

An Analysis of Exercising Behavior in Online Populations

David Jurgens, James McCorrison, Derek Ruths

jurgens@cs.mcgill.ca, james.mccorrison@mail.mcgill.ca,

derek.ruths@mcgill.ca

School of Computer Science

McGill University

Abstract

Exercise plays a central role in many peoples' fitness goals. While prior work has examined how individuals pursue these health and fitness goals on general purpose platforms such as Twitter, the lack of precise activity recording has limited detailed analyses of individual and group behaviors. In this study, we explore a recent social media platform dedicated to exercise and use nearly four years of longitudinal exercising history of over 188,000 users to discover large-scale exercising patterns corresponding to different motivations such as sports or general fitness.

1 Introduction

Social media provides individuals with new ways to achieve their health-related goals, including tracking activities and discussing fitness-related topics with like-minded individuals (Ma, Chen, and Xiao 2010; Kendall et al. 2011; Teodoro and Naaman 2013). As larger segments of the population engage in social media, their online records of exercising activities provide a rich source of data for how different groups seek to attain their goals.

Growing concerns about obesity, heart disease, and other phenomena have brought questions about population health to the forefront. A major challenge, though, involves getting an accurate picture of the habits of a population. Previous work on the fitness behaviors of online populations has largely focused on Twitter, assessing the motivations for why users engage in fitness communities (Teodoro and Naaman 2013) and what content they post (Kendall et al. 2011). However, precisely analyzing fitness behaviors in general platforms such as Twitter has been limited due to the significant noise in measuring what activities were performed and the lack of reliable demographic information.

This paper presents a first study of Fitocracy, a popular social media site focused on fitness, where users post their exercise history and engage in social activities. Using a dataset of over 3M workouts from 188K individuals over nearly four years, we are for the first time able to characterize the different trends in exercising behavior and how those trends vary within a population. We demonstrate that individuals'

exercising habits follow meaningful patterns that match different motivations, such as practicing a sport, casual exercise, or strength training. Furthermore, these patterns reveal strong demographic biases towards male or female, and surprisingly, in some cases, towards height and age. Together, these results suggest that Fitocracy is a data-rich platform for precisely quantifying exercise-related phenomena and will prove valuable for future studies that analyze the relationships between exercise, motivation, and social engagement.

2 Related Work

While no work has yet analyzed the Fitocracy platform or large-scale patterns in exercising behavior, three bodies of work in social media relate to our study. Most related to our work are the study of Teodoro and Naaman (2013), who analyzed the motivations for users engaging in the Twitter health and fitness communities and the study of Kendall et al. (2011), who manually analyzed exercise-related posts on Twitter, finding that with targeted keyword searches, it is possible to identify posts of individuals' plans for exercise or evidence of having completed exercise, though upwards of 53% of exercises-related posts were advertisements.

Second, several works have examined the social structure and behavior patterns of topically-focused platforms similar to Fitocracy, such as youbemom (Schoenebeck 2013) and Pintrest (Ottoni, Pesce, and Las Casas 2013). Our analysis of Fitocracy differs from these works due to the prevalence of the demographic information self-reported by users and our targeted focus on exercise behavior, which is not seen on these platforms. Most related is the study of Ma, Chen, and Xiao (2010), who analyzed the weight-loss focused platform FatSecret, which tracks a user's weight, diet, and exercises. Their analysis focused primarily on social factors for weight-change and not exercise, unlike the present study which examines individual users' exercise behavior.

Third, Fitocracy includes several gamification elements, which potentially encourage healthy behaviors. Multiple works have analyzed users' behaviors and motivations in pedometer-based health games (Xu et al. 2012; Walsh and Golbeck 2014). However, these works have largely focused on social factors for behavior, which are not considered here, and due to the constraints of the games, are focused entirely on walking, whereas our behavioral analyses examines individuals' activity patterns across over a thousand exercises.

3 Fitocracy

Fitocracy is a social networking website targeting individuals interested in exercising and fitness. Users create an account and optionally report their age, gender, and height, with 91% reporting at least two attributes. Users record workouts by selecting from a predefined list of 1,090 exercises, each of which allows them to record its relevant aspects, such as its duration, number of repetitions, or the amount of weight lifted. The requirement of selecting exercises from a predefined list enables precisely measuring the activities that were performed, which is often not possible in general-purpose social media or in exercise-related forum posts. Fitocracy includes gamification elements where users receive points for recording workouts, which contributes to their level, and users are awarded badges upon completing certain exercises or engaging in specific social behaviors. Beyond recording workouts, users may post status updates, comment on others' workouts and statuses, and participate in social groups, which are commonly based on sports, diet, lifestyle, or location.

