Characterizing diabetes, diet, exercise, and obesity comments on Twitter

Amir Karami[a,*], Alicia A. Dahl[b], Gabrielle Turner-McGrievy[b], Hadi Kharrazi[c], George Shaw Jr.[a]

[a] University of South Carolina, School of Library and Information Science, United States
[b] University of South Carolina, Arnold School of Public Health, United States
[c] Johns Hopkins University, Bloomberg School of Public Health, United States

ARTICLE INFO

ABSTRACT

Social media provide a platform for users to express their opinions and share information. Understanding public health opinions on social media, such as Twitter, offers a unique approach to characterizing common health issues such as diabetes, diet, exercise, and obesity (DDEO); however, collecting and analyzing a large scale conversational public health data set is a challenging research task. The goal of this research is to analyze the characteristics of the general public's opinions in regard to diabetes, diet, exercise and obesity (DDEO) as expressed on Twitter. A multi-component semantic and linguistic framework was developed to collect Twitter data, discover topics of interest about DDEO, and analyze the topics. From the extracted 4.5 million tweets, 8% of tweets discussed diabetes, 23.7% diet, 16.6% exercise, and 51.7% obesity. The strongest correlation among the topics was determined between exercise and obesity ($p < .0002$). Other notable correlations were: diabetes and obesity ($p < .0005$), and diet and obesity ($p < .001$). DDEO terms were also identified as subtopics of each of the DDEO topics. The frequent subtopics discussed along with "Diabetes", excluding the DDEO terms themselves, were blood pressure, heart attack, yoga, and Alzheimer. The non-DDEO subtopics for "Diet" included vegetarian, pregnancy, celebrities, weight loss, religious, and mental health, while subtopics for "Exercise" included computer games, brain, fitness, and daily plan. Non-DDEO subtopics for "Obesity" included Alzheimer, cancer, and children. With 2.67 billion social media users in 2016, publicly available data such as Twitter posts can be utilized to support clinical providers, public health experts, and social scientists in better understanding common public opinions in regard to diabetes, diet, exercise, and obesity.

1. Introduction

The global prevalence of obesity has doubled between 1980 and 2014, with more than 1.9 billion adults considered as overweight and over 600 million adults considered as obese in 2014 (World Health Organization Fact Sheet, 2016). Since the 1970s, obesity has risen 37% affecting 25% of the U.S. adults (Flegal, Carroll, Kit, & Ogden, 2012). Similar upward trends of obesity have been found in youth populations, with a 60% increase in preschool aged children between 1990 and 2010 (Harvard HSPH, 2017). Overweight and obesity are the fifth leading risk for global deaths according to the European Association for the Study of Obesity (World Health Organization Fact Sheet, 2016). Excess energy intake and inadequate energy expenditure both contribute to weight gain and diabetes (Hill, Wyatt, & Peters, 2012; Wing et al., 2001).

Obesity can be reduced through modifiable lifestyle behaviors such as diet and exercise (Wing et al., 2001). There are several comorbidities associated with being overweight or obese, such as diabetes (Kopelman,

2000). The prevalence of diabetes in adults has risen globally from 4.7% in 1980 to 8.5% in 2014. Current projections estimate that by 2050, 29 million Americans will be diagnosed with type 2 diabetes, which is a 165% increase from the 11 million diagnosed in 2002 (Boyle et al., 2001). Studies show that there are strong relations among diabetes, diet, exercise, and obesity (DDEO) (Association et al., 2004; Barnard et al., 2009; Hartz, Rupley, Kalkhoff, & Rimm, 1983; Wing et al., 2001); however, the general public's perception of DDEO remains limited to survey-based studies (Tompson et al., 2012).

The growth of social media has provided a research opportunity to track public behaviors, information, and opinions about common health issues. It is estimated that the number of social media users will increase from 2.34 billion in 2016 to 2.95 billion in 2020 (Statista, 2017). Twitter has 316 million users worldwide (Olanoff, 2015) providing a unique opportunity to understand users' opinions with respect to the most common health issues (Mejova, Weber, & Macy, 2015). Publicly available Twitter posts have facilitated data collection and leveraged the research at the intersection of public health and data science; thus,

informing the research community of major opinions and topics of interest among the general population (Nasukawa & Yi, 2003; Wiebe et al., 2003; Zabin & Jefferies, 2008) that cannot otherwise be collected through traditional means of research (e.g., surveys, interviews, focus groups) (Eichstaedt et al., 2015; Wartell, 2015). Furthermore, analyzing Twitter data can help health organizations such as state health departments and large healthcare systems to provide health advice and track health opinions of their populations and provide effective health advice when needed (Mejova et al., 2015).

