hashtags(#)

Rank	Term	Tweets	Users	h_{avg}
1	#surgery	69,599	46,356	5.88
2	#jobs	27,470	3,911	6.18
3	#job	24,389	2,972	6.14
4	#health	15,092	5,956	5.61
5	#news	14,490	4,534	5.65
6	#nfl	10,031	2,801	5.55
7	#mlb	9,175	2,260	5.76
8	#love	8,392	8,215	6.49
9	#hiring	8,387	985	6.24
10	#weightloss	8,320	2,752	5.49
11	#nba	8,204	2,655	5.74
12	#respect	7,280	7,084	4.81
13	#sports	5,469	1,164	5.66
14	#donate	5,249	2,972	6.29
15	#plasticsurgery	5,220	2,931	6.04
16	#healthcare	5,188	1,944	5.91
17	#cancer	5,072	3,213	5.68
18	#kca	5,043	4,330	6.03
19	#prayfortrevor	4,858	2,692	7.04
20	#nursing	4,793	615	6.13
21	#fitness	4,383	999	5.33
22	#diet	4,255	840	5.38
23	#blood	3,990	2,307	5.89
24	#healthy	3,698	523	5.33
25	#workout	3,615	354	5.06
26	#heart	3,591	1,352	6.3
27	#rt	3,472	1,407	5.92
28	#storybehindmyscar	3,442	3,391	5.5
29	#nervous	3,183	3,137	5.97
30	#cosmetic	3,170	1,925	5.99
31	#vote5sos	3,126	2,739	6.01
32	#wwe	3,079	1,628	5.77
33	#loveyou	2,998	2,971	6.39
34	#cosmeticsurgery	2,883	1,000	6.16
35	#plastic	2,871	2,002	5.89
36	#medical	2,864	1,508	5.9
37	#blessed	2,853	2,803	6.38
38	#fb	2,820	1,294	5.86
39	#fit	2,800	270	5.3
40	#fatloss	2,798	223	5.28
41	#gofundme	2,697	1,960	6.32
42	#rn	2,689	412	6.14
43	#mufc	2,688	2,402	5.8
44	#nhl	2,687	975	5.62
45	#tbt	2,587	2,542	6.11
46	#beauty	2,573	1,166	5.85
47	#nhs	2,431	1,857	5.69
48	#breastcancer	2,413	1,625	5.97
49	#nurse	2,398	697	6.16
50	#hospital	2,325	1,819	5.88
*	**Total**	345,424	128,177	5.87

Sentiment Sorted Surgery Hashtags(#)

Rank	Term	Tweets	Users	h_{avg}
1	#prayfortrevor	4,858	2,692	7.04
2	#love	8,392	8,215	6.49
3	#loveyou	2,998	2,971	6.39
4	#blessed	2,853	2,803	6.38
5	#gofundme	2,697	1,960	6.32
6	#heart	3,591	1,352	6.3
7	#donate	5,249	2,972	6.29
8	#hiring	8,387	985	6.24
9	#jobs	27,470	3,911	6.18
10	#nurse	2,398	697	6.16
11	#cosmeticsurgery	2,883	1,000	6.16
12	#jobsearch	2,324	258	6.15
13	#job	24,389	2,972	6.14
14	#rn	2,689	412	6.14
15	#nursing	4,793	615	6.13
16	#tbt	2,587	2,542	6.11
17	#plasticsurgery	5,220	2,931	6.04
18	#kca	5,043	4,330	6.03
19	#vote5sos	3,126	2,739	6.01
20	#cosmetic	3,170	1,925	5.99
21	#nervous	3,183	3,137	5.97
22	#breastcancer	2,413	1,625	5.97
23	#healthcare	5,188	1,944	5.91
24	#medical	2,864	1,508	5.9
25	#blood	3,990	2,307	5.89
26	#plastic	2,871	2,002	5.89
27	#hospital	2,325	1,819	5.88
28	#surgery	69,599	46,356	5.88
29	#scared	2,280	2,240	5.87
30	#fb	2,820	1,294	5.86
31	#beauty	2,573	1,166	5.85
32	#mufc	2,688	2,402	5.8
33	#wwe	3,079	1,628	5.77
34	#mlb	9,175	2,260	5.76
35	#nba	8,204	2,655	5.74
36	#nhs	2,431	1,857	5.69
37	#cancer	5,072	3,213	5.68
38	#sports	5,469	1,164	5.66
39	#news	14,490	4,534	5.65
40	#nhl	2,687	975	5.62
41	#health	15,092	5,956	5.61
42	#nfl	10,031	2,801	5.55
43	#storybehindmyscar	3,442	3,391	5.5
44	#weightloss	8,320	2,752	5.49
45	#diet	4,255	840	5.38
46	#fitness	4,383	999	5.33
47	#healthy	3,698	523	5.33
48	#fit	2,800	270	5.3
49	#fatloss	2,798	223	5.28
50	#workout	3,615	354	5.06
*	**Total**	344,232	128,980	5.87

Table 5.2: **50 Most Frequently Tweeted Hashtags** (#) from all collected surgery tweets (left) and the same list sorted by average happiness score. The relative computed average happiness h_{avg} for each tag is colored relative to the group average (blue- negative, orange - positive).

5.5 Discussion

We have demonstrated the potential of applying hedonometrics to extract emotionally charged themes and compare relevant surgical topics discussed on Twitter. Our framework can be readily applied across healthcare disciplines. Investigating the sentiment landscape over time helped unravel relevant global themes discussed throughout the data. Word shift graphs were essential for comprehending the emotionally charged terms that drove shifts in average computed word happiness. These fundamental graphical tools show great promise for extracting sentiments and insights from large text repositories.

The sentiment analysis may have uncovered an apparent public misconception regarding the scope of plastic surgery and its connotation. In particular, plastic surgery is defined as reconstructive surgical specialty for repairing "facial and body defects due to birth disorder, trauma, burns, and disease" whereas cosmetic surgery focuses on appearance enhancement (e.g. breast augmentation, facial contouring, etc),(The-American-Board-Of-Cosmetic-Surgery, 2017). By definition, plastic and reconstructive surgery are identical disciplines, yet plastic surgery tweets ($h_{\text{avg}} = 5.73$) were notably more prevalent and less positive than reconstructive ($h_{\text{avg}} = 5.94$).

Interestingly, there are more references describing vanity (i.e. more 'ugly', 'addiction', 'fake', and addiction) as well as aesthetics (more 'beautiful', 'pretty') in plastic surgery tweets in comparison to cosmetic surgery tweets. There were also less mentions of clinical words like 'patients', 'cancer', 'hospital' in plastic posts relative to cosmetic. This may be evidence that the public conception of plastic and cosmetic surgery may be erroneous flipped - the scope of plastic surgery is clinically motivated whereas cosmetic is purely vanity. This misinterpretation may be subverted with public service announcements as well as reality television programs, (Fogel and King, 2014; Crockett et al., 2007), which have been shown to potentially influence the patient decision making process.

While our study primarily focused on English posts, online communication among medical institutions and professionals have gained traction outside of the United States, (Sugawara et al., 2016). Thus, a multi-lingual extension of this analysis could be interesting.

Future studies may integrate other social media platforms. An analysis using Instagram data showed relevant indicators for depression, (Reece and Danforth, 2017). In particular, Facebook, Youtube, and

Instagram were found to be the most engaging platform for plastic surgery patients, (Sorice et al., 2017). These other platforms should also be investigated for mining patient sentiments and perceptions.

5.6 Conclusion

Our results validate Twitter as an informative resource for health care professionals. Future studies should investigate how to further integrate social media platforms into the internet of things (IoT) healthcare movement. We have demonstrated the potential of using hedonometrics to evaluate public perceptions regarding surgical disciplines as well as identifying emotional trends in the global online discussion across Twitter.

5.7 Remarks on Extending Our Framework

In this section we provide a generalized outline of our hedonometric approach for mining sentiments and emotional themes within large sets of text. We leave this for future data scientists and enthusiasts interested in applying our framework.

Topic Formulation: In many cases, data scientists can provide their skills in both data mining and computation to synergize with researchers and industry professionals from a multitude of disciplines. It can be beneficial to coordinate focus groups with stakeholders and/or experts within the field of study when framing the research project. In this way, a clear plan of action should be coordinated to identify the topics of interests, potential problems/limitations, and how to provide contributions in the investigated field. The analysis can then be tailored towards answering fundamental questions that may otherwise be unattainable without collaboration. Stakeholders and industry professionals can be particularly useful for providing feedback into key-word term lists for pulling relevant subsets of data in the analysis. The data scientist should also be responsible for managing expectations and providing insight into what may or may not be possible using their skill-set and data available.

User Frequency Distributions: This is a caveat for social media related projects. It is desirable to isolate high frequency tweet entities, since they can non-trivially impact your analysis. Even a small group of these “content polluters” can sway the inferred positivity of results and even undermine the credibility of a sentiment analysis. Hence, it is of utmost importance to consider the distribution of analyzed users as a function of their number of captured tweets. We normally plot this relationship on a log-scale and provide a statistical summary of their tweet frequency behaviors. In most situations, the tail of the distribution will consist of the high frequency contributors. From here, you can establish a tweet threshold for removing these individuals from the analysis. This tweet threshold and number of users removed should be reported in your methodology.

Time Series Analyses: A useful initial approach should include plotting the relationships between tweet frequency and corresponding average happiness score (h_{avg}) of key terms and phrases as a function of time. This can help isolate story arcs and timelines that need further investigation. When plotting h_{avg} be sure to choose appropriate sized bins (i.e., days, weeks, months, ...) to ensure their are enough terms for a meaningful calculation. In practice, we remove neutral words from the time series analysis via a predefined stop window (usually $4.5 <= h_{\text{avg}} <= 5.5$) to bolster any emotional signals. Relative extrema (i.e. peaks and dips) in each feature landscape should be further analyzed as these points tend to correspond to interesting real-world events.

Word Shift Graphs: These are essential tools for deciphering the cause of a shift in average computed word happiness between sets of texts. Word shift graphs list the key terms and their relative weights responsible, which is a distinguishing feature between hedonometrics and other sentiment analysis techniques. For these comparisons, it is important to have a sizable sample of tweets in both the reference and comparison corpora. In general, the metric performs better on larger sets of text, however for less prevalent topics, try to create bins of no less than 1000 tweets for an appropriate comparison.

Word shift graphs can also help recognize prevalent emotionally charged terms that may not be relevant to the studied topic. For example, before language detection was prevalent on Twitter, there was the potential of overlap between Spanish and English tweets being compiled. In some cases, the word ‘sin’ was influencing the sentiment shift since it is a negative word ($h_{\text{avg}} = 2.64$) in English with religious undertones, while it is simply a frequently appearing preposition (translation ‘without’) in Spanish tweets. These types of terms were then excluded from the analysis as a means of noise reduction.

Sentiment Topic Comparisons: It may also be practical to compare sets of tweets mentioning relevant key terms and phrases as well as hashtags. Hashtags are terms preceded by a (#) for self organization of general topics within a post. These concise yet informative attributes can be easily tracked as they percolate throughout the social network. When considering these term lists and topics, it is important to also report the number of individuals (i.e., user count) along with the corresponding tweet frequency and average happiness score. This helps stem potential noise attributed to high frequency bloggers.

Precautionary Disclaimers: It is necessary to be wary of assuming a text's sentiment score allows us to generalize that topic as positive or negative in actuality. This can become particularly ambiguous when the average happiness score fluctuates around the center of the hedonometer's scale ($4.5 <= h_{\text{avg}} <= 5.5$). It is better to think of h_{avg} as a comparative metric to distinguish themes between groups of text, and to understand how their emotional signals may differ. In this way, we can say one group exhibits more or less emotionally charged terms, which drive a measurable difference in their average computed word happiness. We can then further investigate these claims using word shift graphs to identify the terms responsible and try to match them to real world narratives. These principles should be added as a disclaimer to help guide interpretations of hedonometric results.

Concluding Remarks: As we progress through the Age of Information, our world is generating an ever-increasing wealth of linguistic data. Hedonometrics is an effective tool that can help researchers and enthusiasts unlock key insights from large and noisy word frequency distributions. These principles can be readily applied across various disciplines to investigate underlying themes and emotional story arcs within large sets of text. We'll leave this as an exercise for the reader.

5.8 References

Alajajian, S. E., J. R. Williams, A. J. Reagan, S. C. Alajajian, M. R. Frank, L. Mitchell, J. Lahne, C. M. Danforth, and P. S. Dodds (2017). The Lexicocalorimeter: Gauging public health through caloric input and output on social media. *PloS ONE 12*(2), e0168893.

Alshaikh, F., F. Ramzan, S. Rawaf, and A. Majeed (2014). Social network sites as a mode to collect health data: A systematic review. *Journal of medical Internet research 16*(7).

Boots, M. T. (2016, September). Young life leader's recovery after stroke 'defies medical explanation'.

Branford, O. A., P. Kamali, R. J. Rohrich, D. H. Song, P. Mallucci, D. Z. Liu, D. Lang, K. Sun, M. Stubican, and S. J. Lin (2016). # plasticsurgery. *Plastic and reconstructive surgery 138*(6), 1354–1365.

Ceron, E. (2016, June). Cat's owner clarifies rumor about zayn malik paying for its surgery.

Chang, J. B., S. L. Woo, and P. S. Cederna (2015). Worth the "likes"? The use of facebook among plastic surgeons and its perceived impact. *Plastic and reconstructive surgery 135*(5), 909e–918e.

Chopan, M., E. Clark, L. Sayadi, P. Buchanan, and K. Maguire (2018). Plastic surgery and social media - examining perceptions. *Plastic and Reconstructive Surgery.*.

Chou, W.-Y. S., Y. M. Hunt, E. B. Beckjord, R. P. Moser, and B. W. Hesse (2009). Social media use in the united states: Implications for health communication. *Journal of medical Internet research 11*(4).

Christensen, J. (2014, September). Joan rivers: How a minor elective surgery could end in death.

Clark, E. M., C. A. Jones, J. R. Williams, A. N. Kurti, M. C. Norotsky, C. M. Danforth, and P. S. Dodds (2016). Vaporous marketing: Uncovering pervasive electronic cigarette advertisements on twitter. *PloS ONE 11*(7), e0157304.

Clark, E. M., J. R. Williams, C. A. Jones, R. A. Galbraith, C. M. Danforth, and P. S. Dodds (2016). Sifting robotic from organic text: A natural language approach for detecting automation on Twitter. *Journal of Computational Science 16*, 1–7.

Cody, E. M., A. J. Reagan, L. Mitchell, P. S. Dodds, and C. M. Danforth (2015). Climate change sentiment on Twitter: An unsolicited public opinion poll. *PloS ONE 10*(8), e0136092.

Crannell, W. C., E. Clark, C. Jones, T. A. James, and J. Moore (2016). A pattern-matched Twitter analysis of U.S. cancer-patient sentiments. *Journal of Surgical Research 206*(2), 536–542.

Crockett, R. J., T. Pruzinsky, and J. A. Persing (2007). The influence of plastic surgery “reality tv” on cosmetic surgery patient expectations and decision making. *Plastic and reconstructive surgery 120*(1), 316–324.

Derschowitz, J. (2013, February). Lady gaga undergoes surgery after canceling tour.

Dodds, P. S., E. M. Clark, S. Desu, M. R. Frank, A. J. Reagan, J. R. Williams, L. Mitchell, K. D. Harris, I. M. Kloumann, J. P. Bagrow, K. Megerdoomian, M. T. McMahon, B. F. Tivnan, and C. M. Danforth (2015). Human language reveals a universal positivity bias. *Proceedings of the National Academy of Sciences 112*(8), 2389–2394.

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

Eberlin, K. R., G. Perdikis, L. Damitz, D. J. Krochmal, L. K. Kalliainen, S. C. Bonawitz, A. H. P. Committee, et al. (2018). Electronic communication in plastic surgery: Guiding principles from the american society of plastic surgeons health policy committee. *Plastic and reconstructive surgery 141*(2), 500–505.

Fogel, J. and K. King (2014). Perceived realism and twitter use are associated with increased acceptance of cosmetic surgery among those watching reality television cosmetic surgery programs. *Plastic and reconstructive surgery 134*(2), 233–238.

Forsberg, C. (2014, February). Rajon rondo gets surgery on aclmeniscus.

Golliver, B. (2015, February). Bulls’ derrick rose to undergo right knee surgery for torn meniscus.

Grajales III, F. J., S. Sheps, K. Ho, H. Novak-Lauscher, and G. Eysenbach (2014). Social media: A review and tutorial of applications in medicine and health care. *Journal of medical Internet research 16*(2).

Keller, B., A. Labrique, K. M. Jain, A. Pekosz, and O. Levine (2014). Mind the gap: Social media engagement by public health researchers. *Journal of medical Internet research 16*(1).