Dataset All of a user's activities, profile, social data, and group memberships on Fitocracy were crawled over a six month period. As Fitocracy does not have a public API, accessing each type of information took different amounts of time, leading to slower collection rates for activities and group memberships. Ultimately, the complete profiles and workout histories of 441,034 users were collected. We note that users engage in the platform in different ways, including only partaking in social functions and commenting rather than posting workouts. As a result, the historical activity data is the result of crawling over 441,034 users, some of whom never recorded any exercise activity but were potentially nonetheless engaged on the site through commenting and other social activities. Ultimately, 188,265 users recorded at least one workout, with the total dataset comprising 3,130,276 workouts (14.3M activities) over nearly a four year span from February 2011 to January 2015.

Examining the demographic information provided with accounts, the Fitocracy userbase is roughly balanced between genders with slightly more women than men but is skewed towards a younger population. Figures 1a and 1b show the number of individuals for each height and age. A small number of individuals (2%) reported themselves either (a) younger than 15 or older than 65 or (b) below four feet or above eight feet, both of which we interpret as inaccurate and exclude from the plots. The mean reported height for females (64.82in) and males (70.52in) is above those reported for the general population in the United States of 63.8in for females and 69.4in for males (McDowell et al. 2008). However, height estimates are known to be higher when self-reported (Merrill and Richardson 2009).

Fitocracy users engaged in a variety of activities. Figures 1c and 1d show the distribution of number of recorded workouts and number of unique exercises performed for all users. Overall, men recorded slightly more workouts than women and were more likely to perform a wider variety of exercises.

While exercising frequency was similar for both genders, the activities performed by different age groups provide a

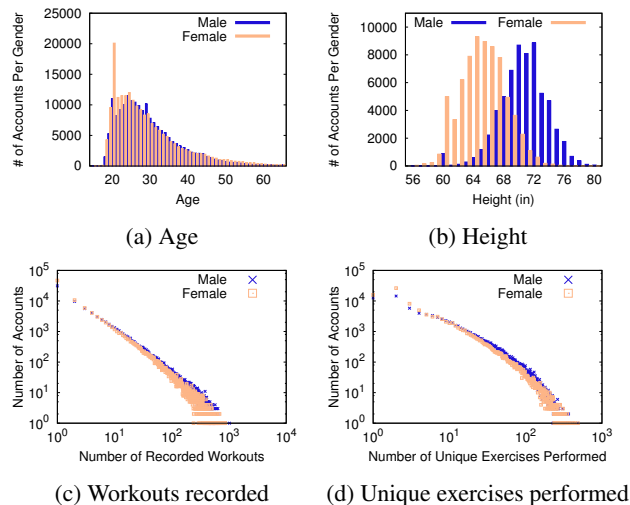


Figure 1: Distributions of users' demographics and exercising frequency, across gender.

insightful view into how behavior changes over time. We divided users by gender and into seven age ranges. Within each gender-age cohort, we computed the probability that a user in that cohort records each exercise and then sorted all exercises according to their average probability of being performed. Table 1 shows the ten exercises that are most likely to be performed by an individual from each cohort. Female exercise selection was highly similar across age groups, with an emphasis on light cardiovascular activities. In contrast, while male exercise selection is initially heavily biased towards strength-training exercises, older cohorts increasingly prefer light cardiovascular exercises. Indeed, for the 50+ cohort, the lists of the ten most-probable exercises for both genders overlapped by nine exercises, only differing in their ranking. This finding reveals that exercising behavior in the general population converges with age and that male habits become more similar to female, which were observed to be relative stable across time.

4 Inferring Fitness Behaviors

Individuals engage in a variety of exercises, as shown in Figure 1d, often as a result of different factors, such as practicing a sport, accessibility to exercising facilities, pursuing health goals, or even latent biological factors. We aim to distill the exercise activities seen in the Fitocracy population into coherent behaviors reflecting the different goals or exercising trends for that subpopulation.

Behavior Model The fundamental assumption behind our behavioral model is that users engage in only a few types of exercise behaviors and that each behavior itself is focused on performing a few types of exercises. Therefore, to infer the latent exercising behaviors, we model the behaviors using Latent Dirichlet Allocation (LDA) (Blei, Ng, and Jordan 2003). Here, each behavior is treated as a multinomial over the exercises, reflecting which are likely to be performed by a practitioner of that behavior. In turn, users are modeled as a