Among computational methods to analyze tweets, computational linguistics is a well-known developed approach to gain insight into a population, track health issues, and discover new knowledge (Moreland-Russell, Tabak, Ruhr, & Maier, 2014; Paul & Dredze, 2011, 2012; Zhao et al., 2011). Twitter data has been used for a wide range of health and non-health related applications, such as stock market (Bollen, Mao, & Zeng, 2011) and election analysis (Tumasjan, Sprenger, Sandner, & Welpe, 2010). Some examples of Twitter data analysis for health-related topics include: flu (Culotta, 2010; Lampos & Cristianini, 2010, 2012; Lampos, De Bie, & Cristianini, 2010; Ritterman, Osborne, & Klein, 2009; Szomszor, Kostkova, & De Quincey, 2010), mental health (Coppersmith, Dredze, Harman, & Hollingshead, 2015), Ebola (Lazard, Scheinfeld, Bernhardt, Wilcox, & Suran, 2015; Odlum & Yoon, 2015), Zika (Fu et al., 2016), medication use (Buntain & Golbeck, 2015; Hanson, Cannon, Burton, & Giraud-Carrier, 2013; Scanfeld, Scanfeld, & Larson, 2010), diabetes (Harris, Mueller, Snider, & Haire-Joshu, 2013), and weight loss and obesity (Dahl, Hales, & Turner-McGrievy, 2016; Ghosh & Guha, 2013; Harris et al., 2014; Turner-McGrievy & Beets, 2015; Vickey, Ginis, & Dabrowski, 2013).

The previous Twitter studies have dealt with extracting common topics of one health issue discussed by the users to better understand common themes; however, this study utilizes an innovative approach to computationally analyze unstructured health related text data exchanged via Twitter to characterize health opinions regarding four common health issues, including diabetes, diet, exercise and obesity (DDEO) on a population level. This study identifies the characteristics of the most common health opinions with respect to DDEO and discloses public perception of the relationship among diabetes, diet, exercise and obesity. These common public opinions/topics and perceptions can be used by providers and public health agencies to better understand the common opinions of their population denominators in regard to DDEO, and reflect upon those opinions accordingly.

2. Methods

Our approach uses semantic and linguistics analyses for disclosing health characteristics of opinions in tweets containing DDEO words. The present study included three phases: data collection, topic discovery, and topic-content analysis.

2.1. Data collection

This phase collected tweets using Twitter's Application Programming Interfaces (API) (Twitter, 2017). Within the Twitter API, diabetes, diet, exercise, and obesity were selected as the related words (Wing et al., 2001) and the related health areas (Paul & Dredze, 2011). Twitter's APIs provides both historic and real-time data collections. The latter method randomly collects 1% of publicly available tweets. This paper used the real-time method to randomly collect 10% of publicly available English tweets using several pre-defined DDEO-related queries (Table 1) within a specific time frame. We used the queries to collect approximately 4.5 million related tweets between 06/01/2016 and 06/30/2016. The data will be available in the first author's website. Fig. 1 shows a sample of collected tweets in this research.

Table 1
DDEO queries.

Health issue	Queries	Number of tweets	Percentage
Diabetes	diabetes OR #diabetes	353,655	8.0%
Diet	diet OR #diet OR dieting	1,045,374	23.7%
Exercise	exercise OR #exercise OR exercising	734,118	16.6%
Obesity	obesity OR #obesity OR fat	2,283,517	51.7%

2.2. Topic discovery

To discover topics from the collected tweets, we used a topic modeling approach that fuzzy clusters the semantically related words such as assigning "diabetes", "cancer", and "influenza" into a topic that has an overall "disease" theme (Karami, 2015; Karami, Gangopadhyay, Zhou, & Kharrazi, 2017). Topic modeling has a wide range of applications in health and medical domains such as predicting protein-protein relationships based on the literature knowledge (Asou & Eguchi, 2008), discovering relevant clinical concepts and structures in patients' health records (Arnold, El-Saden, Bui, & Taira, 2010), and identifying patterns of clinical events in a cohort of brain cancer patients (Arnold & Speier, 2012).

Among topic models, Latent Dirichlet Allocation (LDA) (Blei, Ng, & Jordan, 2003) is the most popular effective model (Lu, Mei, & Zhai, 2011; Paul & Dredze, 2011) as studies have shown that LDA is an effective computational linguistics model for discovering topics in a corpus (Hong & Davison, 2010; Mcauliffe & Blei, 2008). LDA assumes that a corpus contains topics such that each word in each document can be assigned to the topics with different degrees of membership (Karami & Gangopadhyay, 2014; Karami, Gangopadhyay, Zhou, & Kharrazi, 2015a, 2015b).