Linshi, J. (2014, September). Goldfish undergoes surgery for removal of deadly tumor.

Lui, D., J. McDonald, A. de Beaux, B. Tulloh, and R. Brady (2017). Contemporary engagement with social media amongst hernia surgery specialists. *Hernia 21*(4), 509–515.

MANNY FERNANDEZ, R. P.-P. and J. E. BROMWICH (2016, July). Five dallas officers were killed as payback, police chief says.

Mitchell, L., M. R. Frank, K. D. Harris, P. S. Dodds, and C. M. Danforth (2013). The geography of happiness: Connecting twitter sentiment and expression, demographics, and objective characteristics of place. *PloS one 8*(5), e64417.

Reagan, A. J., B. Tivnan, J. R. Williams, C. M. Danforth, and P. S. Dodds (2015). Benchmarking sentiment analysis methods for large-scale texts: A case for using continuum-scored words and word shift graphs. *arXiv preprint arXiv:1512.00531*.

Reece, A. G. and C. M. Danforth (2017). Instagram photos reveal predictive markers of depression. *EPJ Data Science 6*(1), 15.

Reece, A. G., A. J. Reagan, K. L. Lix, P. S. Dodds, C. M. Danforth, and E. J. Langer (2016). Forecasting the onset and course of mental illness with Twitter data. *arXiv preprint arXiv:1608.07740*.

Reporter, D. M. (2014, January). It was a complete success considering the scale': Niall Horan walks with crutches after undergoing major knee surgery 'much larger than anticipated.

Sheets, C. A. (2014, June). Did ronaldo cut his hair in solidarity with a young brain tumor patient?

Sorice, S. C., A. Y. Li, J. Gilstrap, F. L. Canales, and H. J. Furnas (2017). Social media and the plastic surgery patient. *Plastic and reconstructive surgery 140*(5), 1047–1056.

Stephen Jones, M. F. and A. Himmelsbach (2013, March). Lacking the votes for passage, house calls off Obamacare repeal vote.

Sugawara, Y., H. Narimatsu, A. Tsuya, A. Tanaka, and A. Fukao (2016). Medical institutions and twitter: A novel tool for public communication in japan. *JMIR public health and surveillance 2*(1).

Thackeray, R., B. T. Crookston, and J. H. West (2013). Correlates of health-related social media use among adults. *Journal of medical Internet research 15*(1).

The-American-Board-Of-Cosmetic-Surgery (2017). Cosmetic surgery vs. plastic surgery.

Vardanian, A. J., N. Kusnezov, D. D. Im, J. C. Lee, and R. Jarrahy (2013). Social media use and impact on plastic surgery practice. *Plastic and reconstructive surgery 131*(5), 1184–1193.

von Horn, S. (2013, April). Russell westbrook injury: Thunder star has torn right meniscus, out indefinitely.

Zurilla, C. D. (2014, September). Justin bieber 'busted' eardrum cliff diving, he says, may need surgery.

Bibliography

Abadi, M., A. Agarwal, P. Barham, E. Brevdo, Z. Chen, C. Citro, G. S. Corrado, A. Davis, J. Dean, M. Devin, et al. (2016). Tensorflow: Large-scale machine learning on heterogeneous distributed systems. *arXiv preprint arXiv:1603.04467*.

Alajajian, S. E., J. R. Williams, A. J. Reagan, S. C. Alajajian, M. R. Frank, L. Mitchell, J. Lahne, C. M. Danforth, and P. S. Dodds (2017). The Lexicocalorimeter: Gauging public health through caloric input and output on social media. *PloS ONE 12*(2), e0168893.

Alshaikh, F., F. Ramzan, S. Rawaf, and A. Majeed (2014). Social network sites as a mode to collect health data: A systematic review. *Journal of medical Internet research 16*(7).

American Society of Addiction Medicine. (2008). Public policy statement on nicotine addiction and tobacco. *Chevy Chase (MD): American Society of Addiction Medicine*.

Aphinyanaphongs, Y., A. Lulejian, D. P. Brown, R. Bonneau, and P. Krebs (2016). Text classification for automatic detection of e-cigarette use and use for smoking cessation from twitter: A feasibility pilot. In *Pacific Symposium on Biocomputing. Pacific Symposium on Biocomputing*, Volume 21, pp. 480. NIH Public Access.

Ashley, D., D. Burns, M. Djordjevic, E. Dybing, N. Gray, S. Hammond, J. Henningfield, M. Jarvis, K. Reddy, C. Robertson, et al. (2007). The scientific basis of tobacco product regulation. *World Health Organization technical report series* (951), 1–277.

Attai, D. J., M. S. Cowher, M. Al-Hamadani, J. M. Schoger, A. C. Staley, and J. Landercasper (2015). Twitter social media is an effective tool for breast cancer patient education and support: Patient-reported outcomes by survey. *Journal of Medical Internet Research 17*(7).

Avdalovic, M. V. and S. Murin (2012). Electronic cigarettes: no such thing as a free lunch? or puff. *CHEST Journal 141*(6), 1371–1372.

Bender, J. L., M.-C. Jimenez-Marroquin, and A. R. Jadad (2011). Seeking support on facebook: a content analysis of breast cancer groups. *Journal of Medical Internet Research 13*(1).

Benevenuto, F., G. Magno, T. Rodrigues, and V. Almeida (2010). Detecting spammers on twitter. In *In Collaboration, Electronic messaging, Anti-Abuse and Spam Conference (CEAS*.

Blunsom, P., E. Grefenstette, and N. Kalchbrenner (2014). A convolutional neural network for modelling sentences. In *Proceedings of the 52nd Annual Meeting of the Association for Computational Linguistics*. Proceedings of the 52nd Annual Meeting of the Association for Computational Linguistics.

Boots, M. T. (2016, September). Young life leader's recovery after stroke 'defies medical explanation'.

Branford, O. A., P. Kamali, R. J. Rohrich, D. H. Song, P. Mallucci, D. Z. Liu, D. Lang, K. Sun, M. Stubican, and S. J. Lin (2016). # plasticsurgery. *Plastic and reconstructive surgery 138*(6), 1354–1365.

Brenner, J. and A. Smith (2013). 72% of online adults are social networking site users. *Washington, DC: Pew Internet & American Life Project*.

Britz, D. (2015a). Implementing a CNN for text classification in tensorflow.

Britz, D. (2015b, November). Understanding convolutional neural networks for NLP.

Broniatowski, D. A., M. J. Paul, and M. Dredze (2013). National and local influenza surveillance through twitter: An analysis of the 2012-2013 influenza epidemic. *PloS one 8*(12), e83672.

Brown, G., T. Howe, M. Ihbe, A. Prakash, and K. Borders (2008). Social networks and context-aware spam. In *Proceedings of the 2008 ACM Conference on Computer Supported Cooperative Work*, CSCW '08, New York, NY, USA, pp. 403–412. ACM.

Bunnell, R. E., I. T. Agaku, R. Arrazola, B. J. Apelberg, R. S. Caraballo, C. G. Corey, B. Coleman, S. R. Dube, and B. A. King (2014). Intentions to smoke cigarettes among never-smoking us middle and high school electronic cigarette users, national youth tobacco survey, 2011-2013. *Nicotine & Tobacco Research*, ntu166.

CA, L., S. IK, Y. H, G. J, O. DJ, and et al. (2015). Vapors produced by electronic cigarettes and e-juices with flavorings induce toxicity, oxidative stress, and inflammatory response in lung epithelial cells and in mouse lung. *PloS one 10*(2), e0116732.

Callahan-Lyon, P. (2014). Electronic cigarettes: human health effects. *Tobacco control 23*(suppl 2), ii36–ii40.

Cameron, J. M., D. N. Howell, J. R. White, D. M. Andrenyak, M. E. Layton, and J. M. Roll (2014). Variable and potentially fatal amounts of nicotine in e-cigarette nicotine solutions. *Tobacco control 23*(1), 77–78.

Centers for Disease Control and Prevention (CDC and others (2011). Quitting smoking among adults–united states, 2001-2010. *MMWR. Morbidity and mortality weekly report 60*(44), 1513.

Ceron, E. (2016, June). Cat's owner clarifies rumor about zayn malik paying for its surgery.

Chang, J. B., S. L. Woo, and P. S. Cederna (2015). Worth the "likes"? The use of facebook among plastic surgeons and its perceived impact. *Plastic and reconstructive surgery 135*(5), 909e–918e.

Cho, J. H., E. Shin, and S.-S. Moon (2011). Electronic-cigarette smoking experience among adolescents. *Journal of Adolescent Health 49*(5), 542–546.

Chopan, M., E. Clark, L. Sayadi, P. Buchanan, and K. Maguire (2018). Plastic surgery and social media - examining perceptions. *Plastic and Reconstructive Surgery.*.

Chou, W.-Y. S., Y. M. Hunt, E. B. Beckjord, R. P. Moser, and B. W. Hesse (2009). Social media use in the united states: Implications for health communication. *Journal of medical Internet research 11*(4).

Christensen, J. (2014, September). Joan rivers: How a minor elective surgery could end in death.

Chu, Z., S. Gianvecchio, H. Wang, and S. Jajodia (2010). Who is tweeting on twitter: Human, bot, or cyborg? In *Proceedings of the 26th Annual Computer Security Applications Conference*, ACSAC '10, New York, NY, USA, pp. 21–30. ACM.

Chu, Z., S. Gianvecchio, H. Wang, and S. Jajodia (2012). Detecting automation of twitter accounts: Are you a human, bot, or cyborg? *Dependable and Secure Computing, IEEE Transactions on 9*(6), 811–824.

Clark, E. M., C. A. Jones, J. R. Williams, A. N. Kurti, M. C. Norotsky, C. M. Danforth, and P. S. Dodds (2016). Vaporous marketing: Uncovering pervasive electronic cigarette advertisements on twitter. *PloS ONE 11*(7), e0157304.

Clark, E. M., C. A. Jones, J. R. Williams, A. N. Kurti, M. C. Nortotsky, C. M. Danforth, and P. S. Dodds (2015). Vaporous marketing: Uncovering pervasive electronic cigarette advertisements on twitter. *arXiv preprint arXiv:1508.01843*.

Clark, E. M., J. R. Williams, R. A. Galbraith, C. M. Danforth, P. S. Dodds, and C. A. Jones (2015). Sifting robotic from organic text: A natural language approach for detecting automation on twitter. *arXiv preprint arXiv:1505.04342*.

Clark, E. M., J. R. Williams, C. A. Jones, R. A. Galbraith, C. M. Danforth, and P. S. Dodds (2016). Sifting robotic from organic text: A natural language approach for detecting automation on Twitter. *Journal of Computational Science 16*, 1–7.

Cobb, N. K., M. J. Byron, D. B. Abrams, and P. G. Shields (2010). Novel nicotine delivery systems and public health: the rise of the ?e-cigarette? *American journal of public health 100*(12), 2340–2342.

Cody, E. M., A. J. Reagan, L. Mitchell, P. S. Dodds, and C. M. Danforth (2015). Climate change sentiment on Twitter: An unsolicited public opinion poll. *PloS ONE 10*(8), e0136092.

CONTROL, C. F. D., PREVENTION, et al. (2014). The health consequences of smoking–50 years of progress: A report of the surgeon general. *Rockville, MD: US DEPARTMENT OF HEALTH AND HUMAN SERVICES*, 171.

Crannell, W. C., E. Clark, C. Jones, T. A. James, and J. Moore (2016). A pattern-matched Twitter analysis of U.S. cancer-patient sentiments. *Journal of Surgical Research 206*(2), 536–542.

Crockett, R. J., T. Pruzinsky, and J. A. Persing (2007). The influence of plastic surgery "reality tv" on cosmetic surgery patient expectations and decision making. *Plastic and reconstructive surgery 120*(1), 316–324.

Derschowitz, J. (2013, February). Lady gaga undergoes surgery after canceling tour.

Dickerson, J. P., V. Kagan, and V. Subrahmanian (2014). Using sentiment to detect bots on twitter: Are humans more opinionated than bots? In *Advances in Social Networks Analysis and Mining (ASONAM), 2014 IEEE/ACM International Conference on*, pp. 620–627. IEEE.

Dodds, P. S., E. M. Clark, S. Desu, M. R. Frank, A. J. Reagan, J. R. Williams, L. Mitchell, K. D. Harris, I. M. Kloumann, J. P. Bagrow, K. Megerdoomian, M. T. McMahon, B. F. Tivnan, and C. M. Danforth (2015). Human language reveals a universal positivity bias. *Proceedings of the National Academy of Sciences 112*(8), 2389–2394.

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

Dredze, M. (2012). How social media will change public health. *IEEE Intelligent Systems 27*(4), 81–84.

Duke, J. C., Y. O. Lee, A. E. Kim, K. A. Watson, K. Y. Arnold, J. M. Nonnemaker, and L. Porter (2014). Exposure to electronic cigarette television advertisements among youth and young adults. *Pediatrics 134*(1), e29–e36.

Dutra, L. M. and S. A. Glantz (2014). Electronic cigarettes and conventional cigarette use among us adolescents: a cross-sectional study. *JAMA pediatrics 168*(7), 610–617.

Eberlin, K. R., G. Perdikis, L. Damitz, D. J. Krochmal, L. K. Kalliainen, S. C. Bonawitz, A. H. P. Committee, et al. (2018). Electronic communication in plastic surgery: Guiding principles from the american society of plastic surgeons health policy committee. *Plastic and reconstructive surgery 141*(2), 500–505.

Eissenberg, T. (2010). Electronic nicotine delivery devices: ineffective nicotine delivery and craving suppression after acute administration. *Tobacco control 19*(1), 87–88.

Ferrara, E., O. Varol, C. Davis, F. Menczer, and A. Flammini (2014). The rise of social bots. *CoRR abs/1407.5225*.

Fogel, J. and K. King (2014). Perceived realism and twitter use are associated with increased acceptance of cosmetic surgery among those watching reality television cosmetic surgery programs. *Plastic and reconstructive surgery 134*(2), 233–238.

Forsberg, C. (2014, February). Rajon rondo gets surgery on aclmeniscus.

Fussell Sisco, H. and T. McCorkindale (2013). Communicating “Pink”: An analysis of the communication strategies, transparency, and credibility of breast cancer social media sites. *International Journal of Nonprofit and Voluntary Sector Marketing 18*(4), 287–301.

Gao, H., J. Hu, C. Wilson, Z. Li, Y. Chen, and B. Y. Zhao (2010). Detecting and characterizing social spam campaigns. In *Proceedings of the 10th ACM SIGCOMM Conference on Internet Measurement*, IMC ’10, New York, NY, USA, pp. 35–47. ACM.

Genkin, A., D. D. Lewis, and D. Madigan (2007). Large-scale Bayesian logistic regression for text categorization. *Technometrics 49*(3), 291–304.

Golliver, B. (2015, February). Bulls’ derrick rose to undergo right knee surgery for torn meniscus.

Goniewicz, M. L., J. Knysak, M. Gawron, L. Kosmider, A. Sobczak, J. Kurek, A. Prokopowicz, M. Jablonska-Czapla, C. Rosik-Dulewska, C. Havel, et al. (2014). Levels of selected carcinogens and toxicants in vapour from electronic cigarettes. *Tobacco control 23*(2), 133–139.

Goniewicz, M. L. and W. Zielinska-Danch (2012). Electronic cigarette use among teenagers and young adults in poland. *Pediatrics 130*(4), e879–e885.

Grajales III, F. J., S. Sheps, K. Ho, H. Novak-Lauscher, and G. Eysenbach (2014). Social media: A review and tutorial of applications in medicine and health care. *Journal of medical Internet research 16*(2).

Grana, R., N. Benowitz, and S. A. Glantz (2014). E-cigarettes a scientific review. *Circulation 129*(19), 1972–1986.

Grana, R. A. and P. M. Ling (2014). Smoking revolution: a content analysis of electronic cigarette retail websites. *American journal of preventive medicine 46*(4), 395–403.

Harris, D. (2013, Feb). Can evil data scientists fool us all with the world’s best spam? `goo.gl/psEguf`.

Hosmer Jr, D. W., S. Lemeshow, and R. X. Sturdivant (2013). *Applied logistic regression*, Volume 398. John Wiley & Sons.

Huang, J., R. Kornfield, G. Szczypka, and S. L. Emery (2014a). A cross-sectional examination of marketing of electronic cigarettes on twitter. *Tobacco control 23*(suppl 3), iii26–iii30.

Huang, J., R. Kornfield, G. Szczypka, and S. L. Emery (2014b). A cross-sectional examination of marketing of electronic cigarettes on twitter. *Tobacco control 23*(suppl 3), iii26–iii30.

InternetLiveStats (2017). Twitter usage statistics.

