TBM

CORRIGENDUM

In the version of "Twitter classification model: the ABC of two million fitness tweets" by Theodore A. Vickey, Kathleen Martin Ginis, John G. Breslin, Maciej Dabrowski (10.1007/s13142-013-0209-0), that is originally published online, one of the authors' name, John G. Breslin, was inadvertently missed.

The error has now been corrected online.

doi: 10.1093/tbm/iby033

© The Author(s) 2018. Published by Oxford University Press on behalf of the Society of Behavioral Medicine. All rights reserved. For permissions, please email: journals.permissions@oup.com

Downloaded from https://academic.oup.com/tbm/advance-article-abstract/doi/10.1093/tbm/iby033/4972655
by National University of Ireland, Galway user
on 18 April 2018

TBM

ORIGINAL RESEARCH

Twitter classification model: the ABC of two million fitness tweets

Theodore A. Vickey,[1] Kathleen Martin Ginis, PhD,[2] Maciej Dabrowski, PhD[1]

[1]National University of Ireland Galway, Galway, Ireland
[2]McMaster University, Hamilton, Canada

Correspondence to: T A Vickey
ted.vickey@gmail.com

Cite this as: *TBM* 2013;3:304–311
doi: 10.1007/s13142-013-0209-0

ABSTRACT
The purpose of this project was to design and test data collection and management tools that can be used to study the use of mobile fitness applications and social networking within the context of physical activity. This project was conducted over a 6-month period and involved collecting publically shared Twitter data from five mobile fitness apps (Nike+, RunKeeper, MyFitnessPal, Endomondo, and dailymile). During that time, over 2.8 million tweets were collected, processed, and categorized using an online tweet collection application and a customized JavaScript. Using the grounded theory, a classification model was developed to categorize and understand the types of information being shared by application users. Our data show that by tracking mobile fitness app hashtags, a wealth of information can be gathered to include but not limited to daily use patterns, exercise frequency, location-based workouts, and overall workout sentiment.

KEYWORDS
Physical activity, Twitter, Mobile fitness apps, Online social network

Implications
Practice: The fitness tweet classification model can be used by researchers to better understand and classify fitness information collected via Twitter.

Policy: A system was developed whereby policy decisions can be made more effectively by the classification of real-time, on-body data collection rather than self-reported measures.

Research: This study provides a research opportunity between health and exercise science and social networking/social software disciplines.

INTRODUCTION
Technology, health, and physical activity
As much as technology has enriched society and expanded global communication, it can be argued that it has also negatively affected overall global health by lowering opportunities for physical activity [1] and by contributing to an overall secular decline in physical activity participation rates [2]. At the same time, research also indicates that there is a potential for technologies to be used as a means for improving health and increasing physical activity [3].

According to a report issued by mobihealthNews, more than 13,000 health and fitness apps were available via iTunes by August 2012 [4]. The use of smartphones in supporting health behavior change via mobile fitness apps is encouraging. Aside from expanded opportunities for users to access health information, mobile devices are becoming more persuasive behavior change tools, allowing for the facilitation of ongoing collection of personal data and the opportune timing of feedback and education to elicit a change in behavior [5]. The most recent health applications have been smartphone applications for personal health areas such as diabetes care, nutrition tracking, smoking cessation, and fitness [4].

The recent advent of smartphones has greatly enhanced both the reach and realm of mobile apps for health purposes by providing a platform for developers to design third-party applications (apps), which expand the functionality and utility of mobile devices [6]. These applications allow users to track their fitness activities via GPS from their smartphones. They also allow the immediate sharing of details about a workout with friends and family that make up one's online community through a website hosted by the app company or by third-party social networks such as Facebook or Twitter. Indeed, simple mobile devices can function as inexpensive, accessible, and powerful triggers for behavior change and may be a particularly powerful mechanism for delivering social support [1].

Online social networking sites are a relatively new and innovative way to deliver social support for physical activity. Online social networking services have eliminated the four walls of brick and mortar found in traditional networking and social interaction [7] and facilitate the development and maintenance of social contacts. One example of a social network is Twitter. The structure of Twitter is simple–users send messages (a.k.a., tweets) to a network of people (a.k.a., followers) from a variety of devices (desktops, laptops, mobile devices, etc.). Tweets are text-based messages of up to 140 characters in length. The default setting for the sharing of tweets is public, which permits other Twitter users to follow and read each other's tweets. Each user has a personalized Twitter home page where all their tweets are aggregated into a single list [8].