Twitter users can post their opinions or share information about a subject to the public. Identifying the main topics of users' tweets provides an interesting point of reference, but conceptualizing larger subtopics of millions of tweets can reveal valuable insight to users' opinions. The topic discovery component of the study approach uses LDA to find main topics, themes, and opinions in the collected tweets.
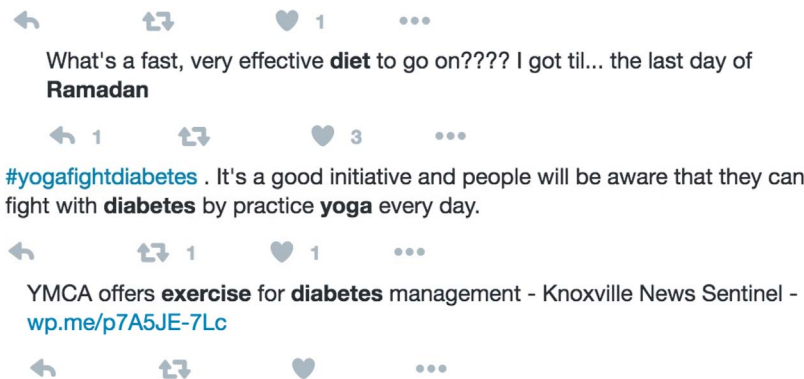
We used the Mallet implementation of LDA (Blei et al., 2003; McCallum, 2002) with its default settings to explore opinions in the tweets. Before identifying the opinions, two pre-processing steps were implemented: (1) using a standard list for removing stop words, that do not have semantic value for analysis (such as "the"); and, (2) finding the optimum number of topics. To determine a proper number of topics, log-likelihood estimation with 80% of tweets for training and 20% of tweets for testing was used to find the highest log-likelihood, as it is the optimum number of topics (Wallach, Murray, Salakhutdinov, & Mimno, 2009). The highest log-likelihood was determined 425 topics.

2.3. Topic content analysis

The topic content analysis component used an objective interpretation approach with a lexicon-based approach to analyze the content of topics. The lexicon-based approach uses dictionaries to disclose the semantic orientation of words in a topic. Linguistic Inquiry and Word Count (LIWC) is a linguistics analysis tool that reveals thoughts, feelings, personality, and motivations in a corpus (Karami & Zhou, 2014a, 2014b, 2015). LIWC has accepted rate of sensitivity, specificity, and English proficiency measures (Golder & Macy, 2011). LIWC has a health related dictionary that can help to find whether a topic contains words associated with health. In this analysis, we used LIWC to find health related topics.

Maternal AND paternal **obesity** can increase subsequent **breast cancer** risk in daughters! Scientific Reports 2016

↩ ⇄ ♥ 1 •••

What's a fast, very effective **diet** to go on???? I got til... the last day of **Ramadan**

↩ 1 ⇄ ♥ 3 •••

#yogafightdiabetes . It's a good initiative and people will be aware that they can fight with **diabetes** by practice **yoga** every day.

↩ ⇄ 1 ♥ 1 •••

YMCA offers **exercise** for **diabetes** management - Knoxville News Sentinel - wp.me/p7A5JE-7Lc

↩ ⇄ ♥ •••

Fig. 1. A sample of tweets.

3. Results

Obesity and Diabetes showed the highest and the lowest number of tweets (51.7% and 8.0%). Diet and Exercise formed 23.7% and 16.6% of the tweets (Table 1).

Out of all 4.5 million DDEO-related tweets returned by Tweeter's API, the LDA found 425 topics. We used LIWC to filter the detected 425 topics and found 222 health-related topics. Additionally, we labeled topics based on the availability of DDEO words. For example, if a topic had "diet", we labeled it as a diet-related topic. As expected and driven by the initial Tweeter API's query, common topics were diabetes, diet, exercise, and obesity (DDEO). Table 2 shows that the highest and the lowest number of topics were related to exercise and diabetes (80 and 21 out of 222). Diet and Obesity had almost similar rates (58 and 63 out of 222).

Each of the DDEO topics included several common subtopics including both DDEO and non-DDEO terms discovered by the LDA algorithm (Table 2). Common subtopics for "Diabetes", in order of frequency, included type 2 diabetes, obesity, diet, exercise, blood pressure, heart attack, yoga, and Alzheimer. Common subtopics for "Diet" included obesity, exercise, weight loss [medicine], celebrities, vegetarian, diabetes, religious diet, pregnancy, and mental health. Frequent subtopics for "Exercise" included fitness, obesity, daily plan, diet, brain, diabetes, and computer games. And finally, the most common subtopics for "Obesity" included diet, exercise, children, diabetes, Alzheimer, and cancer (Table 2). Table 3 provides illustrative examples for each of the topics and subtopics.

Further exploration of the subtopics revealed additional patterns of interest (Tables 2 and 3). We found 21 diabetes-related topics with 8 subtopics. While type 2 diabetes was the most frequent of the sub-topics, heart attack, Yoga, and Alzheimer are the least frequent subtopics for diabetes. Diet had a wide variety of emerging themes ranging from celebrity diet (e.g., Beyonce) to religious diet (e.g., Ramadan). Diet was detected in 63 topics with 10 subtopics; obesity, and pregnancy and mental health were the most and the least discussed obesity-related topics, respectively. Exploring the themes for Exercise subtopics revealed subjects such as computer games (e.g., Pokemon-Go) and brain exercises (e.g., memory improvement). Exercise had 7 subtopics with fitness as the most discussed subtopic and computer games as the least discussed subtopic. Finally, Obesity themes showed topics such as Alzheimer (e.g., research studies) and cancer (e.g., breast cancer). Obesity had the lowest diversity of subtopics: six with diet as the most discussed subtopic, and Alzheimer and cancer as the least discussed subtopics.