Johnson, R. and T. Zhang (2015). Semi-supervised convolutional neural networks for text categorization via region embedding. In *Advances in Neural Information Processing Systems*, pp. 919–927.

Johnston, L. D., J. G. Bachman, et al. (2014). Monitoring the future: National results on adolescent drug use: Overview of key findings.

Kandra, K. L., L. M. Ranney, J. G. Lee, and A. O. Goldstein (2014). Physicians? attitudes and use of e-cigarettes as cessation devices, north carolina, 2013. *PloS one 9*(7), e103462.

Kaplan, T., R. Pear, and E. Huetteman (2017, March). Lacking the votes for passage, house calls off Obamacare repeal vote.

Keller, B., A. Labrique, K. M. Jain, A. Pekosz, and O. Levine (2014). Mind the gap: Social media engagement by public health researchers. *Journal of medical Internet research 16*(1).

Kim, A. E., T. Hopper, S. Simpson, J. Nonnemaker, A. J. Lieberman, H. Hansen, J. Guillory, and L. Porter (2015). Using twitter data to gain insights into e-cigarette marketing and locations of use: An infoveillance study. *Journal of medical Internet research 17*(11).

Kim, Y. (2014). Convolutional neural networks for sentence classification. *arXiv preprint arXiv:1408.5882*.

Kosmider, L., A. Sobczak, M. Fik, J. Knysak, M. Zaciera, J. Kurek, and M. L. Goniewicz (2014). Carbonyl compounds in electronic cigarette vapors: effects of nicotine solvent and battery output voltage. *Nicotine & Tobacco Research 16*(10), 1319–1326.

Lamb, A., M. J. Paul, and M. Dredze (2013). Separating fact from fear: Tracking flu infections on Twitter. In *HLT-NAACL*, pp. 789–795.

Lee, K., J. Caverlee, and S. Webb (2010a). The social honeypot project: Protecting online communities from spammers. In *Proceedings of the 19th International Conference on World Wide Web*, WWW '10, New York, NY, USA, pp. 1139–1140. ACM.

Lee, K., J. Caverlee, and S. Webb (2010b). Uncovering social spammers: Social honeypots + machine learning. In *Proceedings of the 33rd International ACM SIGIR Conference on Research and Development in Information Retrieval*, SIGIR '10, New York, NY, USA, pp. 435–442. ACM.

Lee, K., B. D. Eoff, and J. Caverlee (2010). Devils, angels, and robots: Tempting destructive users in media. In W. W. Cohen and S. Gosling (Eds.), *ICWSM*. The AAAI Press.

Lee, K., B. D. Eoff, and J. Caverlee (2011). Seven months with the devils: a long-term study of content polluters on twitter. In *In AAAI Int?l Conference on Weblogs and Social Media (ICWSM*.

Lee, K., P. Tamilarasan, and J. Caverlee (2013). Crowdturfers, campaigns, and social media: Tracking and revealing crowdsourced manipulation of social media. In *ICWSM*.

Li, H., A. Mukherjee, B. Liu, R. Kornfield, and S. Emery (2014). Detecting campaign promoters on twitter using markov random fields. In *Data Mining (ICDM), 2014 IEEE International Conference on*, pp. 290–299. IEEE.

Linshi, J. (2014, September). Goldfish undergoes surgery for removal of deadly tumor.

Liu, B. (2010). Sentiment analysis and subjectivity. *Handbook of natural language processing 2*, 627–666.

Lui, D., J. McDonald, A. de Beaux, B. Tulloh, and R. Brady (2017). Contemporary engagement with social media amongst hernia surgery specialists. *Hernia 21*(4), 509–515.

MANNY FERNANDEZ, R. P.-P. and J. E. BROMWICH (2016, July). Five dallas officers were killed as payback, police chief says.

ME.ME (2017). The group proposing to cut breast cancer screening maternity care and contraceptive coverage. do you see the problem here? meme.

Mitchell, L., M. R. Frank, K. D. Harris, P. S. Dodds, and C. M. Danforth (2013). The geography of happiness: Connecting twitter sentiment and expression, demographics, and objective characteristics of place. *PloS one 8*(5), e64417.

MJ Lee, Lauren Fox, T. B. P. M. and A. Killough (2017, July). GOP Obamacare repeal bill fails in dramatic late-night vote.

Moorhead, S. A., D. E. Hazlett, L. Harrison, J. K. Carroll, A. Irwin, and C. Hoving (2013). A new dimension of health care: Systematic review of the uses, benefits, and limitations of social media for health communication. *Journal of Medical Internet Research 15*(4).

National Institute on Drug Abuse (2012). Research report series: Is nicotine addictive? *Bethesda (MD): National Institutes of Health, National Institute on Drug Abuse*.

Nielsen, F. Å. (2011). A new anew: Evaluation of a word list for sentiment analysis in microblogs. *arXiv preprint arXiv:1103.2903*.

Palazzolo, D. L. (2013). Electronic cigarettes and vaping: a new challenge in clinical medicine and public health. a literature review. *Frontiers in public health 1*.

Paul, M. J. and M. Dredze (2011). You are what you tweet: Analyzing Twitter for public health. *ICWSM 20*, 265–272.

Pearson, J. L., A. Richardson, R. S. Niaura, D. M. Vallone, and D. B. Abrams (2012). e-cigarette awareness, use, and harm perceptions in us adults. *American journal of public health 102*(9), 1758–1766.

Pennebaker, J. W., M. E. Francis, and R. J. Booth (2001). Linguistic inquiry and word count: Liwc 2001. *Mahway: Lawrence Erlbaum Associates 71*(2001), 2001.

Pepper, J. K., P. L. Reiter, A.-L. McRee, L. D. Cameron, M. B. Gilkey, and N. T. Brewer (2013). Adolescent males' awareness of and willingness to try electronic cigarettes. *Journal of Adolescent Health 52*(2), 144–150.

Ranganath, S., X. Hu, J. Tang, and H. Liu. Understanding and identifying advocates for political campaigns on social media.

Reagan, A. J., B. Tivnan, J. R. Williams, C. M. Danforth, and P. S. Dodds (2015). Benchmarking sentiment analysis methods for large-scale texts: A case for using continuum-scored words and word shift graphs. *arXiv preprint arXiv:1512.00531*.

Reece, A. G. and C. M. Danforth (2017). Instagram photos reveal predictive markers of depression. *EPJ Data Science 6*(1), 15.

Reece, A. G., A. J. Reagan, K. L. Lix, P. S. Dodds, C. M. Danforth, and E. J. Langer (2016). Forecasting the onset and course of mental illness with Twitter data. *arXiv preprint arXiv:1608.07740*.

Reporter, D. M. (2014, January). It was a complete success considering the scale': Niall Horan walks with crutches after undergoing major knee surgery 'much larger than anticipated.

Roesslein, J. (2009). Tweepy documentation. *5*.

Sadilek, A., H. A. Kautz, and V. Silenzio (2012). Modeling spread of disease from social interactions. In *ICWSM*.

Salton, G., A. Wong, and C.-S. Yang (1975). A vector space model for automatic indexing. *Communications of the ACM 18*(11), 613–620.

Sheets, C. A. (2014, June). Did ronaldo cut his hair in solidarity with a young brain tumor patient?

Shive, M., M. Bhatt, A. Cantino, J. Kvedar, and K. Jethwani (2013). Perspectives on acne: what twitter can teach health care providers. *JAMA Dermatology 149*(5), 621–622.

Sorice, S. C., A. Y. Li, J. Gilstrap, F. L. Canales, and H. J. Furnas (2017). Social media and the plastic surgery patient. *Plastic and reconstructive surgery 140*(5), 1047–1056.

Stephen Jones, M. F. and A. Himmelsbach (2013, March). Lacking the votes for passage, house calls off Obamacare repeal vote.

Subrahmanian, V., A. Azaria, S. Durst, V. Kagan, A. Galstyan, K. Lerman, L. Zhu, E. Ferrara, A. Flammini, F. Menczer, et al. (2016). The darpa twitter bot challenge. *arXiv preprint arXiv:1601.05140*.

Sugawara, Y., H. Narimatsu, A. Hozawa, L. Shao, K. Otani, and A. Fukao (2012). Cancer patients on twitter: a novel patient community on social media. *BMC Research Notes 5*(1), 699.

Sugawara, Y., H. Narimatsu, A. Tsuya, A. Tanaka, and A. Fukao (2016). Medical institutions and twitter: A novel tool for public communication in japan. *JMIR public health and surveillance 2*(1).

Sussan, T. E., S. Gajghate, R. K. Thimmulappa, J. Ma, J.-H. Kim, K. Sudini, N. Consolini, S. A. Cormier, S. Lomnicki, F. Hasan, et al. (2015). Exposure to electronic cigarettes impairs pulmonary anti-bacterial and anti-viral defenses in a mouse model. *PloS one 10*(2), e0116861.

Thackeray, R., B. T. Crookston, and J. H. West (2013). Correlates of health-related social media use among adults. *Journal of medical Internet research 15*(1).

The-American-Board-Of-Cosmetic-Surgery (2017). Cosmetic surgery vs. plastic surgery.

Thomas, K., C. Grier, and V. Paxson (2012). Adapting social spam infrastructure for political censorship. In *Presented as part of the 5th USENIX Workshop on Large-Scale Exploits and Emergent Threats*, Berkeley, CA. USENIX.

Thomas, K., C. Grier, D. Song, and V. Paxson (2011). Suspended accounts in retrospect: An analysis of twitter spam. In *Proceedings of the 2011 ACM SIGCOMM Conference on Internet Measurement Conference*, IMC '11, New York, NY, USA, pp. 243–258. ACM.

Trtchounian, A. and P. Talbot (2011). Electronic nicotine delivery systems: is there a need for regulation? *Tobacco control 20*(1), 47–52.

Twitter, I. (2017). Twitter streaming API.

US Department of Health and Human Services and others (2000). *Reducing tobacco use: a report of the Surgeon General*. US Department of Health and Human Services.

US Department of Health and Human Services and others (2004). The health consequences of smoking: a report of the surgeon general. *Atlanta, GA: US Department of Health and Human Services, Centers for Disease Control and Prevention, National Center for Chronic Disease Prevention and Health Promotion, Office on Smoking and Health 62*.

US Department of Health and Human Services and others (2010). How tobacco smoke causes disease: the biology and behavioral basis for smoking-attributable disease: a report of the surgeon general. *Atlanta,* *GA: US Department of Health and Human Services, Centers for Disease Control and Prevention, National Center for Chronic Disease Prevention and Health Promotion, Office on Smoking and Health 2*.

Vansickel, A. R., C. O. Cobb, M. F. Weaver, and T. E. Eissenberg (2010). A clinical laboratory model for evaluating the acute effects of electronic cigarettes: nicotine delivery profile and cardiovascular and subjective effects. *Cancer Epidemiology Biomarkers & Prevention 19*(8), 1945–1953.

Vardanian, A. J., N. Kusnezov, D. D. Im, J. C. Lee, and R. Jarrahy (2013). Social media use and impact on plastic surgery practice. *Plastic and reconstructive surgery 131*(5), 1184–1193.

von Horn, S. (2013, April). Russell westbrook injury: Thunder star has torn right meniscus, out indefinitely.

Wagner, C., S. Mitter, M. Strohmaier, and C. Korner. When social bots attack: Modeling susceptibility of users in online social networks.

Wagstaff, A. and A. J. Culyer (2012). Four decades of health economics through a bibliometric lens. *Journal of health economics 31*(2), 406–439.

Wallner, L. P., K. A. Martinez, Y. Li, R. Jagsi, N. K. Janz, S. J. Katz, and S. T. Hawley (2016). Use of online communication by patients with newly diagnosed breast cancer during the treatment decision process. *JAMA Oncology 2*(12), 1654–1656.

Wang, G., C. Wilson, X. Zhao, Y. Zhu, M. Mohanlal, H. Zheng, and B. Y. Zhao (2012). Serf and turf: crowdturfing for fun and profit. In *Proceedings of the 21st international conference on World Wide Web*, pp. 679–688. ACM.

Warriner, A. B. and V. Kuperman (2015). Affective biases in english are bi-dimensional. *Cognition and Emotion 29*(7), 1147–1167.

Williams, J. R., J. P. Bagrow, C. M. Danforth, and P. S. Dodds (Accepted 17 March 2015). Text mixing shapes the anatomy of rank-frequency distributions. *Physical Review E (in press)*.

Williams, M., A. Villarreal, K. Bozhilov, S. Lin, and P. Talbot (2013). Metal and silicate particles including nanoparticles are present in electronic cigarette cartomizer fluid and aerosol. *PloS one 8*(3), e57987.

Wills, T. A., R. Knight, R. J. Williams, I. Pagano, and J. D. Sargent (2015). Risk factors for exclusive e-cigarette use and dual e-cigarette use and tobacco use in adolescents. *Pediatrics 135*(1), e43–e51.

Wilson, T., J. Wiebe, and P. Hoffmann (2005). Recognizing contextual polarity in phrase-level sentiment analysis. In *Proceedings of the conference on human language technology and empirical methods in natural language processing*, pp. 347–354. Association for Computational Linguistics.

Wu, X., Z. Feng, W. Fan, J. Gao, and Y. Yu (2013). Detecting marionette microblog users for improved information credibility. In H. Blockeel, K. Kersting, S. Nijssen, and F. Yelezn (Eds.), *Machine Learning and Knowledge Discovery in Databases*, Volume 8190 of *Lecture Notes in Computer Science*, pp. 483–498. Springer Berlin Heidelberg.

Yip, H. and P. Talbot (2013). Mining data on usage of electronic nicotine delivery systems (ends) from youtube videos. *Tobacco Control 22*(2), 103–106.

Zezima, K. (2009). Cigarettes without smoke, or regulation.

Zhang, C. M. and V. Paxson (2011). Detecting and analyzing automated activity on twitter. In *Proceedings of the 12th International Conference on Passive and Active Measurement*, PAM'11, Berlin, Heidelberg, pp. 102–111. Springer-Verlag.

Zhang, Y. and B. Wallace (2015). A sensitivity analysis of (and practitioners' guide to) convolutional neural networks for sentence classification. *arXiv preprint arXiv:1510.03820*.

Zhu, S.-H., A. Gamst, M. Lee, S. Cummins, L. Yin, and L. Zoref (2013). The use and perception of electronic cigarettes and snus among the us population. *PloS one 8*(10), e79332.

Zhu, S.-H., J. Y. Sun, E. Bonnevie, S. E. Cummins, A. Gamst, L. Yin, and M. Lee (2014). Four hundred and sixty brands of e-cigarettes and counting: implications for product regulation. *Tobacco control 23*(suppl 3), iii3–iii9.

Zurilla, C. D. (2014, September). Justin bieber 'busted' eardrum cliff diving, he says, may need surgery.

Appendix A: Sifting Robotic from Organic Text

A.0.1 Cross Sectional Classifier Performance

In Figure A.1 Calibrated Classifier Performance on 1 000 User Geo Tweet Dataset. Correctly classified humans (True Negatives), are coded in Green, while correctly identified automatons (True Positives) are coded in red. The 400 tweet average optimal thresholds from the cross validation experiment designate the thresholding for each feature. The black lines demonstrate each feature cutoff.