Recent academic research has explored the role of the Twitter hashtag–a short keyword, prefixed with the hash symbol "#"–as a means of collecting a distributed discussion between groups of users, who do not need to be connected through existing "follower" networks [9]. One function of Twitter is the ability for information to be shared not only to those who are part of the follower network but also to the entire Twitter population by default. Twitter co-founder Evan Williams suggests "Twitter lets people know what's going on about things they care about instantly, as it happens. In the best cases, Twitter makes people smarter and faster and more efficient" [15]. But with over 400 million tweets sent every day [14], individual tweets can be inane; but taken collectively, analysis of a stream of messages can turn Twitter into a useful tool for solving problems, performing research, and providing insights into the digital moods of its users. Twitter hashtags have been studied to garner information on topics such as terrorism informatics, user modeling and personalization, online security, spam detection, and information streaming [10].

In addition, research has focused on how Twitter is used as a communication platform and understanding *why* and *how* people use online social networks. By understanding the reasons, improvements to the overall structure of the network can occur [11]. From this work, researchers have derived standard metrics for measuring a user's Twitter behavior, such as the number of tweets, retweets, and followers, [12] as well as text classification models to help understand the content of each tweet [13]. Retweets are the forwarding of tweets received by one user to their own personal social network, thus allowing for tremendous "virtual" sharing of information. Twitter followers are fellow Twitter users where one user "follows" the tweets of another.

Research on text classification within Twitter has shown that people use Twitter for different reasons. Java et al. [11] identified four main user intentions on Twitter: (1) Daily Chatter–most posts on Twitter talk about daily routine or what people are currently doing and this is the largest and most common user of Twitter; (2) Conversations–about one-eighth of all posts contain a conversation and this form of communication was used by almost 21 % of users; (3) Sharing information–about 13 % of posts contained a URL (i.e., website address), directing readers to another information source; and (4) Reporting news–many Twitter users report latest news or comment about current events on Twitter. Some automated users or agents post updates like weather reports and new stories from RSS feeds.

Text classification is one of the most important research fields in information retrieval and data mining, and its solutions are at the core of several technology applications ranging from the automatic cataloging of newspaper pages and web pages to the management of incoming e-mails and from the annotation of DNA genome sequences to sentiment analysis of tweets [16]. By tapping into the world's collective brain, researchers have found that efforts to dig through the millions of individual tweet can provide a glimpse into public sentiment and activity and perhaps can even help shape it [15]. To the best of our knowledge, no research to date has conducted text classification within the context of Twitter and physical activity. An understanding of how Twitter is used in physical activity contexts could lead to improvements in the development of mobile fitness apps that promote and support physical activity behavior change.

Thus, the overarching purpose of this study was to develop an understanding of the types of information being shared from mobile fitness apps via Twitter. Our specific objectives were (1) to develop and implement a method for collecting fitness tweets sent from mobile fitness apps, (2) to develop a conceptual model to classify tweets, and (3) to analyze and interpret a sample of tweets. Given the preliminary nature of this research, no hypotheses were put forth.

METHODS AND RESULTS

Development and implementation of a fitness tweet collection method

After an online review of online tools that could collect and manage tweets, an open source program called TwapperKeeper was chosen. TwapperKeeper is a web application designed to archive social media data via Twitter to allow for long-term archival and analysis. The application uses a Twitter-enabled API that acts as an interface between the Twitter search function and a cloud database for tweet storage. The application allows users to monitor and archive specific hashtags and to provide additional metadata to describe an archive that can later be viewed in multiple.

Once the hashtags were defined, the application began two archiving processes (TwapperKeeper, 2011):

- "The Crawl"–For the keyword defined (by a hashtag), the crawling processes began to poll the Twitter Search API to find all tweets in the search cache that match the desired hashtag. This allowed for TwapperKeeper to fill in older tweets (limited by the Twitter API) as well as continually monitor tweets that might be missed by "The Stream" archive process. A disruption of service is possible during disconnects/reconnects with the Twitter Streaming API, rate limits imposed by Twitter, and possible service interruptions on the Twitter service itself.
- "The Stream"–A persistent connection was also created with the Twitter Streaming API for the desired hashtags. The archiving process inserted all inbound tweets into a database table for later processing. A second process ran to analyze each

tweet in the table and moved the tweets into the proper archive table.

TwapperKeeper was installed on a cloud server and began collecting tweets in March 2011 from five mobile fitness apps.