Diabetes subtopics show the relation between diabetes and exercise, diet, and obesity. Subtopics of diabetes revealed that users post about the relationship between diabetes and other diseases such as heart attack (Tables 2 and 3). The subtopic Alzheimer is also shown in the obesity subtopics. This overlap between categories prompts the discussion of research and linkages among obesity, diabetes, and Alzheimer's disease. Type 2 diabetes was another subtopic expressed by users and scientifically documented in the literature.

Table 2
DDEO topics and subtopics – diabetes, diet, exercise, and obesity are shown with italic and underline styles in subtopics.

Topics	Frequency	Subtopics	Distributions (%)	Topics	Frequency	Subtopics	Distributions (%)
Diabetes	21	Diabetes type 2	42.87%	Diet	63	*Obesity*	39.69%
		Obesity	14.29%			*Exercise*	15.87%
		Diet	9.52%			Weight loss	12.71%
		Exercise	9.52%			Celebrities	9.52%
		Blood pressure	9.52%			Vegetarian	9.52%
		Heart attack	4.76%			*Diabetes*	3.17%
		Yoga	4.76%			Religious diet	3.17%
		Alzheimer	4.76%			Weight loss medicine	3.17%
						Pregnancy	1.59%
						Mental health	1.59%
Exercise	80	Fitness	32.5%	Obesity	58	*Diet*	43.11%
		Obesity	22.5%			*Exercise*	31.04%
		Daily plan	21.25%			Children	17.24%
		Diet	11.25%			*Diabetes*	5.17%
		Brain	8.75%			Alzheimer	1.72%
		Diabetes	2.50%			Cancer	1.72%
		Computer games	1.25%				

Table 3
Topics examples.

Blood pressure	Heart attack	Diabetes type 2	Yoga	Alzheimer	Obesity	Diet and exercise	Obesity
risk	heart	change	diabetes	medicine	diabetes	helps	health
blood	diabetes	diabetes	#yogafightsdiabetes	diseases	surgery	diabetes	diet
high	cardiovascular	#lifestyle	yoga	common	treatment	children	obesity
diabetes	attack	type	control	drugs	obesity	exercise	immune
pressure	stroke	ii	life	Alzheimer	cure	diet	syndrome

Vegetarian	Pregnancy diet	Celebrities diet	Weight loss diet	Weight loss medicine	Religious diet	Mental health	Exercise and diabetes
diet	pregnancy	diet	weightlose	diet	burning	health	helps
eat	motherhood	beyonce	effective	#weightloss	#weightloss	nutrition	diabetes
fruits	diet	tips	morning	slimming	fasting	benefits	children
vegetables	baby	fatloss	dieting	pills	Ramadan	healing	exercise
fresh	motherhood	#angelinajolie	banana	#fatburners	diets	#mentalhealth	diet

Diet	Daily plan	Computer games	Brain	Fitness	Diet and diabetes	Obesity	Exercise
diet	food	exercise	exercise	fitness	helps	workout	bellyfat
exercise	exercise	finding	brain	#gymlife	diabetes	burning	losing
protein	calorie	pokemon	improve	bodybuilding	children	exercise	exercise
beauty	goal	#pokemongo	memory	gym	exercise	fatburn	ways
muscle	completed	hour	performance	workout	diet	obesity	effective

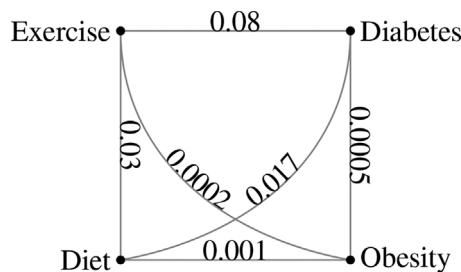| Diet	Alzheimer	Cancer	Children	Diabetes
health	study	cancer	obesity	diabetes
diet	link	breast	kids	surgery
obesity	Alzheimer	study	childhood	treatment
immune	obesity	risk	rates	obesity
syndrome	research	obesity	problem	cure

Fig. 2. DDEO correlation *p*-value.

The main DDEO topics showed some level of interrelationship by appearing as subtopics of other DDEO topics. The words with *italic* and <u>underline</u> styles in Table 2 demonstrate the relation among the four DDEO areas. Our results show users' interest about posting their opinions, sharing information, and conversing about exercise & diabetes, exercise & diet, diabetes & diet, diabetes & obesity, and diet & obesity (Fig. 2). The strongest correlation among the topics was determined to be between exercise and obesity ($p < .0002$). Other notable correlations were: diabetes and obesity ($p < .0005$), and diet and obesity ($p < .001$).