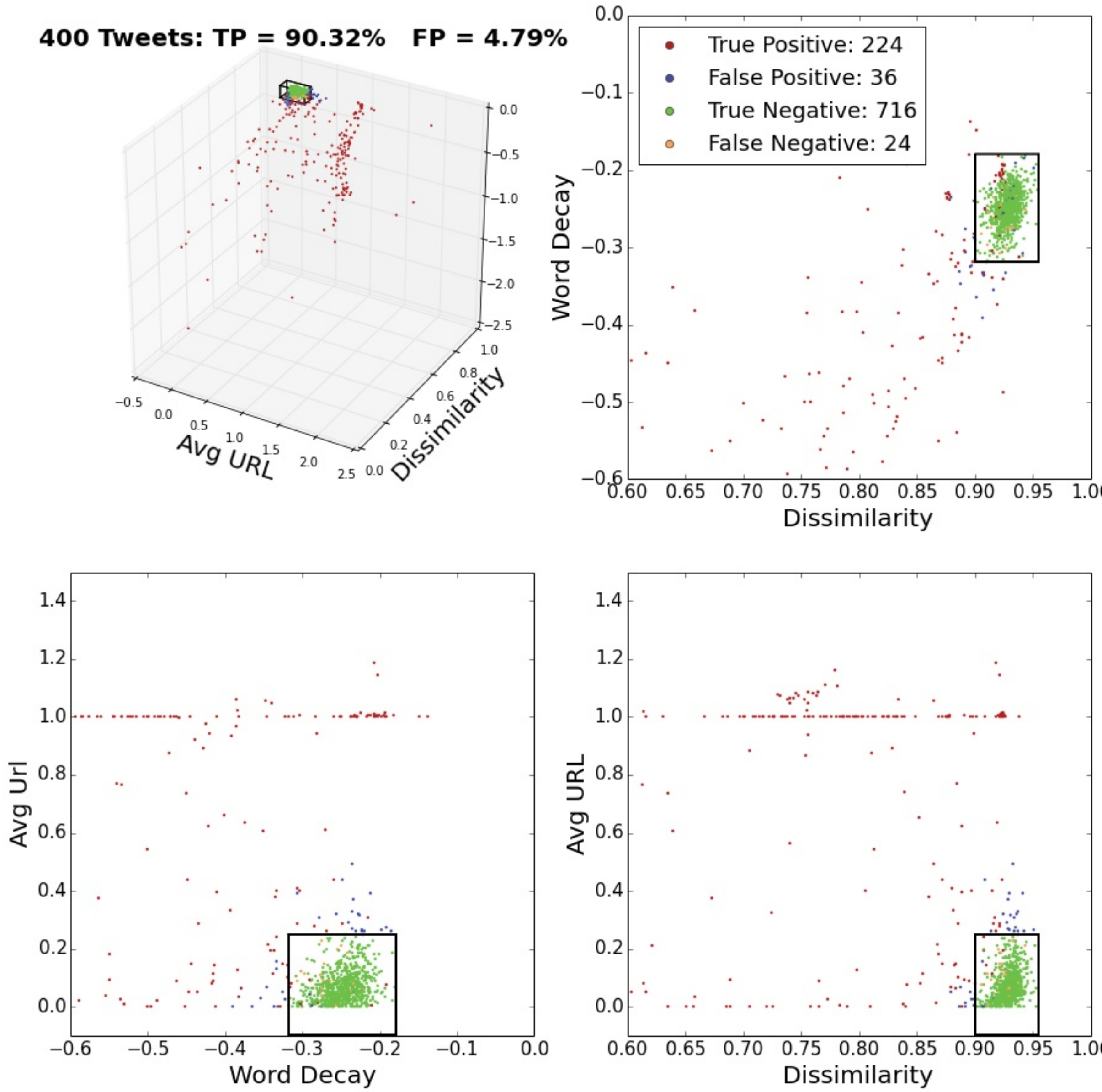


Figure A.1: 1,000 User Geo Tweet Performance

A.0.2 Model Comparisons

Receiver Operator Characteristic (ROC) Curves of the performance of each individual feature are given in Figure A.2 below. We see each feature set performs comparably with accuracies (measured as AUC) ranging from 80% to 91% depending on the number of tweets compiled in the analysis. Combinations of each metric greatly increases the classification accuracy, the apparent most accurate model uses all three features. However, it is notable that combinations of two of these features perform strongly in comparison. It is also notable that the word introduction decay parameter coupled with the average URL rate performs as well as the Dissimilarity-URL model. The Dissimilarity metric requires determining the Longest Common Substring between many sets of tweets which is computationally expensive compared to analytically calculating the Word Introduction Decay Parameter.

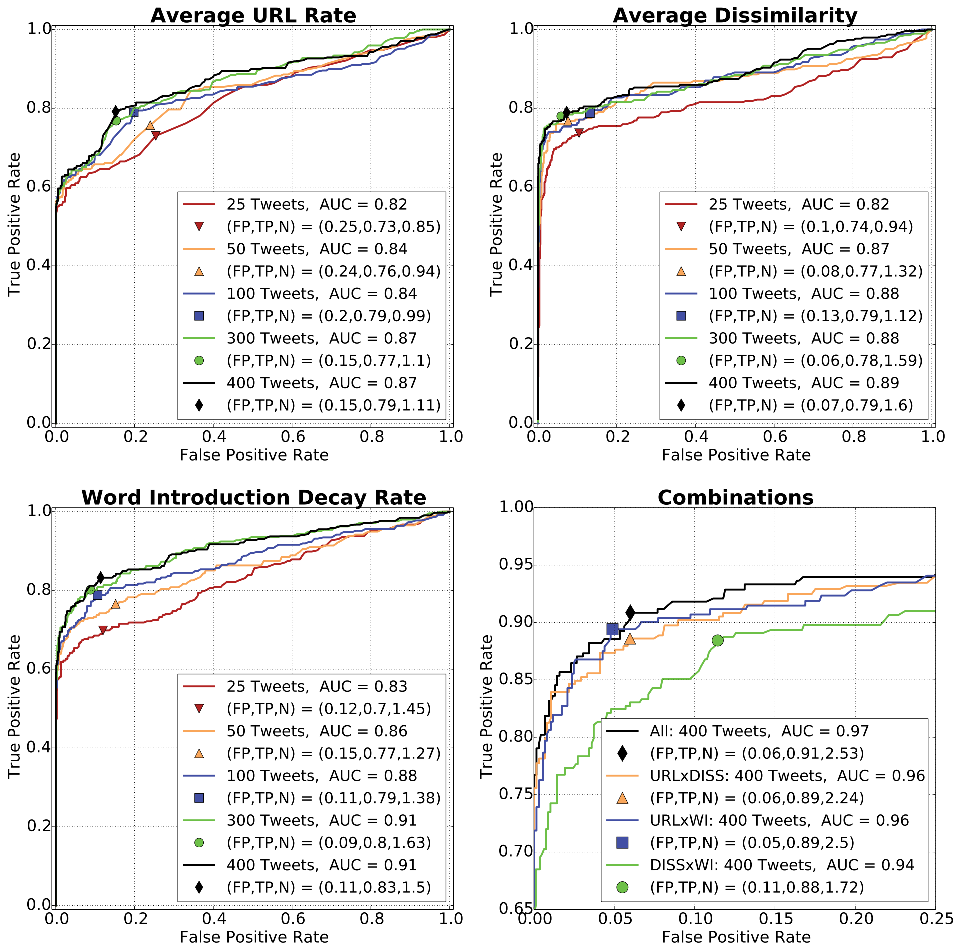


Figure A.2: Feature Receiver Operator Curves

A.0.3 Word Introduction Decay Parameter

Here, we expand upon our description of the Word Introduction Decay Parameter. This parameter is based upon random word shufflings of a text, but is computed via the analytic formula given in Eq 8. of Williams et al. (2015). To determine the decay rate parameter, we: (1) compute the word introduction rate as a function of word number, n, and (2) regress in log-log space for a power law decay rate parameter measuring in the final third for the tail, where the decay rate assumes the form of a power law. While this heuristic is crude and could certainly be refined to more precisely measure the power law region, which can vary with corpus size, the tightness of organic-user clustering afforded by this parameter, coupled with its computationally cheap cost when compared to the pairwise tweet dissimilarity metric affords us great power for bot discrimination.

In Figure A.3 below the unique word introduction gaps are plotted in log-space as a function of unique word introduction number (rank) for each individual in our data set for various numbers of tweets. We see the distribution growing with the number of tweets. However, at each resolution, the human class is very distinctly distributed in comparison to each form of automation. Even within 25 tweets, the human clustering is visually apparent versus their automated counterparts.

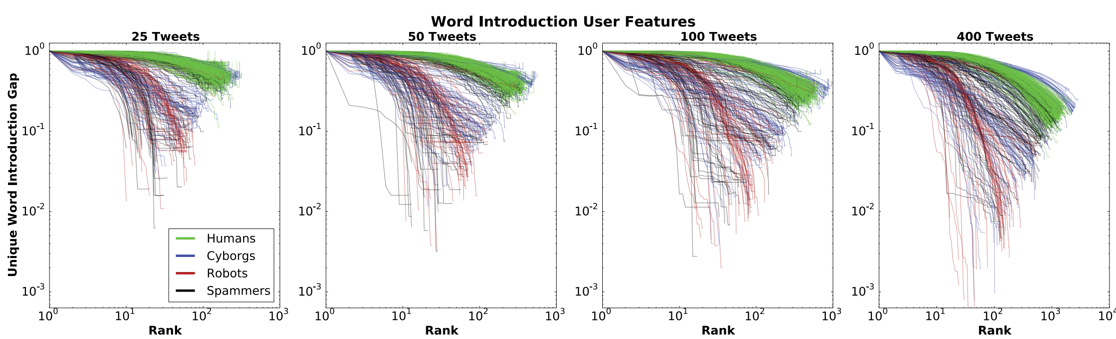


Figure A.3: Unique Word Introduction Gaps per Word Introduction Number (rank)

In Figure A.4 below, we visualize the stability of this parameter between an individual's set of tweets. The tweets from each account in our data set were resampled 100 times to recompute the word introduction decay parameter for 25 tweets (top) and 400 tweets (bottom). The standard deviation between each account's 100 decay parameters is given in the histogram. The average standard deviation across all individuals of each set, μ, is given in the title of each histogram. Notably, the humans have very little deviation (i.e. within the 'window of forgiveness') for both sets of tweets. Automated classes, in particular spammers, can vary quite wildly depending on the sample of tweets that are analyzed. In particular, spammers look similar to (and usually are) humans and if the spamming event is not captured in the sampled data they will be misclassified. This decay parameter for human text is robust for varying sets of tweets and is quite distinguishable from automated accounts.

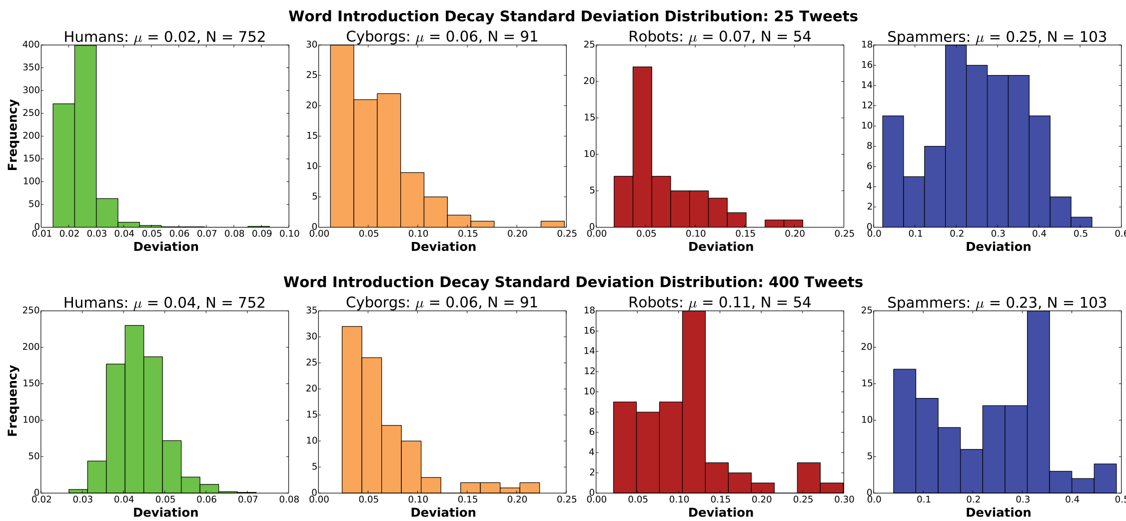


Figure A.4: Word Decay Deviations

A.0.4 Human Annotation of Twitter Accounts

In this section we describe the annotation process for classifying user accounts. Each of the one-thousand accounts were separately classified by two evaluators. All of the collected tweets from each account were assessed until the presence of automation was uncovered. An account was coded as ‘human’ if no automated posting presence was detected (i.e. an algorithm posting on the individual’s behalf). The inter-rater reliability is summarized in the table below by listing the classification discrepancies between account classes. For each class, the counts of the type of rating are displayed. For example, the Human class had a total of 6 account discrepancies which is composed of 12 different scores (6 from each rater) - 5 human codings, 3 robot, 0 cyborg, and 4 spammers.

The reliability between raters was favorable (91.49%) for the entire dataset. The largest source of discrepancies were from the ‘Robot’ and ‘Spammer’ classes. The ’Spammer’ class was most confused with ’Humans’- which is intuitive because many of these individuals were humans that utilized a algorithm to SPAM a particular message. ‘Robots’ were commonly confused with ’cyborgs’. This is most likely due to boundary cases regarding both classes. The boundaries between these classes can at time be ambiguous. We classified cyborgs as automatons that were posting ‘borrowed content‘ from another source or an account that used human assisted automation, i.e. a human that could be overseeing an automated account. Robots were defined as strictly posting structured automated messages in the forms of updates. Perhaps future work can work to sub classify different types of robots and cyborgs to investigate the ecology of these automatons.

Table A.1: Annotation Discrepancies of Twitter Accounts

Class	**Discrepancies**				**Totals**
	Human	Robot	Cyborg	Spammer	
Human	5	3	0	4	0.79%
Cyborg	2	9	6	1	9.89%
Robot	0	31	32	1	60.38%
Spammer	31	4	4	37	36.89%
All	6	9	32	38	8.51%

Each discrepancy was revisited by both annotators and discussed until a class was determined. For extreme boundary cases, the account ID was searched via the hyperlink:
`https://twitter.com/intent/user?user_id=#####`. This helped observe other user features

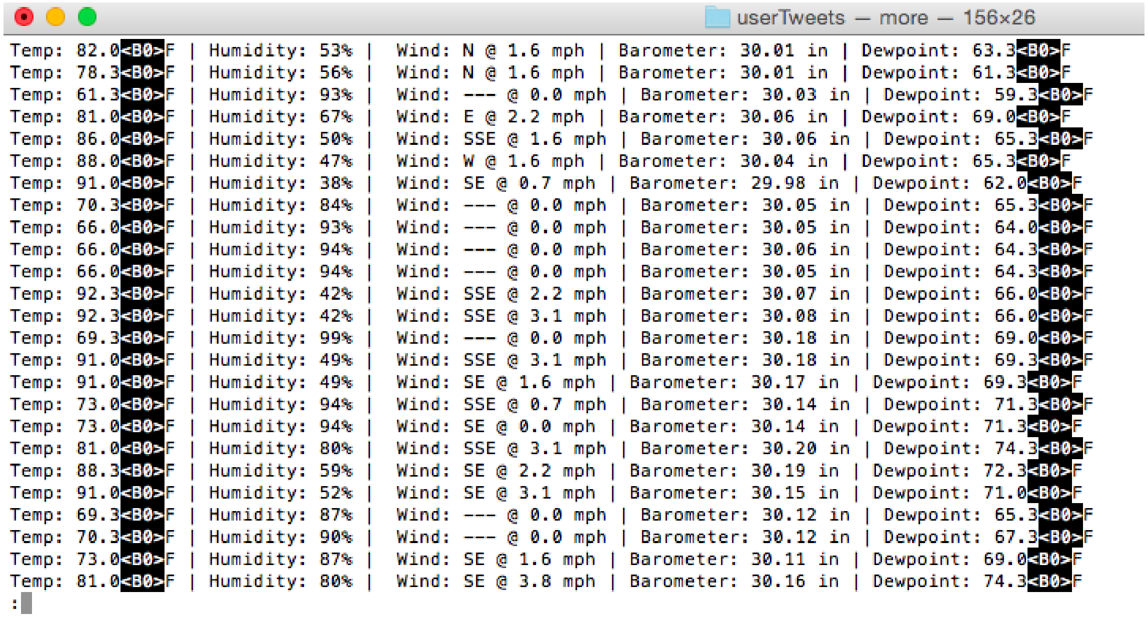
(screen name, description, etc.) to make a better decision about the user. This was especially helpful for identifying promotional accounts or news sources.

Screen shots of particular accounts are given below to help describe the annotation process. Each annotator scrolled through a terminal interface containing each individual's tweets. Scrolling through 'human' text appears un-ordered and chaotic with very little structure. Automated accounts have very structured messages, hence these patterns become very apparent in comparison to human accounts.

Cyborg Account Example: A canonical cyborg's tweets are given below. This particular automaton is a news promotional account that is tweeting links to articles. Notice the description tailors off when it reaches the character limit and shows this with an ellipses (...) next to a URL.