Of the thousands of mobile fitness apps available, we chose five for analysis based on their availability via iPhone, the ability of the mobile fitness app to share workout information through Twitter, and also apps that represented larger versus smaller corporations, based in the USA versus abroad, and targeting beginner versus experienced exercisers. We used these criteria to narrow possible choices, reviewed additional academic research for previously used applications, researched publicly available reviews on different mobile fitness apps, interviewed both developers and users of mobile fitness apps to obtain their input, and met as a group to finalize the selected mobile fitness apps to study. The five apps chosen were Endomondo, MyFitnessPal, Nike+, RunKeeper, and dailymile.

We then collected tweets from the five mobile fitness apps by gathering tweets that used the following hashtags: #endomondo, #myfitnesspal, #nikeplus, #runkeeper, and #dailymile. These are the hashtags that the apps automatically attach to a tweet to indicate it has come from that particular application. It is through these hashtags that common themes or information can be grouped within Twitter. Tweet collection was done by TwapperKeeper which began the archiving process by searching publicly available tweets, identifying tweets that contained the desired hashtags, and inserting identified tweets into a database for later processing. The type of information collected from each tweet is shown in Table 1.

In addition, supplementary information was collected regarding demographics and Twitter usage. To collect and process such information, a JavaScript was created. The JavaScript extracted Twitter information collected from the TwapperKeeper database and requested specific information about the publicly available Twitter user account. After limiting the process to unique users, the script has the ability to send Twitter user information to websites such as Twitter, Klout, and other information websites to collect general information about the Twitter users such as their start date on Twitter, number of total sent tweets, their Klout score, and frequency of tweets.

The information collected about the Twitter user is shown in Table 2. All information collected was publicly available with each user of Twitter agreeing to this public sharing of information by their agreement to the terms and conditions of their Twitter account.

Development of a conceptual model to classify fitness tweets

Our second research objective was to develop and validate a strategy to classify the collected tweets. This strategy was based on two fields of exploration within the Twitter research tracks that seek to better understand the context of tweets: data mining and text classification. Data mining is a relatively young and interdisciplinary field of computer science with the process that results in the discovery of new patterns in large datasets by using methods at the intersection of artificial intelligence, machine learning, statistics, and database systems, with the overall goal being to extract knowledge from an existing dataset and transform it into a human-understandable structure for further use [17]. Text classification is the labeling of natural language texts (in this case, a tweet) into one or more categories drawn from a predefined set. This may be done manually or algorithmically. For the purposes of this research, the complete dataset was sorted and evaluated manually to determine any apparent similarities. This evaluation allowed the algorithms to be established that were then used to classify the entire database of fitness tweets. Data verification tests

Table 1 | Collected Twitter data point descriptions

Data point	Description
Archive source	Twitter Search or Twitter Stream
Text	The actual tweet
To_User	Name of recipient user if the tweet was sent to a specific Twitter user
From_User	Name of the Twitter user that sent the tweet
ID	Specific Twitter identification number for the associated tweet
From_User_ID	Specific Twitter identification number for the associated Twitter user name that sent tweet
Iso_Language_Code	Identified language of the tweet
Source	Twitter platform used to send tweet
Profile_Img_URL	URL to the picture of the tweeter
Geo_Type	Either "point" if geolocation was used with tweet or blank if not
Geo_Coordinates_0	Latitude of the location where the tweet was sent
Geo_Coordinates_1	Longitude of the location where the tweet was sent
Created_At	Day, date, and time the tweet was sent
Time	UNIX time the tweet was sent

Table 2 | Additional Twitter data point descriptions from the Fitness Tweet Crawler

Data point	Description
User_Name	Twitter user name of the person that sent the tweet (same user name as the From_User listed in TwapperKeeper)
Location	Location of the tweeter as recorded in their Twitter user profile
Tweets	Number of total tweets sent by user at the time of the query
Following	Number of people the user is following at the time of the query
Followers	Number of people that follow the user
Klout	Klout score of the user
Style	Klout style classification of the user
Access_Date	Date of the query
Access_Time	Time of the query
Twitter_Startdate	Date the user started using Twitter
Times_Per_Day	Number of times per day the user sends any tweet

were run throughout the process to ensure proper automatic classifications. Enhancements to the algorithms and reclassification of the entire database were conducted as needed.