4. Discussion

Diabetes, diet, exercise, and obesity are common public health related opinions. Analyzing individual-level opinions by automated algorithmic techniques can be a useful approach to better characterize health opinions of a population. Traditional public health polls and surveys are limited by a small sample size; however, Twitter provides a platform to capture an array of opinions and shared information a expressed in the words of the tweeter. Studies show that Twitter data can be used to discover trending topics, and that there is a strong correlation between Twitter health conversations and Centers for Disease

Control and Prevention (CDC) statistics (Prier, Smith, Giraud-Carrier, & Hanson, 2011).

This research provides a computational content analysis approach to conduct a deep analysis using a large data set of tweets. Our framework decodes public health opinions in DDEO related tweets, which can be applied to other public health issues. Among health-related subtopics, there are a wide range of topics from diseases to personal experiences such as participating in religious activities or vegetarian diets.

Diabetes subtopics showed the relationship between diabetes and exercise, diet, and obesity (Tables 2 and 3). Subtopics of diabetes revealed that users posted about the relation between diabetes and other diseases such as heart attack. The subtopic Alzheimer is also shown in the obesity subtopics. This overlap between categories prompts the discussion of research and linkages among obesity, diabetes, and Alzheimer's disease. Type 2 diabetes was another subtopic that was also expressed by users and scientifically documented in the literature. The inclusion of Yoga in posts about diabetes is interesting. While yoga would certainly be labeled as a form of fitness, when considering the post, it was insightful to see discussion on the mental health benefits that yoga offers to those living with diabetes (Ross & Thomas, 2010).

Diet had the highest number of subtopics. For example, religious diet activities such as fasting during the month of Ramadan for Muslims incorporated two subtopics categorized under the diet topic (Tables 2 and 3). This information has implications for the type of diets that are being practiced in the religious community, but may help inform religious scholars who focus on health and psychological conditions during fasting. Other religions such as Judaism, Christianity, and Taoism have periods of fasting that were not captured in our data collection, which may have been due to lack of posts or the timeframe in which we collected the data. The diet plans of celebrities were also considered influential to explaining and informing diet opinions of Twitter users (Boyington et al., 2008).

Exercise themes show the Twitter users' association of exercise with "brain" benefits such as increased memory and cognitive performance

(Tables 2 and 3) (Cotman & Berchtold, 2002). The topics also confirm that exercising is associated with controlling diabetes and assisting with meal planning (Association et al., 2004; Laaksonen et al., 2005), and obesity (Ross et al., 2000). Additionally, we found the Twitter users mentioned exercise topics about the use of computer games that assist with exercising. The recent mobile gaming phenomenon Pokeman-Go game (Pokeman-Go Game, 2017) was highly associated with the exercise topic. Pokemon-Go allows users to operate in a virtual environment while simultaneously functioning in the real word. Capturing Pokemons, battling characters, and finding physical locations for meeting other users required physically activity to reach predefined locations. These themes reflect on the potential of augmented reality in increasing patients' physical activity levels (Schwarzer, 2008).

Obesity had the lowest number of subtopics in our study. Three of the subtopics were related to other diseases such as diabetes (Tables 2 and 3). The scholarly literature has well documented the possible linkages between obesity and chronic diseases such as diabetes (Flegal et al., 2012) as supported by the study results. The topic of children is another prominent subtopic associated with obesity. There has been an increasing number of opinions in regard to child obesity and national health campaigns that have been developed to encourage physical activity among children (PLAY 60 Challenge, 2017). Alzheimer was also identified as a topic under obesity. Although considered a perplexing finding, recent studies have been conducted to identify possible correlation between obesity and Alzheimer's disease (Kivipelto et al., 2005; Luchsinger, Cheng, Tang, Schupf, & Mayeux, 2012; Verdile et al., 2015). Indeed, Twitter users have expressed opinions about the study of Alzheimer's disease and the linkage between these two topics.

This paper addresses a need for clinical providers, public health experts, and social scientists to utilize a large conversational data set to collect and utilize population level opinions and information needs. Although our framework is applied to Twitter, the applications from this study can be used in patient communication devices monitored by physicians or weight management interventions with social media accounts, and support large scale population-wide initiatives to promote healthy behaviors and preventative measures for diabetes, diet, exercise, and obesity.

This research has some limitations. First, our DDEO analysis does not take geographical location of the Twitter users into consideration and thus does not reveal if certain geographical differences exists. Second, we used a limited number of queries to select the initial pool of tweets, thus perhaps missing tweets that may have been relevant to DDEO but have used unusual terms referenced. Third, our analysis only included tweets generated in one month; however, as our previous work has demonstrated (Turner-McGrievy & Beets, 2015), public opinion can change during a year. Additionally, we did not track individuals across time to detect changes in common themes discussed. Our future research plans includes introducing a dynamic framework to collect and analyze DDEO related tweets during extended time periods (multiple months) and incorporating spatial analysis of DDEO-related tweets.