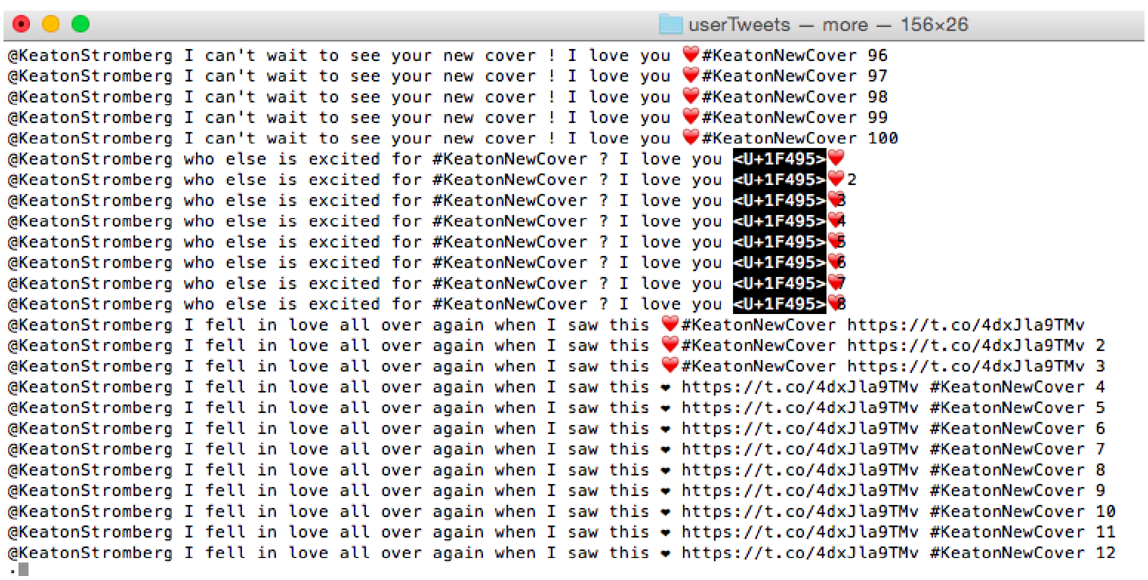
```
userTweets — more — 156×26
Miley's Meadowlands show is surreal, lacks substance: review: Say what you will about Miley Cyrus, but she knows... http://t.co/CCQI3P7LLO
Bushwick warehouse owner invites tenants to 'join the gentrification': Yeah, it's gentrification – get over it.... http://t.co/axMKzW5TYo
Yankees Insider: No Mo? No problem for Diamond Dave: In his first opportunity as the full-time closer for the... http://t.co/VcrHPFOiyp
Phew! Yanks avoid sweep as Ichiro and Solarte power Bombers to first victory: Breathe easy, Yankees fans. The... http://t.co/k3Qv30yD21
'If you don't like it, sit your a-- on a couch and eat a donut': CrossFit mom whose pregnant weightlifting photos... http://t.co/PHz83m5YgN
Bronx Urban Cowboy back in the saddle: HE'S A real urban cowboy. Bronx-born rodeo star AJ Silver has lassoed and... http://t.co/93jAT4QnFd
Wake-Up Call: De Blasio and Cuomo spat over speed enforcement cameras: Good morning and welcome to Wake-Up Call,... http://t.co/UwQqQqmQmA
Cabbies' $1.4M health care fund remains frozen by city lawyers: A long-sought health care and disability fund for... http://t.co/yT4TAeqPhC
Judge tosses out Brooklyn gun seizure over cop's false testimony: The decision – which will likely result in the... http://t.co/6yIctqdXzE
From Winter's Bite to Spring's: New Yorkers took advantage of the warmer weather in Hudson River Park in New York. http://t.co/Vphj1R7IpW
New York Today: Floating Your Boat: What you need to know for Friday: the city's boat launches open, dismal... http://t.co/WBg3Pz6TRi
Michael Schumacher 'shows moments of consciousness and awakening' three months after falling into coma following... http://t.co/BUVxRDMdu8
Sen. Charles Schumer could help to save Cit Bike: A commuter benefit introduced by Sen. Charles Schumer would... http://t.co/z23ltjyKHA
B'klyn Housing Court can't find a home: Brooklyn Housing Court is looking for a new home. The court has outgrown... http://t.co/ax1lFtlzP4
It's Still Bad for the Long-Term Unemployed: The evidence remains that the so-so recovery is not enough to help... http://t.co/AX6cUmfoCK
Most Details in Jobs Report Are Positive: The trend toward a rising labor force in 2014 suggests that America is... http://t.co/wI1Tw9eMGN
:BUBBA BEWARE! Bill Clinton asked aides to investigate aliens while in charge of the free world – 'We're not alone... http://t.co/Tetxao8AjM
Theater Review: 'Raisin in the Sun' Brings Denzel Washington Back to Broadway: Starring Denzel Washington, "A... http://t.co/whyDHcaebV
Panel looks to shelve big-spending library director: The Queens Library Board retreated into a private meeting... http://t.co/bRNhbmqFsP
Derek Jeter avoids pesky adult fan to give souvenir ball to young girl: When Derek Jeter tells you to get out of... http://t.co/89msl7GqzF
Raissman: Time to lower Boomer on pompous Pope!: For sports radio Gasbags, unintended consequences are a good... http://t.co/fh9xxJuF2K
Phew! Yanks avoid sweep as Ichiro and Solarte power Bombers to first victory: Breathe easy, Yankees fans. The... http://t.co/gTMNbIAlhD
Groups spend $210M lobbying state, local governments: Firms spend big bucks petitioning their causes to state and... http://t.co/DdXVsWxaaN
New mom Newser weighs in on WFAN baby flap: On his WFAN radio show, Esiason famously mocked Daniel Murphy's... http://t.co/T8tKpQyBx9
Lupica: After Fort Hood, time to invest in the mental health of our vets: The first reports out of Fort Hood... http://t.co/M7S6eeYUN0
Gov. Cuomo wants to reform Port Authority: Gov. Cuomo wants to reform the scandal-tainted Port Authority of New... http://t.co/5Lid7sTHlQ
```

Robot Account Example: Robots tweet generically structured messages, usually as a form of update. These automatons have a very limited vocabulary and in general only change a few characters per tweet. This robot (below) is an example of a weather update bot that is tweeting statistics about the weather at regular intervals.



```
userTweets — more — 156×26
Temp: 82.0<B0>F | Humidity: 53% |  Wind: N @ 1.6 mph | Barometer: 30.01 in | Dewpoint: 63.3<B0>F
Temp: 78.3<B0>F | Humidity: 56% |  Wind: N @ 1.6 mph | Barometer: 30.01 in | Dewpoint: 61.3<B0>F
Temp: 61.3<B0>F | Humidity: 93% |  Wind: --- @ 0.0 mph | Barometer: 30.03 in | Dewpoint: 59.3<B0>F
Temp: 81.0<B0>F | Humidity: 67% |  Wind: E @ 2.2 mph | Barometer: 30.06 in | Dewpoint: 69.0<B0>F
Temp: 86.0<B0>F | Humidity: 50% |  Wind: SSE @ 1.6 mph | Barometer: 30.06 in | Dewpoint: 65.3<B0>F
Temp: 88.0<B0>F | Humidity: 47% |  Wind: W @ 1.6 mph | Barometer: 30.04 in | Dewpoint: 65.3<B0>F
Temp: 91.0<B0>F | Humidity: 38% |  Wind: SE @ 0.7 mph | Barometer: 29.98 in | Dewpoint: 62.0<B0>F
Temp: 70.3<B0>F | Humidity: 84% |  Wind: --- @ 0.0 mph | Barometer: 30.05 in | Dewpoint: 65.3<B0>F
Temp: 66.0<B0>F | Humidity: 93% |  Wind: --- @ 0.0 mph | Barometer: 30.05 in | Dewpoint: 64.0<B0>F
Temp: 66.0<B0>F | Humidity: 94% |  Wind: --- @ 0.0 mph | Barometer: 30.06 in | Dewpoint: 64.3<B0>F
Temp: 66.0<B0>F | Humidity: 94% |  Wind: --- @ 0.0 mph | Barometer: 30.05 in | Dewpoint: 64.3<B0>F
Temp: 92.3<B0>F | Humidity: 42% |  Wind: SSE @ 2.2 mph | Barometer: 30.07 in | Dewpoint: 66.0<B0>F
Temp: 92.3<B0>F | Humidity: 42% |  Wind: SSE @ 3.1 mph | Barometer: 30.08 in | Dewpoint: 66.0<B0>F
Temp: 69.3<B0>F | Humidity: 99% |  Wind: --- @ 0.0 mph | Barometer: 30.18 in | Dewpoint: 69.0<B0>F
Temp: 91.0<B0>F | Humidity: 49% |  Wind: SSE @ 3.1 mph | Barometer: 30.18 in | Dewpoint: 69.3<B0>F
Temp: 91.0<B0>F | Humidity: 49% |  Wind: SE @ 1.6 mph | Barometer: 30.17 in | Dewpoint: 69.3<B0>F
Temp: 73.0<B0>F | Humidity: 94% |  Wind: SSE @ 0.7 mph | Barometer: 30.14 in | Dewpoint: 71.3<B0>F
Temp: 73.0<B0>F | Humidity: 94% |  Wind: SE @ 0.0 mph | Barometer: 30.14 in | Dewpoint: 71.3<B0>F
Temp: 81.0<B0>F | Humidity: 80% |  Wind: SSE @ 3.1 mph | Barometer: 30.20 in | Dewpoint: 74.3<B0>F
Temp: 88.3<B0>F | Humidity: 59% |  Wind: SE @ 2.2 mph | Barometer: 30.19 in | Dewpoint: 72.3<B0>F
Temp: 91.0<B0>F | Humidity: 52% |  Wind: SE @ 3.1 mph | Barometer: 30.15 in | Dewpoint: 71.0<B0>F
Temp: 69.3<B0>F | Humidity: 87% |  Wind: --- @ 0.0 mph | Barometer: 30.12 in | Dewpoint: 65.3<B0>F
Temp: 70.3<B0>F | Humidity: 90% |  Wind: --- @ 0.0 mph | Barometer: 30.12 in | Dewpoint: 67.3<B0>F
Temp: 73.0<B0>F | Humidity: 87% |  Wind: SE @ 1.6 mph | Barometer: 30.11 in | Dewpoint: 69.0<B0>F
Temp: 81.0<B0>F | Humidity: 80% |  Wind: SE @ 3.8 mph | Barometer: 30.16 in | Dewpoint: 74.3<B0>F
:
```

Spammer Account Example: Tweets from a spamming human account are given below. This individual has utilized an algorithm to tweet at a musical celebrity. Many of these spam algorithms try to fool Twitter's detector by including a different number or symbol at the end of the tweet.



```
userTweets — more — 156×26
@KeatonStromberg I can't wait to see your new cover ! I love you ❤#KeatonNewCover 96
@KeatonStromberg I can't wait to see your new cover ! I love you ❤#KeatonNewCover 97
@KeatonStromberg I can't wait to see your new cover ! I love you ❤#KeatonNewCover 98
@KeatonStromberg I can't wait to see your new cover ! I love you ❤#KeatonNewCover 99
@KeatonStromberg I can't wait to see your new cover ! I love you ❤#KeatonNewCover 100
@KeatonStromberg who else is excited for #KeatonNewCover ? I love you <U+1F495>❤
@KeatonStromberg who else is excited for #KeatonNewCover ? I love you <U+1F495>❤2
@KeatonStromberg who else is excited for #KeatonNewCover ? I love you <U+1F495>❤3
@KeatonStromberg who else is excited for #KeatonNewCover ? I love you <U+1F495>❤4
@KeatonStromberg who else is excited for #KeatonNewCover ? I love you <U+1F495>❤5
@KeatonStromberg who else is excited for #KeatonNewCover ? I love you <U+1F495>❤6
@KeatonStromberg who else is excited for #KeatonNewCover ? I love you <U+1F495>❤7
@KeatonStromberg who else is excited for #KeatonNewCover ? I love you <U+1F495>❤8
@KeatonStromberg I fell in love all over again when I saw this ❤#KeatonNewCover https://t.co/4dxJla9TMv
@KeatonStromberg I fell in love all over again when I saw this ❤#KeatonNewCover https://t.co/4dxJla9TMv 2
@KeatonStromberg I fell in love all over again when I saw this ❤#KeatonNewCover https://t.co/4dxJla9TMv 3
@KeatonStromberg I fell in love all over again when I saw this ❤ https://t.co/4dxJla9TMv #KeatonNewCover 4
@KeatonStromberg I fell in love all over again when I saw this ❤ https://t.co/4dxJla9TMv #KeatonNewCover 5
@KeatonStromberg I fell in love all over again when I saw this ❤ https://t.co/4dxJla9TMv #KeatonNewCover 6
@KeatonStromberg I fell in love all over again when I saw this ❤ https://t.co/4dxJla9TMv #KeatonNewCover 7
@KeatonStromberg I fell in love all over again when I saw this ❤ https://t.co/4dxJla9TMv #KeatonNewCover 8
@KeatonStromberg I fell in love all over again when I saw this ❤ https://t.co/4dxJla9TMv #KeatonNewCover 9
@KeatonStromberg I fell in love all over again when I saw this ❤ https://t.co/4dxJla9TMv #KeatonNewCover 10
@KeatonStromberg I fell in love all over again when I saw this ❤ https://t.co/4dxJla9TMv #KeatonNewCover 11
@KeatonStromberg I fell in love all over again when I saw this ❤ https://t.co/4dxJla9TMv #KeatonNewCover 12
:
```

A.0.5 Mixed User Sample Tweets

Table A.2: **Robot Sample Tweets**

No.	Tweet
1	#SuspiciousPerson 3955 W D JUDGE DR 32808 (7/26 22:19) #Orlando #MercyDrive
2	#AccidentWithRoadBlockage LEEVISTA BV & S SEMORAN BV N/A (6/24 18:07) #Orlando #CentralBusinessDistrict
3	@USER the 1st mention of #sunnysmiles appears on your TL. Now is Trending Topic in United States! #trndnl
4	TRAFFIC STOP at SE 181ST AVE / SE PINE ST, GRESHAM, OR [Gresham Police #PG14000039852] 23:58 #pdx911
5	A 2002 Ford Ranger was just scanned near Cleveland, TN 37311 URL #myvinny #startup #buyacar
6	Visiting #SantaCruz, #California? Check out this great new app for news, weather, hotels, and food here! URL
7	Trend Alert: #heatJustinBieber. More trends at URL #trndnl URL
8	Temp: 76.5F — Humidity: 70% — Wind: — @USER 0.0 mph — Barometer: 30.01 in — Dewpoint: 66.0F
9	On Sunday 4, #WinUgly was Trending Topic in Pittsburgh for 10 hours: URL #trndnl
10	Wind 0.0 mph —. Barometer 1016.0 mb, Rising slowly. Temperature 67.6 F. Rain today 0.46 in. Humidity 96%

Table A.3: **Cyborg Sample Tweets**

No.	Tweet
1	Indianapolis, IN suburb Family practice physi... - Soliant Health: (#Indianapolis, IN) URL #FamilyPractice #Job
2	Barnabas Health: Patient Care Associate (#LongBranch, NJ) URL #Nursing #Job #Jobs #TweetMyJobs
3	Overlake offers a low-cost way to check lungs for cancer early: Doctors at Overlake say they?re tired of waiting... URL
4	#TweetMyJobs #Nursing #Job alert: Opening — Accountable Healthcare Staffing — #Glendale, AZ URL #Jobs
5	Soliant Health #IT #Job: Cerner Jobs - Cerner Analyst - San Diego, CA (#SanDiego , CA) URL #Jobs #TweetMyJobs
6	Tyco #Marketing #Job: Digital Marketing Specialist (#Monroe , NC) URL #Jobs #TweetMyJobs
7	@USER Timing is everything when announcing a breakup URL
8	Southwest flights briefly diverted to DFW Airport on Friday: A Southwest Airlines plane experiencing mechanical... URL
9	Fort Carson To Welcome Home About 225 Soldiers: FORT CARSON, Colo. (AP) ? Fort Carson will welcome home about 225... URL
10	Joint venture secures $97M in financing for two Boston hotels: Commonwealth Ventures and Ares Management have... URL

Table A.4: **Spammer Sample Tweets**

No.	Tweet
1	#CallMeCam n#CallMeCam n @USER n nIf Cameron called me it'll seriously make my day I love you please call me! 100
2	#CallMeCam n#CallMeCam n @USER n nIf Cameron called me it'll seriously make my day I love you please call me! 321
3	#CallMeCam n#CallMeCam n @USER n nIf Cameron called me it'll seriously make my day I love you please call me! 167
4	S/o to @USER thanks for the support. Check out my music @USER URL I promise u won't be disappointed.
5	S/o to @USER destiiny thanks for the support. Check out my music @USER URL I promise u won't be disappointed.
6	S/o to @USER thanks for the support. Check out my music @USER URL I promise u won't be disappointed.
7	nAshton Irwin from 5SOS n nMy birthday is in 11 days, nAnd it would be an amazing gift, nIf you could follow me. Ily n @USER n nX3126
8	nAshton Irwin from 5SOS n nMy birthday is today, nAnd it would be an amazing gift, nIf you could follow me. Ily n @USER n nX5408
9	nAshton Irwin from 5SOS n nMy birthday is in 22 days, nAnd it would be an amazing gift, nIf you could follow me. Ily n @USER n nX765
10	nAshton Irwin from 5SOS n nMy birthday is in 8 days, nAnd it would be an amazing gift, nIf you could follow me. Ily n @USER n nX3422