The development of the fitness tweet classification model was based on available macro topic classification models where tweets were categories into broad categories of content, based on prior literature [18], and sourced from other works [8, 11, 13, 19]. These prior literatures indicated four major categories for sharing on Twitter–Conversational, Pass-Along, News, and Status–categories that are consistent with Java's [11] research on the primary purposes of tweets. These four categories provided the starting point for the development of our framework. Once the theoretical foundation was established, a custom computer program was created that incorporated data mining and text classification of the collected tweets.

Consistent with previous research [13], we used a grounded theory approach to develop the framework. The grounded theory research approach is opposite to other traditional social science research where the researcher chooses a theoretical framework and then applies the model to the phenomenon to be studied. Rather than beginning with a hypothesis, the grounded theory starts with data collection. From the collected data, key points are marked with a series of codes, which are extracted from the text. These codes are then grouped into similar concepts, making them more workable. From these concepts, categories are formed, which are the basis for the creation of a theory or a reverse-engineered hypothesis [20].

Development of the framework began by taking a random sample of 500 public tweets (100 from each of the five apps), tweeted over a 2-week period. The researchers sorted the tweets into groups with two general themes emerging: tweets about a recent workout (i.e., "Activity") and tweets about other non-exercise-related conversational topics (i.e., "Conversation"). Tweets that share a person's workout, specific to the tweet structure as defined by the five different mobile fitness apps, were classified as Activity. Each mobile fitness app used a different data structure that was able to be defined. Table 2 provides examples within the Activity category. These 500 tweets were then submitted to a computerized text classification procedure programmed to identify Activity and Conversation tweets. This procedure revealed subcategories within the Activity and Conversation groupings as well as a third category, subsequently labeled "Blarney."

Specifically, further analysis of the Activity tweets showed that some users added additional messages along with the information about their actual workout (e.g., I just ran 4 mi using #RunKeeper in the sunshine of San Diego, felt great); thus, the "Workout Plus" subcategory was added. A Workout Plus tweet has the same foundation of a Workout tweet but adds the additional variable of information.

Further analysis of the Conversation category indicated tweets pertaining to four areas: requests for technical support (requests to the app company or the broader community), marketing (e.g., press releases and updates that came from the app company itself or the community), statements of support (where people within the app community congratulated others on reaching milestones, personal bests, etc.), or information sharing (e.g., those within the app community that wanted to run together in an upcoming 10-km race would post messages using the hashtag per the app). Thus, the following subcategories were added to the Conversation category: Technical Support, Marketing, Statements of Support, and Information Sharing. In addition, a new third category was added (Blarney) that tagged spam tweets (tweets with only a URL) or tweets that had little relevance to exercise (e.g., Test FB http://t.co/IKIQjTi #myfitnesspal). Blarney is defined as skillful flattery, nonsense, or blandishment [22]. Tweets that were classified as Blarney fit the same definition but were further classified into Pointless Babble or Spam. Any tweet that contained just a URL link or appeared to be an unsolicited commercial tweet was classified as Spam, and all other tweets classified as Blarney were subclassified as Pointless Babble. Table 2 provides examples of Pointless Babble and Spam within the Blarney category. The full classification framework is shown in Fig. 1.

To examine the strength of the framework, six coders were given the same 500 tweets previously used and were asked to classify them according to the framework. Examples of messages placed into each category are shown in Table 3. Agreement among raters was high. An intraclass correlation coefficient, using the two-way mixed-effects model, yielded an ICC=.925 (95 % confidence interval=.914–.935). With the general recommendation for reliability being .7 [21], reliability among the raters was high, providing strong evidence that the framework could be reliably used to classify fitness tweets. This framework was then used to generate computer code for computerized classification of the tweets, as reported in the next section.

Analysis and interpretation of a sample of tweets

The third objective was to provide an analysis of tweets using our classification framework. Data collection using TwapperKeeper began on Thursday April 21, 2011 at 00:00 Greenwich mean time and continued until September 21, 2011 at 23:59 for a total collection of Twitter data of 184 days. During this period, 2,856,534 tweets were collected in 23 different languages. However, for the purposes of this study, only the English tweets were analyzed. After reviewing the human classification of the tweets, minor adjustments to the code enhanced the overall reliability of the computer classification of the tweets.