5. Conclusion

This study represents the first step in developing routine processes to collect, analyze, and interpret DDEO-related posts to social media around health-related topics and presents a transdisciplinary approach to analyzing public discussions around health topics. With 2.34 billion social media users in 2016, the ability to collect and synthesize social media data will continue to grow. Developing methods to make this process more streamlined and robust will allow for more rapid identification of public health trends in real time.

Conflict of interest

The authors state that they have no conflict of interest.

References

Arnold, C., & Speier, W. (2012). *A topic model of clinical reports. Proceedings of the 35th international ACM SIGIR conference on research and development in information retrieval.* ACM1031–1032.

Arnold, C. W., El-Saden, S. M., Bui, A. A., & Taira, R. (2010). *Clinical case-based retrieval using latent topic analysis. AMIA annual symposium proceedings, Vol. 2010.* American Medical Informatics Association26.

Asou, T., & Eguchi, K. (2008). *Predicting protein–protein relationships from literature using collapsed variational latent Dirichlet allocation. Proceedings of the 2nd international workshop on data and text mining in bioinformatics.* ACM77–80.

Association, A. D., et al. (2004). Physical activity/exercise and diabetes. *Diabetes Care, 27*(suppl 1), s58–s62.

Barnard, N. D., Cohen, J., Jenkins, D. J., Turner-McGrievy, G., Gloede, L., Green, A., et al. (2009). A low-fat vegan diet and a conventional diabetes diet in the treatment of type 2 diabetes: A randomized, controlled, 74-wk clinical trial. *The American Journal of Clinical Nutrition,* 26736H.

Blei, D. M., Ng, A. Y., & Jordan, M. I. (2003). Latent Dirichlet allocation. *Journal of Machine Learning Research, 3,* 993–1022.

Bollen, J., Mao, H., & Zeng, X. (2011). Twitter mood predicts the stock market. *Journal of Computational Science, 2*(1), 1–8.

Boyington, J. E., Carter-Edwards, L., Piehl, M., Hutson, J., Langdon, D., & McManus, S. (2008). Cultural attitudes toward weight, diet, and physical activity among overweight African American girls. *Preventing Chronic Disease, 5*(2).

Boyle, J. P., Honeycutt, A. A., Narayan, K. V., Hoerger, T. J., Geiss, L. S., Chen, H., et al. (2001). Projection of diabetes burden through 2050. *Diabetes Care, 24*(11), 1936–1940.

Buntain, C., & Golbeck, J. (2015). *This is your Twitter on drugs: Any questions? Proceedings of the 24th international conference on World Wide Web.* ACM777–782.

Coppersmith, G., Dredze, M., Harman, C., & Hollingshead, K. (2015). *From ADHD to SAD: Analyzing the language of mental health on Twitter through self-reported diagnoses. NAACL HLT 2015*1.

Cotman, C. W., & Berchtold, N. C. (2002). Exercise: A behavioral intervention to enhance brain health and plasticity. *Trends in Neurosciences, 25*(6), 295–301.

Culotta, A. (2010). *Towards detecting influenza epidemics by analyzing Twitter messages. Proceedings of the first workshop on social media analytics.* ACM115–122.

Dahl, A. A., Hales, S. B., & Turner-McGrievy, G. M. (2016). Integrating social media into weight loss interventions. *Current Opinion in Psychology, 9,* 11–15.

Eichstaedt, J. C., Schwartz, H. A., Kern, M. L., Park, G., Labarthe, D. R., Merchant, R. M., et al. (2015). Psychological language on Twitter predicts county-level heart disease mortality. *Psychological Science, 26*(2), 159–169.

Flegal, K. M., Carroll, M. D., Kit, B. K., & Ogden, C. L. (2012). Prevalence of obesity and trends in the distribution of body mass index among US adults, 1999–2010. *The Journal of the American Medical Association, 307*(5), 491–497.

Fu, K.-W., Liang, H., Saroha, N., Tse, Z. T. H., Ip, P., & Fung, I. C.-H. (2016). How people react to Zika virus outbreaks on Twitter? A computational content analysis. *American Journal of Infection Control, 44*(12), 1700–1702.

Ghosh, D., & Guha, R. (2013). What are we 'tweeting' about obesity? Mapping tweets with topic modeling and geographic information system. *Cartography and Geographic Information Science, 40*(2), 90–102.

Golder, S. A., & Macy, M. W. (2011). Diurnal and seasonal mood vary with work, sleep, and daylength across diverse cultures. *Science, 333*(6051), 1878–1881.

Hanson, C. L., Cannon, B., Burton, S., & Giraud-Carrier, C. (2013). An exploration of social circles and prescription drug abuse through Twitter. *Journal of Medical Internet Research, 15*(9), e189.

Harris, J. K., Moreland-Russell, S., Tabak, R. G., Ruhr, L. R., & Maier, R. C. (2014). Communication about childhood obesity on Twitter. *American Journal of Public Health, 104*(7), e62–e69.