Table A.5: **Human Sample Tweets**

No.	Tweet
1	I'll marry whoever comes thru with some food
2	Ewwww them seats #BlackInkCrew
3	@USER if you only knew ???? i like you
4	Really wish he wasn't so **** busy ??
5	My son's name is Gabriel.
6	guess I need to get up and get ready then
7	Grandma stayed on me bout not wearing socks in her house aint nobody got time for that
8	Thank you for reading. ??
9	@USER: If only I knew.. ???
10	WHY ARE YOU SO HOT URL

Appendix B: Vaporous marketing

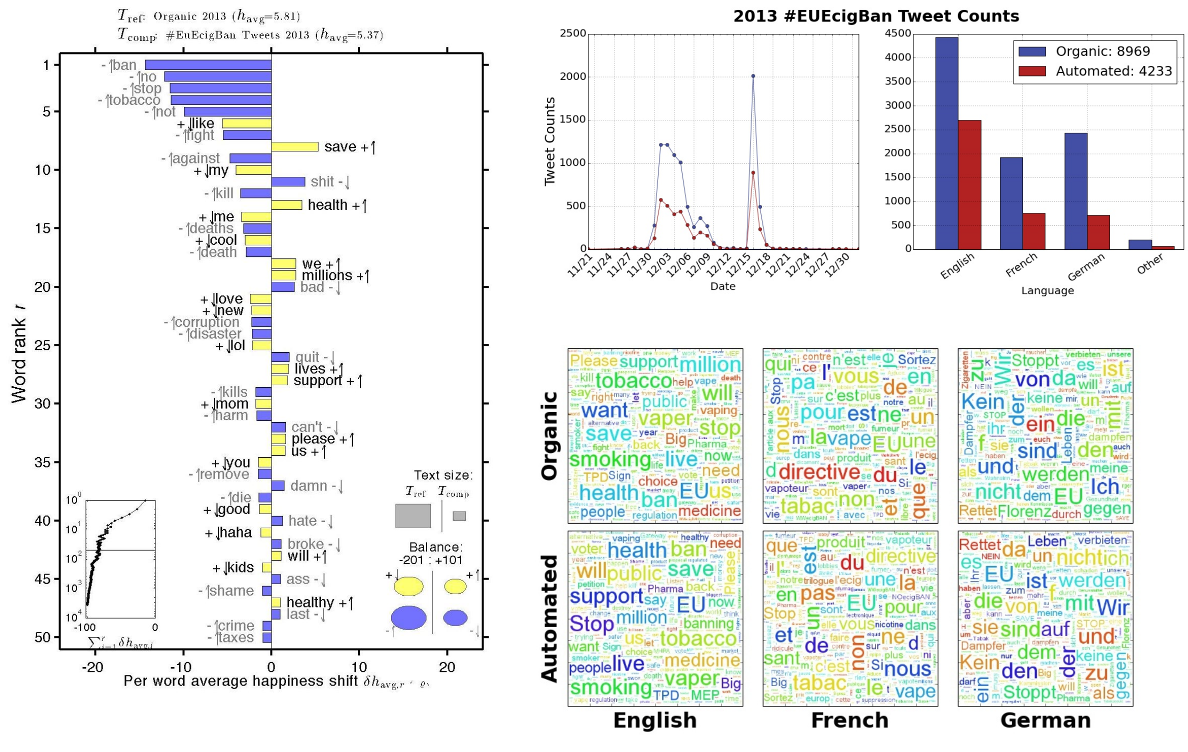


Figure B.1: **European Union E-cigarette Ban Political Debate (#EUecigBan)** (Left) Word shift graph comparing tweets tagged #EUecigBan against 2013 English Organic User Tweets (untagged). (top-right) The automated and Organic tagged tweet distributions are plotted. A histogram displays the counts per language and user class. (bottom-right) Word clouds compare ranked-word frequencies across language and user type. Each categorical time-series exhibits a severe negative trend occurring between December 2013 and January 2014. There is an inverse relationship with the average happiness scores during this time period. This was during the time that the EU was debating strict regulation and a possible ban on specific e-cigarette products. Hashtags (#) allow users to categorize the content of their tweets. During this period, 13,227 sampled tweets were tagged with #EUecigBan. In S1 Fig, a word shift graph (left) visualizes the sentiments from English Organic users using #EUecigBan versus the remaining Organic tweets from 2013. English Tweets tagged #EuEcigBan are the comparison distribution in reference to all other tweets from 2013. Tweets containing #EuEcigBan are on average much more negative (h_{avg} 5.81 versus 5.37) due to an increase in the negative words 'ban', 'stop', 'no', 'not', 'fight', 'against', 'disaster', 'death', 'corruption', 'tobacco', 'kills', etc. The positive words also disfavor the legislation, with the words 'save', 'millions', 'lives', 'support', 'healthy' occurring more frequently. English, French, and German tagged tweets were the most prevalent, and word clouds help visualize themes between language and user class. This shows that Twitter sentiments can be useful in gauging public opinion toward regulation of electronic cigarettes. There is also a heavy automated tweet presence in each language with a similar attitude regarding the legislation, as depicted in the word clouds. Future work should also investigate if and how automated users can impact organic opinion on legislation.

APPENDIX B. VAPOROUS MARKETING

Electronic Cigarette Table of Key Words List of all key words used in the analysis. Flavors compiled from `https://crazyvapors.com/e-liquid-flavor-list/`Keywords other than 'General Twitter Scrape' were applied to categorize automated account tweets.

Table B.1: **Electronic Cigarette Table of Key Words**

Type	Keywords
General Twitter Scrape (includes hashtag variants)	ecig, e cig, e-cig, ecigs, e cigs, e-cigs, e ciggs, e ciggs, e-ciggs, eciggs, e cigg, ecigg, e-cigarette e cigarette, e cigarettes, e-cigarettes, electronic cigarette blucigs, blucig, blu cig, blu cigs, blu ciggs, electronic cigarettes
Commercial	buy, save, coupon, coupons, discount, price, cost, deal, promo, money, sale ,purchase, offer, review, code ,win, winner, starter kit, starter kits, premium, $, kit, %, sales,voucher, brand, free e cigarette, free electronic cigarette, free e cig, free ecig
Cessation	quit, quitting, quits, stop smoking, smoke free, quitter, safe, safest, safer, quitsmoking, give up smoking
Discount	free trial, free shipping, free sample ,free samples, coupon, discount, discounts, save, sale, coupons, deal, deals, free e cigarette, free electronic cigarette, free e cig, free ecig
Flavors*	flavor, flavour, flavors, flavours, flavored, flavoured Cherry, Lime, Almond Coconut Bar, Alpine Fresh, Amaretto, Apple Pie (Ala Mode), Banana, Banana Cream, Banana Graham, Banana Nut Bread ,Banana Pudding, Banana Split, Bavarian Cream, Belgian Waffle Berry Blast, Black Cherry, Black Berry, Black Honey, Blazing Frost, Blueberry,Blueberry Cheesecake, Blueberry Cinnamon Crumble, Blueberry Cotton Candy Blueberry Delight,Brandy, Bubble Gum, Butterscotch Butter Rum, Buttered Popcorn, Cafe Latte, Cake Batter, Candy Cane, Candy Apple, Cantaloupe, Caramel Caramel Cappuccino, Cappuccino,Champagne, Cheesecake, Chocolate Covered Raspberries

Type	Keywords
Flavors* (continued)	Cinnamon Coffee Cake, Cinnamon Danish,
	Cinnamon Sugar Cookie, Circus Cotton Candy
	Clove, Coconut, Coconut Candy, Coffee
	Coffee&Cream, Cola, Cool, Cotton Candy
	Cranberry, Crazy Berry, Crazy Chill, Crazy Dew
	Crazy Freeze, Crazy Grass, Crazy Hump
	Crazy Pep, Crazy Rainbow, Crazy Watermelon
	Cream Cheese Frosting, Cream de Menthe
	Creamy Fruit Smoothie, Cuban Cigar
	Cured TobaccoDaquiri, DK-Tab, Double Chocolate
	Dragon’s Blood, Dragon Fruit, Dulce De Leche
	Egg Nog, English Toffee, Espresso, Extreme Ice
	Flaming Peach, French Toast, French Vanilla,
	French Vanilla Deluxe, Fresh Apple, Fresh-N-Fruity
	Fudge Brownie, Fruit Rocket, Georgia Peach, Gingerbread
	Goblin Goo, Golden Pineapple, Graham Cracker, Green Apple
	Green Tea, Gummy Candy, Harvest Berry, Hazelnut,
	Hot Chocolate, Hot Cinnamon Candy, Hypnotic, Irish Cream,
	Island Getaway, Jamaican Rum, Java Shake, Jungle Juice,
	Kentucky Bourbon, Kettle Corn, Khaluah & Cream, Kiwi,
	Lemon Drop, Lemon Lime, Lemon Meringue Pie, Mango,
	Marshmallow, Melon, Menthol, Mint Patty, Milk Chocolate,
	M-Mix Menthol, M-Mix Special Blend, Mocha, Mojito, Mummy Mint
	Munster, N-Mix, N-Mix Menthol, NY Cheesecake,
	Orange Creamsicle, P-Mix, P-Mix Menthol, Papaya
	Passion Fruit, Peanut Butter, Peanut Buttercup,
	Honey Dew Melon, Margarita, M-Mix, Orange Cognac

Appendix C: Breast Cancer Treatment Experiences and Healthcare Perceptions

C.0.1 Raw ‘Cancer’ Twitter Data Overview

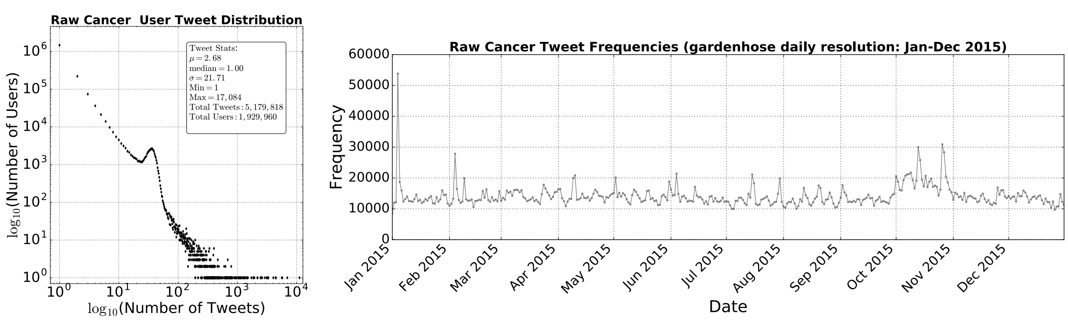


Figure C.1: A Frequency time-series of raw ‘cancer’ tweets collected, binned by day. This sample was compiled from a 10% random sample of Twitter, the ‘Gardenhose’ feed. (left) The distribution of tweets per given user is plotted on a log axis. The tail tends to be high frequency automated accounts, some of which provide daily updates on horoscope information, or about news related to cancer. The kink in the center is also abnormal and could be representative of other classes of automation. This shows the necessity to sift irrelevant tweets using combinations of keyword removal and content classifiers.

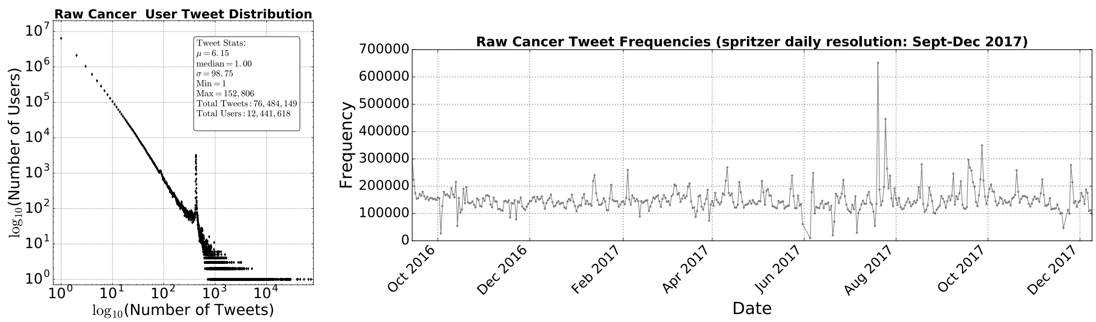


Figure C.2: Another frequency time-series of raw 'cancer' tweets collected, binned by day. This sample was compiled from a 1% random sample of Twitter, the 'Spritzer' feed concentrated on keyword 'cancer', during the same time interval as Figure 4.1. We collected over 76 million tweets, which accounted for approximately 65.2% of all tweets mentioning 'cancer' while the data stream was active (i.e., not accounting for power/network outages). The kink that was visible in the previous figure seems to moved outward by almost a factor of 10, since this is a much larger sample of 'cancer' tweets (10% versus $\approx$ 65%). This serves as a comparison to the tweets collected using keywords 'breast' and 'cancer' and to raw 'cancer' tweets collected from the Gardenhose feed.

C.0.2 Calculating the Tweet Sampling Proportion

There are three types of endpoints to access data from Twitter. The ‘spritzer’ (1%) and ‘gardenhose’ (10%) endpoints were both implemented to collect publicly posted relevant data for our analysis. The third type of endpoint is the ‘Firehose’ feed, a full 100% sample, which can be purchased via subscription from Twitter. This was unnecessary for our analysis, since our set of keywords yielded a high proportion of the true tweet sample. We quantified the sampled proportion of tweets using overflow statistics provided by Twitter. These ‘limit tweets’, L, issue a timestamp along with the approximate number of posts withheld from our collected sample, T_s. The sampling percentage, $\tilde{\rho}_s$, of keyword tweets is approximated as the collected tweet total, $|T_s|$, as a proportion of itself combined with the sum of the limit counts, each $\ell \in L$:

$$\tilde{\rho}_s = \frac{|T_s|}{|T_s| + \sum_{\ell \in L} \ell} = \frac{\text{total collected tweets}}{\text{total collected tweets + overflow limit sum}} \approx \text{sampling proportion} \tag{C.1}$$

By the end of 2017, Twitter was accumulating an average of 500 million tweets per day, InternetLiveStats (2017). Our topics were relatively specific, which allowed us to collect a large sample of tweets. For the singular search term, ‘cancer’, the keyword sampled proportion, $\tilde{\rho}_s$, was approximately 65.21% with a sample of 89.2 million tweets. Our separate Twitter spritzer feed searching for keywords ‘breast AND cancer‘ OR ‘lymphedema’ rarely surpassed the 1% limit. We calculated a 96.1% sampling proportion while our stream was active (i.e. not accounting for network or power outages). We present the daily overflow limit counts of tweets not appearing in our data-set, and the approximation of the sampling size in Fig SI 2.

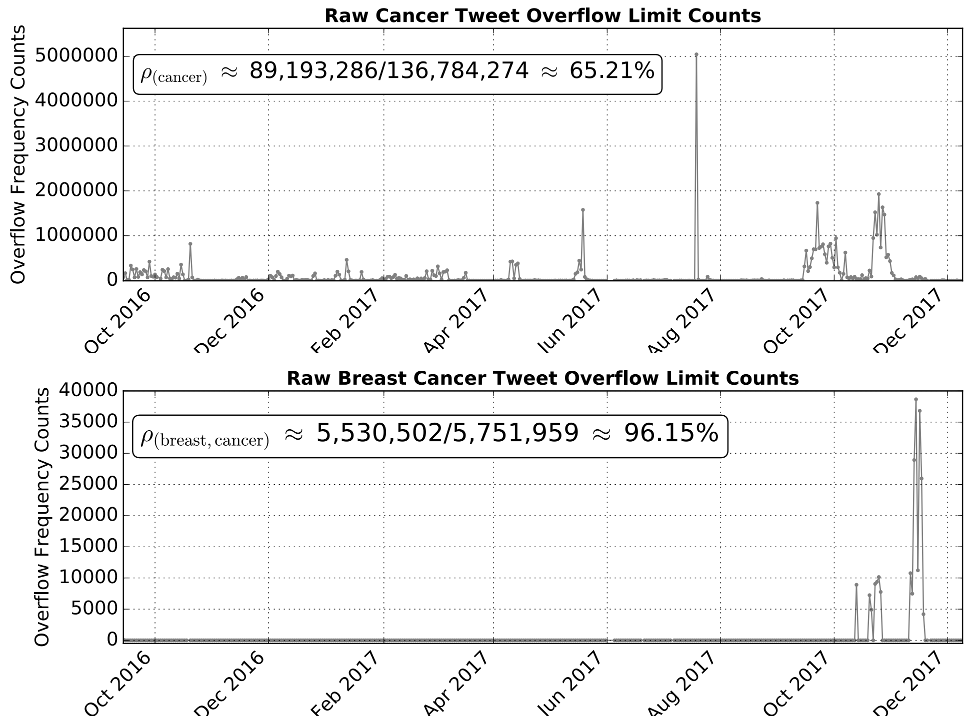


Figure C.3: Overflow limit statistics, plotted per day for both the cancer and breast cancer Twitter feeds with the corresponding approximation of the sampling proportion over the study time frame.