| Age: 0-20 | 20-25 | 25-30 | 30-35 | 35-40 | 40-50 | 50+ |
|---------------------|----------------------|----------------------|---------------------|----------------------|----------------------|---------------------------|
| Walking | Walking | Walking | Walking | Walking | Walking | Walking |
| Running | Running | Running | Running | Running | Running | Walking (treadmill) |
| Crunch | Elliptical Trainer | Elliptical Trainer | Elliptical Trainer | Elliptical Trainer | Elliptical Trainer | Running |
| Plank | Crunch | Running (treadmill) | Running (treadmill) | Running (treadmill) | Walking (treadmill) | Elliptical Trainer |
| Body Weight Squat | Running (treadmill) | Push-Up | Push-Up | Walking (treadmill) | Running (treadmill) | Cycling (stationary) |
| Push-Up | Push-Up | Plank | Plank | Push-Up | Push-Up | Walked up the stairs |
| Elliptical Trainer | Plank | Crunch | Crunch | Body Weight Squat | Plank | Cycling |
| Running (treadmill) | Body Weight Squat | Body Weight Squat | Body Weight Squat | Plank | Cycling (stationary) | Swimming |
| Jumping Jacks | Cycling (stationary) | Cycling (stationary) | Walking (treadmill) | Crunch | Crunch | Running (treadmill) |
| Sit-Up | Stretching | Stretching | Cycling | Cycling (stationary) | Stretching | Push-Up |
| Push-Up | Push-Up | Running | Running | Running | Running | Walking |
| Barbell Squat | Running | Push-Up | Push-Up | Walking | Walking | Running |
| Barbell Bench Press | Barbell Squat | Barbell Squat | Walking | Push-Up | Push-Up | Cycling |
| Running | Barbell Bench Press | Barbell Bench Press | Barbell Squat | Running (treadmill) | Cycling | Push-Up |
| Barbell Deadlift | Barbell Deadlift | Walking | Barbell Bench Press | Cycling | Running (treadmill) | Elliptical Trainer |
| Pull-Up | Pull-Up | Barbell Deadlift | Running (treadmill) | Barbell Bench Press | Elliptical Trainer | Cycling (stationary) |
| Dumbbell Bicep Curl | Dumbbell Bicep Curl | Pull-Up | Cycling | Elliptical Trainer | Cycling (stationary) | Running (treadmill) |
| Walking | Walking | Running (treadmill) | Elliptical Trainer | Barbell Squat | Barbell Bench Press | Walking (treadmill) |
| Crunch | Crunch | Cycling | Barbell Deadlift | Cycling (stationary) | Barbell Squat | Light Walking (secondary) |
| Overhead Press | Sit-Up | Dumbbell Bicep Curl | Pull-Up | Pull-Up | Crunch | Walked up the stairs |

Table 1: The ten exercises most likely to be performed by an individual in each age cohort for women (top) and men (bottom).

sparse multinomial over behaviors, reflecting those in which they take part. In the generative explanation of a model, a user’s exercise history can be interpreted as the user repeatedly sampling a behavior from those they want to perform and then sampling exercises from that behavior.

A key point in this model is what constitutes the observable data from which the behavior and user multinomials are inferred (i.e., the data analogous to terms in a text-based LDA model). Ultimately, a user was modeled as the frequency distribution over the exercises they performed, independent of how an exercise was performed (e.g., its duration, repetitions, etc.), which resulted in highly-interpretable models. Several other alternatives for representing user activity were tested, such as encoding the number of sets done per exercise; however, these alternatives significantly increased data sparsity, leading to lower-quality behaviors.

Setup The LDA behavior model was inferred from the complete workout history of 188,265 individuals. The behavior model requires specifying the number of behaviors to be learned. Multiple models were built for 10 to 100 behaviors and an analysis by experts showed that all models produced highly salient behaviors capturing different behavior aspects. Here, we present results for 20 behaviors in order to comment on the high-level trends seen in the data but full results for all models are made available online.

Results The model identified multiple types of exercising behaviors characteristic of different fitness objectives that were not reflected in the most probable exercises for the entire population (cf. Table 1). Figure 2 illustrates six behaviors identified by the model and their most-probable exercises, with corresponding demographic statistics for each in Table 2. Demographic information was created for each behavior by aggregating individuals’ demographic information relative to the normalized probability mass assigned to that behavior for each user, thereby indicating the population segment likely to engage in the behavior.