The total number of processed tweets in English was 1,982,653, which were tweeted by 165,768 unique users. Figure 2 displays the breakdown of the fitness tweets categorization using the fitness tweet classification model. Of the English language tweets, 1,446,462 (73 %) were classified as Activity, 104,360 (5 %) as Blarney, and 420,603 (21 %) as Conversation. Of additional interest is the subclassification breakdown. Figure 3 displays the breakdown of fitness tweets in the subclassification. Of the Activity tweets, 53 % of the total tweets were Workout, with 21 % as Workout Plus. There was a small sample of Blarney tweets with .1 % Pointless Babble and 5.2 % Spam tweets from the total dataset of tweets. Of Conversation tweets, .4 % was of Technical Support, .5 % was of Corporate Marketing, 1.3 % was of Statements of Support, and 19 % was of Information Sharing relative to the total number of tweets in the dataset. Of the Activity fitness tweets, over 76,192,059 min of exercise was shared via the five mobile apps via Twitter equaling over 145 years of physical activity.

DISCUSSION

Using the data collection and data processing tools described in this paper, we have been able to create a growing dataset of information that people publically share from their smartphones and other devices, via Twitter, about their workout activities. This information includes data collected by the app itself–such as exercise type, length, day of the week, mood, geographical location, and time–as well as data on how people use fitness apps to share information and engage in social networking regarding their fitness activities. Together, this information can facilitate research on how technology can be used to monitor and motivate physical activity and how online social networks may play a role in physical activity promotion and adherence.

We have created a Twitter classification model that allows for analysis of mobile fitness app tweets through data collection, data processing, and data