C.0.3 Interpreting Word Shift Graphs

Word shift graphs are essential tools for analyzing which terms are affecting the computed average happiness scores between two text distributions, Reagan et al. (2015). The reference word distribution, T_{ref}, serves as a lingual basis to compare with another text, T_{comp}. The top 50 words causing the shift in computed word happiness are displayed along with their relative weight. The arrows ($\uparrow, \downarrow$) next to each word mark an increase or decrease in the word's frequency. The $+$,$-$, symbols indicate whether the word contributes positively or negatively to the shift in computed average word happiness.

In Fig SI 3, word shift graphs compare tweets mentioning ‘breast’ ‘cancer’ and a random 10% ‘Gardenhose’ sample of non filtered tweets. On the left, ‘breast’,‘cancer’ tweets were slightly less positive due to an increase in negative words like ‘fight’, ‘battle’, ‘risk’, and ‘lost’. These distributions had similar average happiness scores, which was in part due to the relatively more positive words ‘women’, mom’, ‘raise’, ‘awareness’, ‘save’, ‘support’, and ‘survivor’. The word shift on the right compares breast cancer patient tweets to non filtered tweets. These were more negative (h_{avg} = 5.78 v. 6.01) due a relative increase in words like ‘fighting’, ‘surgery’, ‘against’, ‘dying’, ‘sick’, ‘killing’, ‘radiation’, and ‘hospital’. This tool helped identify words that signal emotional themes and allow us to extract content from large corpora, and identify thematic emotional topics within the data.

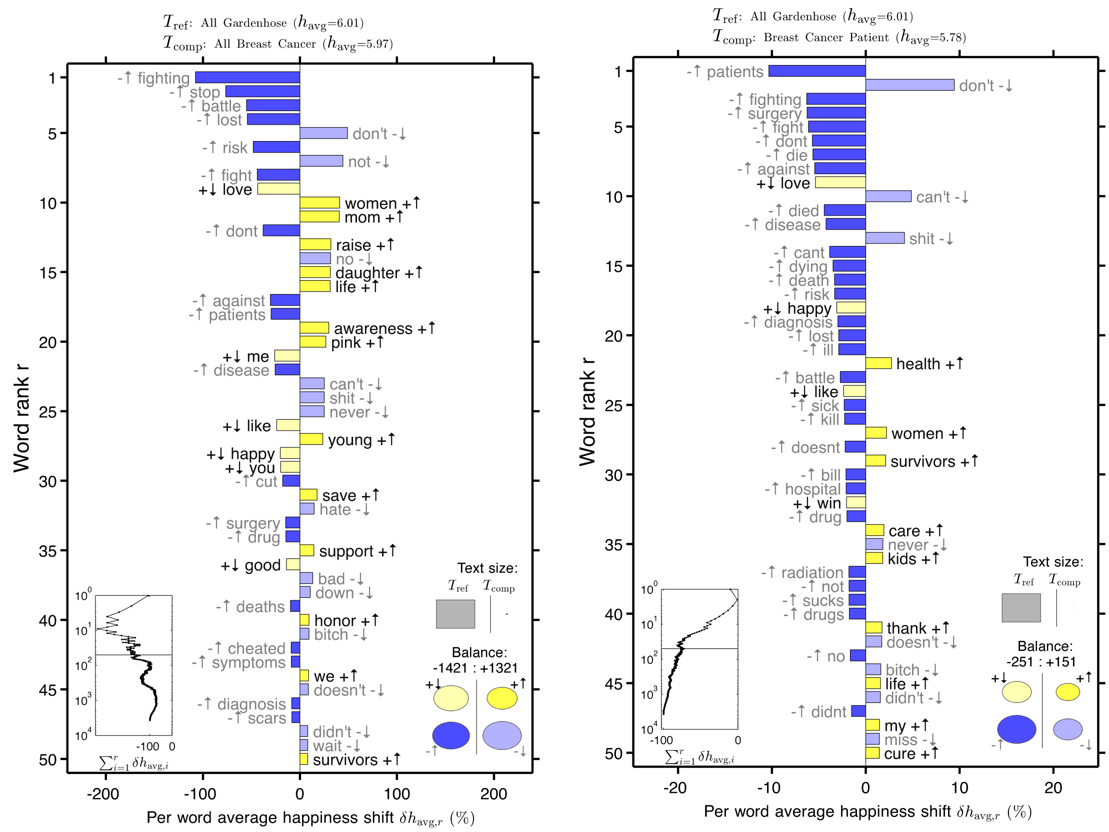


Figure C.4: (Left) A word shift graph comparing tweets collected mentioning breast cancer, T_{comp}, to a random unfiltered reference sample of tweets along the same time period. Breast cancer tweets were slightly less positive ($h_{\text{avg}} = 5.97$ v. 6.01) due to an increase in negative words 'fight(ing)', 'stop' , 'battle', 'lost', and 'risk'. This set of tweets featured a relative increase in positive words 'women', 'mom', 'daughter', 'awareness', 'pink', 'save', 'support', and 'survivors', which are referencing aspects of breast cancer awareness, support, and the experiences of survivors and patients. (Right) A word shift graph comparing breast cancer patient tweets to the unfiltered sample. These were more negative (h_{avg} = 5.78 v. 6.01) due to a relative increase in negative words such as 'dying', 'sick', 'killing', 'radiation', and 'hospital' among other terms similar to the figure on the left.

C.0.4 Sentence Classification Methodology

We built the vocabulary corpus for the logistic model by tokenizing the annotated set of patient tweets by word, removing punctuation, and lowercasing all text. We also included patient unrelated 'cancer' tweets collected as a frame of reference to train the classifier. This set of tweets was not annotated, so we made the assumption that tweets not validated by, Crannell et al. (2016) were patient unrelated. The proportion, α, of unrelated to related tweets has a profound effect on the vocabulary of the logistic model, so we experimented with various ranges of α and settled on a 1:10 ratio of patient related to unrelated tweets. We then applied the tf-idf statistic to build the binary classification logistic model.

The Tensorflow open source machine learning library has previously shown great promise when applied to NLP benchmark data-sets, Kim (2014) . The CNN loosely works by implementing a filter, called convolution functions, across various subregions of the feature landscape, Johnson and Zhang (2015); Britz (2015b), in this case the tweet vocabulary. The model tests the robustness of different word embeddings (e.g., phrases) by randomly removing filtered pieces during optimization to find the best predictive terms over the course of training. We divided the input labeled data into training and evaluation to successively test for the best word embedding predictors. The trained model can then be applied for binary classification of text content.

C.0.5 Hashtag Table Sorted by Average Word Happiness

Top Hashtags(#): All Breast Cancer					Top Hashtags(#): Breast Cancer Patient Sample				
Rank	Term	Tweets	Users	h_{avg}	Rank	Term	Tweets	Users	h_{avg}
1	#1savetatas	3,051	1,040	6.51	1	#crucialcatch	107	3	6.58
2	#nbcf	3,445	1,965	6.49	2	#mbcproject	39	11	6.28
3	#spas4acause	2,585	1,615	6.48	3	#idrivefor	40	19	6.27
4	#pink	3,480	2,763	6.34	4	#breast	79	24	6.2
5	#giveaway	4,861	1,284	6.31	5	#childhoodcancer	42	9	6.14
6	#obamacare	2,240	2,059	6.2	6	#clinicaltrials	51	12	6.13
7	#awareness	3,697	1,369	6.19	7	#kissthis4mbc	61	14	6.13
8	#twibbon	16,809	14,332	6.18	8	#breastcancerawarenessmonth	62	41	6.04
9	#bcam	2,555	1,961	6.14	9	#research	46	17	6.02
10	#breastcancerawareness	13,429	8,820	6.13	10	#lifeofafourthstager	41	3	6.0
11	#nfl	2,188	647	6.13	11	#aacr17	51	7	5.99
12	#pinkout	1,946	1,778	6.12	12	#curechat	70	3	5.97
13	#savethetatas	5,551	5,390	6.11	13	#mylymphedemalife	170	4	5.95
14	#nyfw	2,060	1,886	6.11	14	#worldcancerday	134	54	5.94
15	#thinkpink	2,707	2,209	6.1	15	#malebreastcancer	155	21	5.94
16	#unitedbyher	2,117	602	6.1	16	#lymphedema	680	12	5.93
17	#ad	3,458	1,383	6.08	17	#metastatic	161	17	5.92
18	#idrivefor	13,562	8,331	6.06	18	#bcsm	1,220	61	5.92
19	#breastcancerawarenessmonth	20,961	13,491	6.06	19	#metastaticbc	41	19	5.9
20	#aca	8,903	8,105	6.05	20	#lcsm	108	14	5.89
21	#worldcancerday	2,936	2,430	6.05	21	#breastcancerawareness	45	36	5.89
22	#himinitiative	7,294	572	6.04	22	#f***cancer	44	21	5.88
23	#ai	1,937	1,241	6.03	23	#cpat17	61	3	5.87
24	#walk	9,344	246	6.0	24	#saveaca	115	47	5.85
25	#malebreastcancer	5,821	1,469	6.0	25	#breastcancer	568	112	5.84
26	#breast	35,544	11,115	6.0	26	#oncology	54	11	5.84
27	#survivor	14,500	1,107	5.98	27	#survivor	92	33	5.83
28	#breastcancer	66,400	22,247	5.97	28	#immunotherapy	52	18	5.81
29	#research	1,912	1,634	5.96	29	#brca	42	17	5.81
30	#bcsm	14,955	4,644	5.95	30	#obamacare	132	42	5.77
31	#cancer	67,111	23,171	5.92	31	#bccww	112	24	5.77
32	#cnndebatenight	2,097	2,029	5.9	32	#cancer	2,063	239	5.76
33	#pinkribbon	2,201	1,104	5.9	33	#protectourcare	91	37	5.75
34	#brca	3,652	1,284	5.89	34	#cancersucks	57	34	5.75
35	#lymphedema	13,263	2,274	5.88	35	#acaworks	47	11	5.72
36	#women	2,078	1,231	5.85	36	#aca	469	88	5.69
37	#healthcare	2,261	1,396	5.85	37	#mbc	59	21	5.69
38	#oncology	2,188	762	5.85	38	#breastcancerrealitycheck	64	17	5.66
39	#health	17,484	5,696	5.82	39	#chokecancer	303	1	5.64
40	#news	6,435	1,680	5.79	40	#iamapreexistingcondition	82	55	5.63
41	#nobraday	23,406	16,785	5.76	41	#nhs	42	16	5.62
42	#donate	2,016	1,279	5.76	42	#amsm	165	25	5.61
43	#exercise	2,740	1,492	5.71	43	#projectpinkblue	178	1	5.56
44	#keepkadcyla	3,822	3,064	5.68	44	#healthcare	62	32	5.44
45	#sabcs16	2,104	828	5.67	45	#ahca	165	45	5.4
46	#ahca	2,607	2,403	5.67	46	#trumpcare	168	70	5.39
47	#thegoodlie	2,314	474	5.63	47	#maga	53	34	5.38
48	#trumpcare	4,778	4,331	5.53	48	#grahamcassidy	40	23	5.3
49	#avonrep	3,517	1,620	5.49	49	#stageivneedsmore	51	8	5.29
50	#iamapreexistingcondition	7,604	6,215	5.41	50	#gop	41	18	5.28
*	**Total**	462,192	155,218	5.96	*	**Total**	8,928	396	5.8

Table C.1: A table 50 frequently tweeted hashtags (#) sorted by average word happiness from all collected breast cancer tweets (left) and from sampled breast cancer patients (right).

Figure C.5: This word cloud displays the most prominent hashtags from all collected “breast cancer” tweets. The hashtag sizes are proportionate to their relative frequencies and colors represent their average ambient happiness scores. Here, light blue terms appear with the most positive LabMT words while purple hashtags appear with relatively more negative terms.

Appendix D: Sentiments and Public Perceptions of Surgery

D.0.1 Surgical Term List Twitter Counts and Sentiments

A surgical term list- constructed by medical professionals- for comparing prevalence and sentiments of tweets between disciplines. The tables below include each terms tweet count, it's relative weight among the sub-list, and it's relative average happiness score - colored relative to the sub-list's average computed word happiness (blue - negative, orange- positive).

Broad Symptom Terms				
Rank	Term	Tweet Count	Percent	h_{avg}
1	pain	71,780	29.8%	5.71
2	injury	50,688	21.1%	5.66
3	disease	22,858	9.5%	5.92
4	hurt	22,140	9.2%	5.77
5	sick	17,722	7.4%	5.83
6	condition	16,212	6.7%	5.34
7	suffering	5,740	2.4%	5.39
8	issue	5,474	2.3%	5.72
9	wound	3,433	1.4%	5.66
10	difficult	3,178	1.3%	5.97
11	illness	2,021	0.8%	5.57
12	complication	2,004	0.8%	5.43
13	trouble	1,940	0.8%	5.87
14	disorder	1,731	0.7%	5.8
15	struggle	1,679	0.7%	5.83
16	struggling	1,416	0.6%	5.82
17	defect	1,067	0.4%	6.22
18	sickness	1,008	0.4%	5.65
19	ailment	878	0.4%	5.84
20	strain	856	0.4%	5.35
21	discomfort	813	0.3%	5.63
22	deformed	666	0.3%	5.77
23	deformity	640	0.3%	5.95
24	abnormal	628	0.3%	6.17
25	agony	550	0.2%	5.47
26	embarrassed	499	0.2%	5.9
27	irregular	420	0.2%	5.9
28	soreness	408	0.2%	5.62
29	complaint	347	0.1%	5.65
30	frustration	284	0.1%	5.69
31	misery	283	0.1%	5.51
32	difficulty	257	0.1%	5.81
33	tender	252	0.1%	5.81
34	distress	246	0.1%	5.57
35	irritated	237	0.1%	5.57
36	dilemma	235	0.1%	5.9
37	impaired	120	0.0%	5.74
*	**Total**	240,710	100%	5.73

Broad Symptom Terms				
Rank	Term	Tweet Count	Percent	h_{avg}
1	stone	3,251	100.0%	5.69
1	hernia	15,092	99.0%	5.79
2	bulge	146	1.0%	5.63
*	**Total**	15,238	100%	5.71
1	blockage	1,243	42.0%	5.91
2	blocked	1,054	35.6%	5.85
3	obstruction	403	13.6%	5.73
4	clogged	263	8.9%	5.77
*	**Total**	2,963	100%	5.81
1	attack	9,385	66.3%	5.75
2	stroke	4,774	33.7%	5.73
*	**Total**	14,159	100%	5.74
1	pregnant	3,679	60.0%	5.88
2	pregnancy	2,326	38.0%	5.65
3	with child	122	2.0%	6.11
*	**Total**	6,127	100%	5.88
1	inflammation	567	63.6%	5.33
2	inflamed	325	36.4%	5.53
*	**Total**	892	100%	5.43
1	pathology	651	100.0%	5.65
*	**Total**	651	100%	5.65
1	fat	13,678	39.7%	5.65
2	obesity	12,201	35.4%	5.56
3	obese	8,557	24.8%	5.35
*	**Total**	34,436	100%	5.52
1	cancer	97,357	60.8%	6.02
2	tumor	36,354	22.7%	5.65
3	disease	22,858	14.3%	5.92
4	big c	1,125	0.7%	5.94
5	sickness	1,008	0.6%	5.65
6	malignant	690	0.4%	5.43
7	carcinoma	503	0.3%	5.71
8	corruption	296	0.2%	5.35
*	**Total**	160,191	100%	5.71

Table D.1: **Broad Symptom Terms** A table key word Twitter stats related to relevant surgical symptoms. The relative computed average happiness h_{avg} for each term is colored relative to the group's average (blue-negative, orange - positive).