Harris, J. K., Mueller, N. L., Snider, D., & Haire-Joshu, D. (2013). Peer reviewed: Local health department use of Twitter to disseminate diabetes information, United States. *Preventing Chronic Disease, 10.*

Hartz, A. J., Rupley, D. C., Kalkhoff, R. D., & Rimm, A. A. (1983). Relationship of obesity to diabetes: Influence of obesity level and body fat distribution. *Preventive Medicine, 12*(2), 351–357.

Harvard HSPH (2017). *Obesity trends.* https://www.hsph.harvard.edu/obesity-prevention-source/obesity-trends/.

Hill, J. O., Wyatt, H. R., & Peters, J. C. (2012). Energy balance and obesity. *Circulation, 126*(1), 126–132.

Hong, L., & Davison, B. D. (2010). *Empirical study of topic modeling in Twitter. Proceedings of the first workshop on social media analytics.* ACM80–88.

Karami, A. (2015). *Fuzzy topic modeling for medical corpora (Ph.D. thesis).* Baltimore

County: University of Maryland.

Karami, A., & Gangopadhyay, A. (2014). FFTM: A fuzzy feature transformation method for medical documents. *Proceedings of the conference of the Association for Computational Linguistics (ACL), Vol. 128*.

Karami, A., Gangopadhyay, A., Zhou, B., & Kharrazi, H. (2015a). *FLATM: A fuzzy logic approach topic model for medical documents. Proceedings of the annual meeting of the North American Fuzzy Information Processing Society (NAFIPS)*. IEEE.

Karami, A., Gangopadhyay, A., Zhou, B., & Kharrazi, H. (2015b). A fuzzy approach model for uncovering hidden latent semantic structure in medical text collections. *Proceedings of the iConference*.

Karami, A., Gangopadhyay, A., Zhou, B., & Kharrazi, H. (2017). Fuzzy approach topic discovery in health and medical corpora. *International Journal of Fuzzy Systems, 1–12*.

Karami, A., & Zhou, B. (2015). Online review spam detection by new linguistic features. *iConference 2015 proceedings*.

Karami, A., & Zhou, L. (2014a). Exploiting latent content based features for the detection of static SMS spams. *The 77th annual meeting of the Association for Information Science and Technology (ASIST)*.

Karami, A., & Zhou, L. (2014b). Improving static SMS spam detection by using new content-based features. *The 20th Americas Conference on Information Systems (AMCIS)*.

Kivipelto, M., Ngandu, T., Fratiglioni, L., Viitanen, M., Kåreholt, I., Winblad, B., et al. (2005). Obesity and vascular risk factors at midlife and the risk of dementia and Alzheimer disease. *Archives of Neurology, 62*(10), 1556–1560.

Kopelman, P. G. (2000). Obesity as a medical problem. *Nature, 404*(6778), 635–643.

Laaksonen, D. E., Lindström, J., Lakka, T. A., Eriksson, J. G., Niskanen, L., Wikström, K., et al. (2005). Physical activity in the prevention of type 2 diabetes. *Diabetes, 54*(1), 158–165.

Lampos, V., & Cristianini, N. (2010). *Tracking the flu pandemic by monitoring the social web. 2010 2nd international workshop on cognitive information processing*. IEEE411–416.

Lampos, V., & Cristianini, N. (2012). Nowcasting events from the social web with statistical learning. *ACM Transactions on Intelligent Systems and Technology, 3*(4), 72.

Lampos, V., De Bie, T., & Cristianini, N. (2010). *Flu detector-tracking epidemics on Twitter. Joint European conference on machine learning and knowledge discovery in databases*. Springer599–602.

Lazard, A. J., Scheinfeld, E., Bernhardt, J. M., Wilcox, G. B., & Suran, M. (2015). Detecting themes of public concern: A text mining analysis of the Centers for Disease Control and Prevention's Ebola live Twitter chat. *American Journal of Infection Control, 43*(10), 1109–1111.

Lu, Y., Mei, Q., & Zhai, C. (2011). Investigating task performance of probabilistic topic models: An empirical study of PLSA and LDA. *Information Retrieval, 14*(2), 178–203.

Luchsinger, J. A., Cheng, D., Tang, M. X., Schupf, N., & Mayeux, R. (2012). Central obesity in the elderly is related to late onset Alzheimer's disease. *Alzheimer Disease and Associated Disorders, 26*(2), 101.

Mcauliffe, J. D., & Blei, D. M. (2008). Supervised topic models. *Proceedings of the advances in neural information processing systems, 121–128*.

McCallum, A. K. (2002). *Mallet: A machine learning for language toolkit*.

Mejova, Y., Weber, I., & Macy, M. W. (2015). *Twitter: A digital socioscope*. Cambridge University Press.

Nasukawa, T., & Yi, J. (2003). *Sentiment analysis: Capturing favorability using natural language processing. Proceedings of the 2nd international conference on knowledge capture*. ACM70–77.