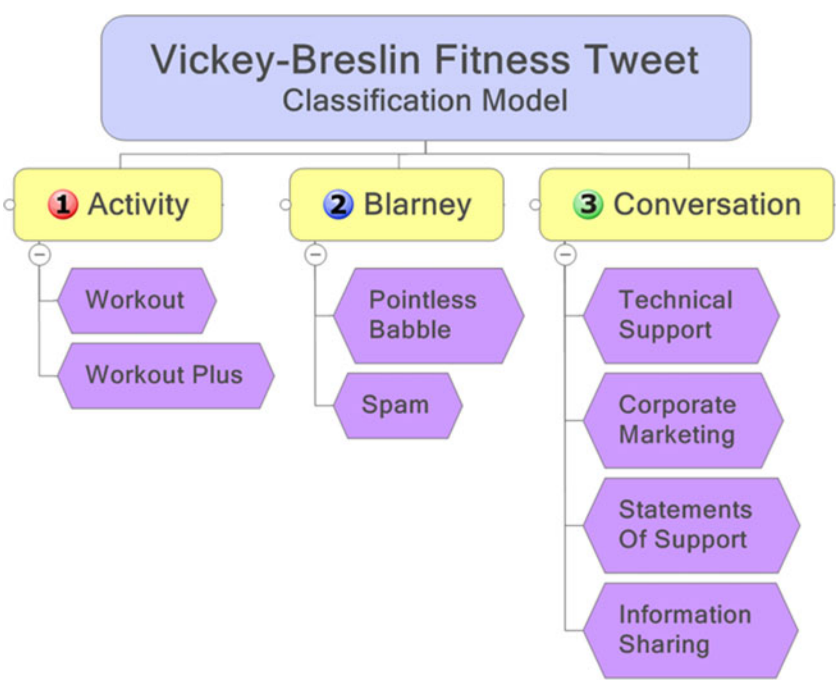

Fig 1 | Fitness tweet classification model

Table 3 | Examples of fitness tweets by mobile fitness app

		dailymile	Endomondo	RunKeeper	Nike+	MyFitnessPal
Activity	Workout	Ran.http://dmile.com/e/Ok7N	Was out cycling 0.32 km with #Endomondo. See it here: http://bit.ly/hNKF3y	Just completed a 5.69 km walk with @runkeeper. Check it out! http://mkpr.com/ahr0yw #RunKeeper	I just finished a 2.00 mi run with a time of 20:05 with Nike+GPS. #nikeplus	burned 157 calories doing 60 minute of "Yoga" #myfitnesspal
	Workout Plus	Did a cross training workout for 45 mins and felt good. Morning upper: 2 × (15 L-pullups +25 bicep curl http://dmile.com/e/Olsl	Was out running 4.4 miles with #Endomondo. Had PB 2nd mile. See it here: http://bit.ly/dRLrgn	Just posted a 2.25 mi run - just some miles to stretch the legs before the race tomorrow. http://mkpr.com/ahm649 #RunKeeper	9.69 miles this morning. Longest run since the move to Colorado. #nikeplus #fb	
Blarney	Pointless Babble	Test #dailymile	Rated on LUUUX http://bit.ly/ohjo71	Other Activity 0.00 km | RunKeeper http://t.co/24Oy6Mvj via @addthis	#nikeplus software is crap. Runs far too slow?	Test FB http://t.co/IKIOjTi #myfitnesspal
	Spam	#dailymile http://bit.ly/iJh2Wn	http://bit.ly/bhtVdB #endomondo	@0591R Congrats! Your RunKeeper Rank has been updated! http://t.co/ymRjs0b #WorldRankin	I just finished not caring w/a time of 2.5 seconds using #nikeplus	MUST SEE! http://bit.ly/bhtVdB #myfitnesspal
Conversation	Technical Support	@aReyo we'll look into that! We do have a way to export all your data.	#Endomondo now supports #Facebook integration for your facebook apps: http://bit.ly/kZ9SyK	@bshermcincy strange it's buggy on runkeeper. Like I said, it's perfect on my iphone4.	@northone A good place to start when you experience weird iPod behavior is with a reset	@MyFitnessPal can the BB app scan barcodes?
	Corporate Marketing	Couch to 5 k Coach+GPS Run Tracker with DailyMile for iPhone on Sale ($1.99) http://goo.gl/fb/G	#Endomondo Celebrates 5 Million #Downloads Of Fitness App With $2.3 M In Seed Funding http://t.co/Z3a2MYwV	The iPhone App RunKeeper allows endurance athletes to track http://bit.ly/9w2kZl	Make your miles count! 200,000 miles=$200,000 towards relief efforts in Japan. http://bit.ly/all4japan #SupportJapan#nikeplus	Count your calories on the go with Calorie Counter by MyFitnessPal - http://bit.ly/lLbyPz
	Statements of Support	@dailymile I can do it. Only so many miles to go!	:) Yay to moving my lazy ass! RT @ishsal: Was out running 4.04 km with #Endomondo. See it here: http://bit.ly/ogZ4	Achieved a new personal record with @RunKeeper: Farthest distance... http://bit.ly/hy8xUz #FitnessAlerts	Good job! RT @Storm21 I just finished a 3.51 mi run with a time of 36:29 with Nike+GPS. #nikeplus	#myfitnesspal New Member Struggle with losing weight and keeping it off http://bit.ly/hx2hbZ
	Information Sharing	#dailymission Too many of us focus on the training, what about core? I work on core everyday. It's what I... http://dailymile.com/e/QhSH	My last Endomondo was walking, not cycling. I forgot to change.	Watch my bike ride right now with @RunKeeper Live http://mkpr.com/ahstp9 #RKLive #RunKeeper	RT @sneakernoize: Reading Material For the Sneaker Fiend Kicksclusive Magazine & http://bit.ly/gsc4ZL #nikeplus	lost 4 pounds since his last weigh-in! He's lost 28.8 pounds so far. #myfitnesspal

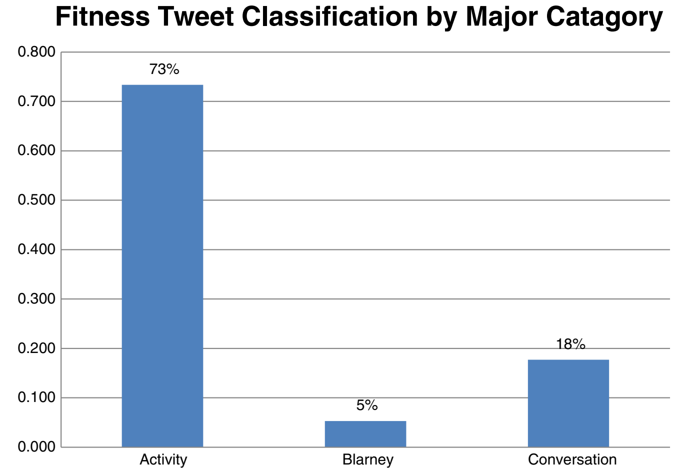

Fig 2 | Fitness tweet classification

analysis. We have shown that a simple 140-character Twitter message can lead to a wealth of pertinent and valuable demographic- and action-oriented information that when processed through the fitness tweet classification model can show patterns of associations between user, online fitness communities, and Twitter as a whole. Our research extends the Twitter classification models established by Naaman et al. [13] who created models within the context of how to establish categorization of tweets. Our classification model was created using a grounded theory approach and its reliability was confirmed across six raters who successfully coded 500 sample tweets. Given the breadth of mobile fitness apps included in our analysis, we are confident that the classification model can be applied to categorize data obtained from other mobile fitness apps that have the ability to share information via Twitter. Indeed, an important contribution of this project is the identification of data structures from within mobile fitness applications when sharing via Twitter. These structures can now be analyzed using text classification processes. Given the tremendous amount of data generated by Twitter (e.g., we have collected over 12 million tweets from just five mobile apps over the past 15 months), researchers need tools to manage and analyze these data in order to address research questions regarding the use of technology and social networks to promote health behavior change. Our work has yielded such tools.