Action Terms

Rank	Term	Tweet Count	Percent	h_{avg}
1	doctor	64,960	50.5%	5.83
2	repair	43,639	33.9%	5.55
3	improve	7,995	6.2%	5.82
4	correct	6,045	4.7%	5.62
5	restore	1,822	1.4%	5.86
6	patch	1,082	0.8%	5.85
7	sew	632	0.5%	5.78
8	remedy	609	0.5%	5.84
9	rebuild	523	0.4%	5.82
10	reform	375	0.3%	5.88
11	rectify	233	0.2%	5.46
12	revive	200	0.2%	5.82
13	rejuvenate	189	0.1%	5.67
14	overhaul	167	0.1%	6.14
15	refresh	152	0.1%	5.77
16	renew	121	0.1%	6.12
*	**Total**	128,744	100%	5.8
1	remove	67,498	52.5%	5.78
2	cut	25,891	20.2%	5.43
3	raise	15,695	12.2%	6.11
4	get rid of	10,575	8.2%	5.06
5	transfer	1,630	1.3%	5.46
6	separate	1,359	1.1%	6.09
7	eliminate	1,281	1.0%	4.81
8	take out	1,078	0.8%	5.77
9	discharge	669	0.5%	5.85
10	cut out	584	0.5%	5.73
11	amputate	556	0.4%	5.89
12	transport	533	0.4%	5.65
13	extract	399	0.3%	5.91
14	pull out	362	0.3%	5.55
15	erase	245	0.2%	5.8
16	rip out	113	0.1%	5.96
*	**Total**	128,468	100%	5.68
1	free	45,313	74.7%	5.86
2	heal	11,541	19.0%	5.85
3	attend	2,117	3.5%	5.74
4	alleviate	760	1.3%	4.88
5	settle	387	0.6%	5.75
6	minister to	207	0.3%	5.97
7	regenerate	205	0.3%	5.95
8	soothe	115	0.2%	6.16
*	**Total**	60,645	100%	5.77
1	fix	31,292	74.6%	5.55
2	sort	8,724	20.8%	5.21
3	reconstruct	909	2.2%	5.95
4	see to	314	0.7%	5.9
5	adjust	263	0.6%	5.97
6	face-lift	175	0.4%	5.65
7	revise	159	0.4%	5.98
8	regulate	111	0.3%	5.72
*	**Total**	41,947	100%	5.74

Action Terms

Rank	Term	Tweet Count	Percent	h_{avg}
1	better	77,400	80.3%	5.95
2	get better	6,220	6.4%	6.03
3	service	4,122	4.3%	5.91
4	gain	3,561	3.7%	5.36
5	mend	2,796	2.9%	5.86
6	aid	2,056	2.1%	5.92
7	recuperate	289	0.3%	6.14
*	**Total**	96,444	100%	5.88
1	deal	9,261	44.9%	5.64
2	survive	4,118	20.0%	5.81
3	handle	3,298	16.0%	5.78
4	suffer	1,791	8.7%	5.42
5	cope	671	3.3%	5.66
6	endure	464	2.2%	5.71
7	wrestle	385	1.9%	5.56
8	dispatch	221	1.1%	6.13
9	encounter	185	0.9%	5.74
10	get by	133	0.6%	5.84
11	buffet	114	0.6%	6.02
*	**Total**	20,641	100%	5.76
6.04				
*	**Total**	287,861	100%	5.89
1	nurse	57,284	82.6%	6.07
2	treat	6,873	9.9%	5.6
3	operate	2,321	3.3%	5.9
4	care for	2,303	3.3%	5.58
5	prescribe	234	0.3%	5.5
6	minister to	207	0.3%	5.97
7	administer	157	0.2%	5.71
*	**Total**	69,379	100%	5.76

Action Terms (continued)				
Rank	Term	Tweet Count	Percent	h_{avg}
1	nurse	57,284	85.5%	6.07
2	cure	7,000	10.5%	5.45
3	relieve	1,686	2.5%	5.29
4	dose	794	1.2%	5.59
5	minister to	207	0.3%	5.97
*	**Total**	66,971	100%	5.67
1	recover	17,401	47.9%	6.03
2	increase	7,550	20.8%	5.44
3	grow	3,673	10.1%	5.7
4	overcome	2,313	6.4%	5.27
5	get over	1,405	3.9%	5.59
6	pick up	1,274	3.5%	5.91
7	bounce back	1,019	2.8%	6.05
8	pull through	729	2.0%	6.11
9	rally	430	1.2%	5.96
10	make a comeback	217	0.6%	6.1
11	rebound	167	0.5%	5.89
12	get in shape	155	0.4%	5.91
*	**Total**	36,333	100%	5.83

Action Terms (continued)				
Rank	Term	Tweet Count	Percent	h_{avg}
1	help	157,162	54.6%	6.0
2	save	39,085	13.6%	6.27
3	support	37,571	13.1%	6.3
4	second	25,033	8.7%	5.93
5	further	4,927	1.7%	5.47
6	benefit	4,320	1.5%	5.6
7	work for	3,391	1.2%	5.85
8	cheer	3,138	1.1%	6.02
9	boost	2,889	1.0%	5.88
10	ease	2,240	0.8%	5.73
11	push	2,044	0.7%	5.85
12	go with	1,669	0.6%	5.9
13	promote	990	0.3%	5.91
14	serve	764	0.3%	6.02
15	maintain	584	0.2%	5.84
16	plug	511	0.2%	5.87
17	advocate	500	0.2%	5.85
18	encourage	483	0.2%	5.92
19	see through	127	0.0%	5.62
20	sustain	122	0.0%	5.56
21	stimulate	109	0.0%	5.99
22	bolster	101	0.0%	5.98
23	stand by	101	0.0%	6.04

Table D.2: **Action Terms :** A table of key word Twitter stats related to relevant to surgical actions.

Basic Symptom Terms

Rank	Term	Tweet Count	Percent	h_{avg}
1	weight loss	60,969	46.0%	5.95
2	nervous	22,268	16.8%	5.99
3	infection	9,111	6.9%	5.51
4	sore	5,826	4.4%	5.87
5	bleeding	4,674	3.5%	5.6
6	swelling	4,020	3.0%	5.48
7	anxiety	3,314	2.5%	5.62
8	fever	2,614	2.0%	5.44
9	growth	2,370	1.8%	5.81
10	anxious	2,111	1.6%	5.86
11	depression	1,617	1.2%	5.58
12	headaches	1,594	1.2%	5.24
13	bleed	1,050	0.8%	5.52
14	knee pain	925	0.7%	5.61
15	numbness	862	0.7%	5.26
16	cough	763	0.6%	5.53
17	vomiting	719	0.5%	5.34
18	allergic	629	0.5%	5.63
19	blood clots	596	0.4%	5.54
20	burning	582	0.4%	5.63
21	varicose veins	572	0.4%	5.89
22	nausea	471	0.4%	5.64
23	shoulder pain	442	0.3%	5.72
24	sore throat	393	0.3%	5.47
25	low back pain	353	0.3%	5.69

Basic Symptom Terms

Rank	Term	Tweet Count	Percent	h_{avg}
26	allergy	306	0.2%	5.63
27	neck pain	291	0.2%	5.56
28	tingling	282	0.2%	4.87
29	chest pain	265	0.2%	5.69
30	heartburn	253	0.2%	5.52
31	allergic reaction	246	0.2%	5.49
32	constipation	195	0.1%	5.39
33	hip pain	191	0.1%	5.65
34	chills	161	0.1%	5.83
35	thirst	160	0.1%	5.38
36	joint pain	155	0.1%	5.65
37	abdominal pain	146	0.1%	5.82
38	vision problems	143	0.1%	5.93
39	hearing loss	137	0.1%	5.94
40	nervousness	119	0.1%	6.03
41	blister	118	0.1%	5.7
42	urgency	116	0.1%	5.8
43	dizziness	113	0.1%	5.22
44	foot pain	106	0.1%	5.51
45	diarrhea	105	0.1%	5.58
*	**Total**	132,453	100%	5.61

Table D.3: **Basic Symptom Terms :** A table key word Twitter stats related to relevant surgical symptoms.

Body Region Terms

Rank	Term	Tweet Count	Percent	h_{avg}
1	brain	141,717	23.1%	5.76
2	eye	108,269	17.7%	5.86
3	face	71,279	11.6%	5.71
4	nose	46,301	7.6%	5.72
5	eyes	33,956	5.5%	5.82
6	hair	33,445	5.5%	5.46
7	head	32,960	5.4%	5.7
8	facial	30,471	5.0%	5.97
9	jaw	22,622	3.7%	5.7
10	ear	20,125	3.3%	5.23
11	mouth	18,081	3.0%	5.71
12	lip	9,192	1.5%	5.88
13	lips	8,410	1.4%	5.81
14	chin	5,208	0.8%	5.71
15	skull	4,781	0.8%	5.87
16	tongue	4,275	0.7%	5.86
17	ears	3,687	0.6%	5.86
18	forehead	3,031	0.5%	5.72
19	nasal	2,205	0.4%	5.93
20	wrinkles	2,139	0.3%	5.56
21	cheeks	2,095	0.3%	5.74
22	cheek	1,871	0.3%	5.8
23	hairline	1,301	0.2%	5.41
24	cerebral	1,223	0.2%	5.86
25	eyebrows	1,174	0.2%	5.87
26	eyebrow	665	0.1%	5.8
27	cranial	652	0.1%	5.95
28	eyeball	615	0.1%	5.78
29	orbit	575	0.1%	4.11
30	jaws	362	0.1%	5.62
31	cranium	101	0.0%	6.01
*	**Total**	612,788	100%	5.7
1	neck	30,740	55.8%	5.74
2	throat	13,581	24.6%	5.67
3	thyroid	4,273	7.8%	5.57
4	vocal cord	3,876	7.0%	5.96
5	vocal cords	972	1.8%	5.69
6	adams apple	525	1.0%	6.0
7	adam's apple	419	0.8%	6.15
8	voice box	293	0.5%	5.65
9	trachea	292	0.5%	5.58
10	larynx	161	0.3%	5.38
*	**Total**	55,132	100%	5.74
1	back	209,822	83.4%	5.86
2	spinal	16,870	6.7%	5.88
3	spine	15,814	6.3%	5.91
4	thoracic	3,848	1.5%	5.8
5	spinal cord	1,656	0.7%	5.82
6	lumbar	1,496	0.6%	5.67
7	cervical	1,365	0.5%	5.67
8	vertebra	390	0.2%	5.64
9	tailbone	239	0.1%	5.51
10	vertebral	105	0.0%	4.78
*	**Total**	251,605	100%	5.65

Body Region Terms

Rank	Term	Tweet Count	Percent	h_{avg}
1	heart	212,976	63.1%	5.92
2	breast	60,097	17.8%	5.59
3	breasts	12,184	3.6%	5.95
4	boobs	11,037	3.3%	5.75
5	lung	7,251	2.1%	5.63
6	chest	6,360	1.9%	5.69
7	boob	6,080	1.8%	5.91
8	tits	3,166	0.9%	5.69
9	coronary	2,677	0.8%	5.9
10	clavicle	2,211	0.7%	5.44
11	lungs	1,906	0.6%	5.79
12	nipples	1,784	0.5%	5.74
13	ribs	1,381	0.4%	5.51
14	rib	1,273	0.4%	5.72
15	airway	1,164	0.3%	6.29
16	nipple	1,037	0.3%	5.84
17	titties	900	0.3%	5.67
18	collar bone	733	0.2%	5.57
19	sternum	534	0.2%	5.44
20	boobies	517	0.2%	5.94
21	pulmonary	512	0.2%	5.64
22	esophageal	447	0.1%	5.48
23	tit	378	0.1%	5.71
24	trachea	292	0.1%	5.58
25	esophagus	284	0.1%	5.59
26	boobie	147	0.0%	5.92
27	mammary	133	0.0%	5.56
*	**Total**	337,461	100%	5.72
1	kidney	15,073	18.6%	5.7
2	stomach	12,818	15.8%	5.55
3	belly	10,126	12.5%	5.58
4	liver	9,060	11.2%	5.76
5	abdominal	5,591	6.9%	5.54
6	gallbladder	4,820	5.9%	5.7
7	tummy	4,666	5.8%	5.79
8	colon	4,321	5.3%	5.59
9	six pack	2,470	3.0%	7.0
10	bowel	2,339	2.9%	5.6
11	gall bladder	2,299	2.8%	5.74
12	abdomen	1,477	1.8%	5.48
13	gut	1,091	1.3%	5.61
14	spleen	1,026	1.3%	5.46
15	pancreatic	955	1.2%	5.21
16	intestine	837	1.0%	5.58
17	pancreas	601	0.7%	5.91
18	renal	591	0.7%	5.25
19	inguinal	321	0.4%	5.9
20	bile	292	0.4%	5.65
21	love handle	125	0.2%	5.42
22	love handles	121	0.1%	5.38
23	bile duct	117	0.1%	5.63
*	**Total**	81,137	100%	5.65

Body Region Terms (continued)				
Rank	Term	Tweet Count	Percent	h_{avg}
1	penis	14,248	17.9%	5.79
2	prostate	11,196	14.1%	5.15
3	groin	6,426	8.1%	5.34
4	dick	6,058	7.6%	5.65
5	vagina	5,776	7.3%	5.84
6	testicle	5,703	7.2%	5.05
7	bladder	5,549	7.0%	5.59
8	pussy	3,843	4.8%	5.74
9	genital	3,315	4.2%	5.69
10	ovarian	2,999	3.8%	5.25
11	anal	2,385	3.0%	5.99
12	rectal	1,682	2.1%	5.41
13	ovaries	1,435	1.8%	5.13
14	pelvic	1,337	1.7%	5.7
15	uterus	1,207	1.5%	5.7
16	uterine	1,180	1.5%	5.33
17	scrotum	1,157	1.5%	5.77
18	pelvis	724	0.9%	5.61
19	ovary	588	0.7%	5.72
20	anus	483	0.6%	5.65
21	testicles	482	0.6%	5.67
22	rectum	435	0.5%	5.66
23	fallopian tube	367	0.5%	4.76
24	fallopian tubes	307	0.4%	4.65
25	cervix	196	0.2%	5.63
26	genitalia	190	0.2%	5.81
27	ureter	142	0.2%	6.1
28	urethra	118	0.1%	5.77
*	**Total**	79,528	100%	5.54
1	artery	2,606	30.4%	5.8
2	lymph	1,944	22.7%	5.27
3	vein	1,096	12.8%	5.75
4	arteries	750	8.8%	5.95
5	node	676	7.9%	5.27
6	carotid	464	5.4%	5.57
7	aorta	404	4.7%	5.86
8	vessel	366	4.3%	5.46
9	femoral	261	3.0%	5.69
*	**Total**	8,567	100%	5.62

Body Region Terms (continued)				
Rank	Term	Tweet Count	Percent	h_{avg}
1	joint	5,241	88.8%	5.66
1	knee	249,927	47.2%	5.77
2	hip	52,489	9.9%	5.86
3	foot	49,009	9.3%	5.66
4	leg	43,789	8.3%	5.64
5	ankle	40,987	7.7%	5.66
6	ass	26,190	4.9%	5.62
7	butt	18,655	3.5%	5.7
8	toe	9,064	1.7%	5.75
9	feet	6,471	1.2%	5.86
10	knees	6,368	1.2%	5.71
11	tibia	4,489	0.8%	5.49
12	fibula	2,787	0.5%	5.48
13	toes	2,747	0.5%	5.94
14	booty	2,522	0.5%	5.79
15	thigh	2,456	0.5%	5.33
16	patella	1,660	0.3%	5.49
17	femur	1,552	0.3%	5.64
18	calf	1,545	0.3%	5.68
19	ankles	1,500	0.3%	5.68
20	buttocks	1,451	0.3%	5.67
21	heels	1,417	0.3%	5.88
22	heel	1,034	0.2%	5.64
23	thighs	502	0.1%	5.67
24	tibial	249	0.0%	5.62
25	sole	241	0.0%	5.6
26	sciatic	183	0.0%	5.42
27	calves	102	0.0%	5.66
*	**Total**	529,386	100%	5.66

Table D.4: **Body Region Terms :** A table key word Twitter stats related to relevant body region.

Body Region Terms (continued)				
Rank	Term	Tweet Count	Percent	h_{avg}
1	shoulder	71,469	29.4%	5.75
2	hand	40,432	16.7%	5.9
3	wrist	32,290	13.3%	5.73
4	elbow	22,982	9.5%	5.6
5	arm	21,915	9.0%	5.54
6	thumb	13,804	5.7%	5.5
7	finger	10,763	4.4%	5.58
8	hands	7,996	3.3%	6.0
9	fingers	5,341	2.2%	5.91
10	arms	2,725	1.1%	5.84
11	forearm	2,429	1.0%	5.08
12	carpal	2,115	0.9%	5.7
13	pinky	2,078	0.9%	5.48
14	triceps	1,253	0.5%	5.64
15	shoulders	1,100	0.5%	5.64
16	biceps	1,004	0.4%	5.21
17	ulnar	937	0.4%	5.54
18	wrists	412	0.2%	5.79
19	elbows	270	0.1%	5.78
20	knuckle	258	0.1%	5.31
21	radial	239	0.1%	4.91
22	pinkie	229	0.1%	5.71
23	radius	193	0.1%	5.57
24	knuckles	162	0.1%	5.87
25	humerus	131	0.1%	5.71
26	ulna	130	0.1%	5.29
27	scapula	123	0.1%	5.22
*	**Total**	242,780	100%	5.59

Table D.5: **Body Region Terms :** A table key word Twitter stats related to relevant body region.