Odlum, M., & Yoon, S. (2015). What can we learn about the Ebola outbreak from tweets? *American Journal of Infection Control, 43*(6), 563–571.

Olanoff, D. (2015). *Twitter Monthly Active Users Crawl To 316M, Dorsey: We are not satisfied*. http://apps.who.int/iris/bitstream/10665/204871/1/9789241565257_eng. pdf.

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

Paul, M. J., & Dredze, M. (2012). A model for mining public health topics from Twitter. *Health, 11* 16-6.

PLAY 60 Challenge (2017). *NFL*. http://www.nflrush.com/play60/.

Pokeman-Go Game (2017). *Niantic, Inc*. http://www.pokemongo.com/.

Prier, K. W., Smith, M. S., Giraud-Carrier, C., & Hanson, C. L. (2011). *Identifying health-related topics on Twitter. International conference on social computing, behavioral-cultural modeling, and prediction*. Springer18–25.

Ritterman, J., Osborne, M., & Klein, E. (2009). Using prediction markets and Twitter to predict a swine flu pandemic. *1st international workshop on mining social media, Vol. 9,* 9–17.

Ross, A., & Thomas, S. (2010). The health benefits of yoga and exercise: A review of comparison studies. *The Journal of Alternative and Complementary Medicine, 16*(1), 3–12.

Ross, R., Dagnone, D., Jones, P. J., Smith, H., Paddags, A., Hudson, R., et al. (2000). Reduction in obesity and related comorbid conditions after diet-induced weight loss or exercise-induced weight loss in men. A randomized, controlled trial. *Annals of Internal Medicine, 133*(2), 92–103.

Scanfeld, D., Scanfeld, V., & Larson, E. L. (2010). Dissemination of health information through social networks: Twitter and antibiotics. *American Journal of Infection Control, 38*(3), 182–188.

Schwarzer, R. (2008). Modeling health behavior change: How to predict and modify the adoption and maintenance of health behaviors. *Applied Psychology, 57*(1), 1–29.

Statista (2017). *Number of social media users worldwide from 2010 to 2020*. https://www. statista.com/statistics/278414/number-of-worldwide-social-network-users/.

Szomszor, M., Kostkova, P., & De Quincey, E. (2010). *# swineflu: Twitter predicts swine flu outbreak in 2009. International conference on electronic healthcare*. Springer18–26.

Tompson, T., Benz, J., Agiesta, J., Brewer, K., Bye, L., Reimer, R., et al. (2012). Obesity in the United States: Public perceptions. *The Food Industry, 53*(26), 21.

Tumasjan, A., Sprenger, T. O., Sandner, P. G., & Welpe, I. M. (2010). *Predicting elections with Twitter: What 140 characters reveal about political sentiment. ICWSM 10*.

Turner-McGrievy, G. M., & Beets, M. W. (2015). Tweet for health: Using an online social network to examine temporal trends in weight loss-related posts. *Translational Behavioral Medicine, 5*(2), 160–166.

Twitter (2017). *Twitter developer documentation*. https://dev.twitter.com/docs.

Verdile, G., Keane, K. N., Cruzat, V. F., Medic, S., Sabale, M., Rowles, J., et al. (2015). Inflammation and oxidative stress: The molecular connectivity between insulin resistance, obesity, and Alzheimer's disease. *Mediators of Inflammation, 2015*.

Vickey, T. A., Ginis, K. M., & Dabrowski, M. (2013). Twitter classification model: The ABC of two million fitness tweets. *Translational Behavioral Medicine, 3*(3), 304–311.

Wallach, H. M., Murray, I., Salakhutdinov, R., & Mimno, D. (2009). *Evaluation methods for topic models. Proceedings of the 26th annual international conference on machine learning*. ACM1105–1112.

Wartell, D. (2015). *The geography of obesity: Predicting obesity rates in California based on access to health care*. http://dx.doi.org/10.7910/DVN/LQ4DY6.

Wiebe, J., Breck, E., Buckley, C., Cardie, C., Davis, P., Fraser, B., et al. (2003). *Recognizing and organizing opinions expressed in the world press. New directions in question answering*12–19.

Wing, R. R., Goldstein, M. G., Acton, K. J., Birch, L. L., Jakicic, J. M., Sallis, J. F., et al. (2001). Behavioral science research in diabetes lifestyle changes related to obesity, eating behavior, and physical activity. *Diabetes Care, 24*(1), 117–123.

World Health Organization Fact Sheet (2016). *Obesity and overweight*. http://www.who. int/mediacentre/factsheets/fs311/en/.

Zabin, J., & Jefferies, A. (2008). Social media monitoring and analysis: Generating consumer insights from online conversation. *Aberdeen Group Benchmark Report, 37*(9).

Zhao, W. X., Jiang, J., Weng, J., He, J., Lim, E.-P., Yan, H., et al. (2011). *Comparing Twitter and traditional media using topic models. European conference on information retrieval*. Springer338–349.