We have also provided preliminary data on how people are engaging with their online social communities to share information on their fitness activities. While there is a substantial amount of information being shared via Twitter regarding actual workouts (i.e., Activity tweets), there appears to be only a small amount of conversation about the workout themselves (Conversation tweets). For those users who track their workouts for internal quantified self reasons, our data would indicate that the mobile fitness apps can provide such a tool. However, it is unclear regarding the reason

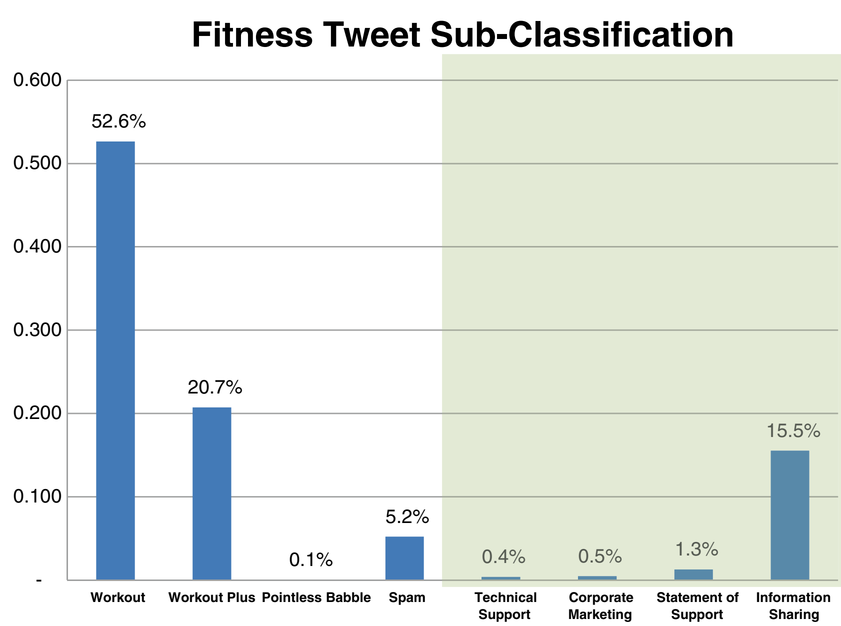

Fig 3 | Fitness tweet subclassification

why these people would decide to share their workouts on a social networking service such as Twitter. The lack of meaningful and engaged conversation between mobile fitness app users is an important question to be addressed for future research. If it is the intent of mobile fitness app developers who have the objective of using these apps to increase physical activity behavior by way of strengthening social support via one's social network, then having a true understanding of why conversation is not occurring is critical. However, our findings would indicate similar usage patterns for general Twitter usage. General Twitter research suggests that Twitter is used more as a one-way, one-to-many publishing service than a two-way, peer-to-peer communication network [23].

In summary, this study has provided tools for advancing research on mobile fitness app use, social networking, and physical activity. With these tools available, researchers can now examine a wide range of questions such as how the use of mobile fitness apps and the sharing of workout information using Twitter is related to possible exercise motivation within one's social network, how mobile fitness apps are related to the possible influence of social support by using Twitter, and what mobile fitness app sharing features are most appropriate with regard to using technology to impact physical inactivity. Addressing these issues will lay the foundation for understanding and potentially improving the role that technology and social networking can play in improving health and fitness behavior.

1. Foster D, Linehan C, Kirman B, et al. Motivating physical activity at work: using persuasive social media for competitive step counting. *14th MindTrek Conference*, 6th–8th October, Tampere, Finland.
2. Nigg CR. Technology's influence on physical activity and exercise science: the present and the future. *Psychol Sport Exerc*. 2003;4(1):57-65.
3. Ranck J. *Connected Health: How Mobile Phones, Cloud and Big Data Will Reinvent Healthcare*. San Francisco, CA: GigaOM Books; 2012.
4. Dolan B. 13K iPhone consumer health apps in 2012. *mobihealthNews*. 2012. http://mobihealthnews.com/13368/report-13k-iphone-consumer-health-apps-in-2012/. Accessed 24 Aug 2012.
5. Patrick K, Intille SS, Zabinski MF. An ecological framework for cancer communication: implications for research. *J Med Internet Res*. 2005;7(3):e23.
6. West JH, Hall PC, Hanson CL, et al. There's an app for that: content analysis of paid health and fitness apps. *J Med Internet Res*. 2012;14(3):e72.
7. Vickey T. *Social Capital and the Role of LinkedIn to Form, Develop and Maintain Irish Entrepreneurial Business Networks*. Newcastle upon Tyne, UK: Cambridge Scholars Publishing; 2011. http://www.c-s-p.org/flyers/Social-Capital-and-the-Role-of-LinkedIn-to-Form–Develop-and-Maintain-Irish-Entrepreneurial-Business1-4438-2904-8.htm. Accessed 24 Aug 2012.
8. Jansen BJ, Zhang M. Twitter power: tweets as electronic word of mouth. *J Am Soc Inf Sci*. 2009;60(11):2169-2188.
9. Bruns A, Burgess J. The use of Twitter hashtags in the formation of ad hoc publics. *6th Eur Consort Polit Res Gen Conf*. 2011;64:1-9.
10. Cheong M, Ray S. A literature review of recent microblogging developments. Technical report. (p. 43). Victoria, Australia, 2001. http://www.csse.monash.edu.au/publications/2011/tr-2011-263-full.pdf. Accessed 27 Aug 2012.
11. Java A, Song X, Finin T. Why we twitter: understanding microblogging usage and communities. *Joint 9th WEBKDD and 1st SNA-KDD Workshop '07*. 2007. http://portal.acm.org/citation.cfm?id=1348556. Accessed 21 Aug 2012.
12. Vega E, Parthasarathy R. Where are my tweeps?: Twitter usage at conferences. *Paper*, *Personal Information*. 2010; pp 1–6. http://www.socialcouch.com/demos/final_paper_twitter.pdf. Accessed 5 July 2012.
13. Naaman M, Boase J, Lai C. Is it really about me? Message content in social awareness streams. In: *CSCW '10 Proceedings of the 2010 ACM conference on Computer supported cooperative work*. New York: NY: ACM; 2010;189-192.
14. Bennett S. Twitter now seeing 400 million tweets per day, increased mobile ad revenue, says CEO. *All Twitter—The Unofficial Twitter Resource*. 2012. http://www.mediabistro.com/alltwitter/twitter-400-million-tweets_b23744. Accessed 3 Sep 2012.
15. Miller C. Finding utility in the jumble of tweeted thoughts. *New York Times*. 2009. http://www.nytimes.com/2009/04/14/technology/internet/14twitter.html?pagewanted=all. Accessed 3 Sep 2012.
16. Vitale D, Ferragina P, Scaiella, M. Classification of short texts by deploying topical annotations. In: Baeza-Yates, R, de Vries, AP, Zaragoza, H, Cambazoglu, BB, Murdock, V, Lempel, R, Silvestri, F, eds. *Advances in Information Retrieval: 34th European Conferences on IR Research ECIR 2012, Barcelona, Spain, April 1-5, 2012, Proceedings*. Vol. 7224. New York, NY: Springer; 2012:376-387.
17. Chakrabarti S, Ester M, Fayyad U, et al. Data mining curriculum: a proposal. *Intensive Working Group of ACM SIGKDD Curriculum Committee*. 2006.
18. Dann S. Twitter content classification. *First Monday*. 2010;*15*(12): 1–11. http://firstmonday.org/htbin/cgiwrap/bin/ojs/index.php/fm/article/view/2745/2681. Accessed 24 August 2012.
19. Kelly R. *Twitter Study—August 2009 Introduction*. San Antonio, TX: Pear Analytics; 2009:2011.
20. Corbin J, Strauss A. Grounded theory method: procedures, canons, and evaluative criteria. *Qual Sociol*. 1990;13:3-21.
21. Shrout PE, Fleiss JL. Intraclass correlations: uses in assessing rater reliability. *Psychol Bull*. 1979;86:420-3428.
22. Merriam-Webster. Blarney—definition and more from the Merriam-Webster Free Dictionary. *Merriam-Webster Free Dictionary*. 2012.
23. Heil, B, Piskorski, M. HBR Blog Network: New Twitter research: men follow men and nobody tweets. Harvard Business Review (June 1, 2009). Cambridge, MA: Harvard Business Publishing; 2009. http://blogs.hbr.org/cs/2009/06/new_twitter_research_men_follo.html. Accessed 6 June 2012.