The six behaviors in Figure 2 differ significantly in their population demographics. Behavior A shows exercises typical of an individual competing in the sport of CrossFit,

which is one of the most-practiced sports on the site. Notably, the population is almost evenly split between genders, which mirrors the known gender inclusivity of the sport (Knapp 2014). Behavior B offers another view of a gender balanced behavior, with slightly more males than females. Surprisingly, the most-probable activities for this behavior contradict perceived gender norms by including exercises such as dumbbell curls and dumbbell bench press that are typically associated with males. Examining the full list of exercises reveals many exercises use dumbbells, which are often available in a variety of weights at most facilities, making it easy for any individual to adapt the exercise’s weight to their strength level, in contrast to barbell-based exercises that, due to the barbell, have a higher minimum weight to perform. Thus, we posit that Behavior B reflect a adaptable strength-training behavior for any individual, independent of their strength level. Behaviors C and D reflect the most gender-skewed behaviors found by the model for men and women, respectively. Behavior C features upper-body, barbell-based strength training exercises, whereas Behavior D features light cardiovascular exercises and group activities such as yoga and Pilates.

Behaviors E and F reveal two curious demographic aspects. While Behavior E has noticeable skew towards women, the individuals of either gender who are mostly likely to participate in this behavior are shorter than the general population, with differences of 5.1in for males and 3.1in for female. A manual inspection for why this behavior would select for shorter physiological characteristics revealed several thousand accounts that had varied demographics but which recorded only a few workouts consisting of the most-probable exercises for this behavior, with no other distinguishable characteristic such as sport engagement; this limited posting and anomalous height suggests that these accounts may be fraudulent. In contrast to the height trend seen in Behavior E, Behavior F shows the genders varying in the age of peak adoption, with women most likely to engage in the behavior in their early 20s and men in their late 20s.

For the general population, Table 1 showed high-level trends in how exercising behavior changed. Therefore, we

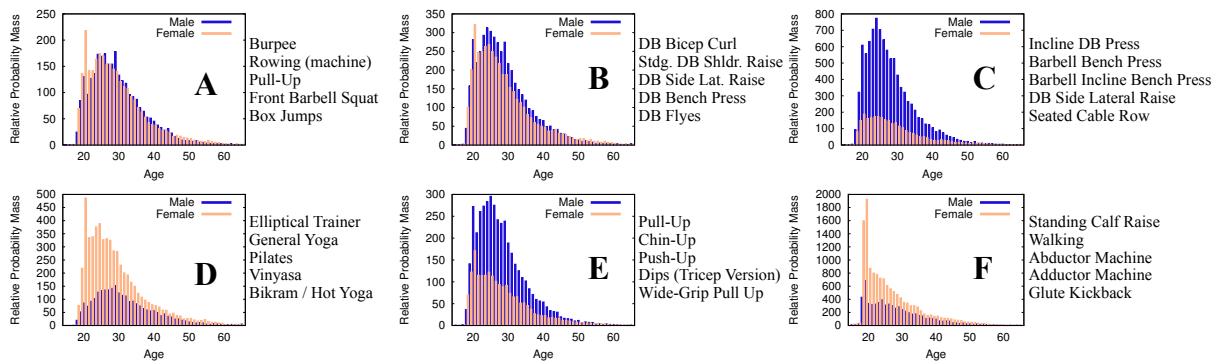


Figure 2: Examples of the most-probable exercises and population demographics of inferred behaviors. (DB denotes Dumbbell.)

| Behav. | % | | Avg. Age | | Avg. Height (in.) | |
|--------|-------|--------|----------|--------|-------------------|--------|
| | Male | Female | Male | Female | Male | Female |
| A | 0.489 | 0.511 | 29.9 | 29.3 | 69.3 | 63.8 |
| B | 0.532 | 0.468 | 29.3 | 28.4 | 69.2 | 63.8 |
| C | 0.776 | 0.224 | 28.2 | 28.6 | 69.4 | 63.9 |
| D | 0.304 | 0.696 | 31.9 | 29.0 | 69.2 | 64.0 |
| E | 0.680 | 0.320 | 28.7 | 28.2 | 69.3 | 63.2 |
| F | 0.329 | 0.671 | 29.0 | 26.5 | 65.1 | 61.7 |

Table 2: Population demographics for behaviors in Fig. 2.

considered a follow-up experiment to test whether behavior adoption rates varied across time. For all individuals of a given age, we computed the relative probability mass assigned to each of the twenty behaviors. Surprisingly, for the majority of behaviors, we did not observe significant differences in adoption rates across ages 20–60, which accounts for nearly all of the Fitocracy user base. Only two behaviors showed moderate changes: (1) a decrease in the behavior associated with body weight exercises, such as push-ups, crunches, and jumping jacks, and (2) an increase in the behavior associated with light cardiovascular activities. While our analysis did not track longitudinal changes in behavior for individuals, as a whole, these results still suggest that exercising behaviors remain relatively constant for the majority of the population across their active lifespan.

5 Conclusion

This paper has presented a first look at Fitocracy, a social media site oriented around the fitness and exercise community. Using a history of over 3M workouts from 188K individuals, we demonstrate that individuals’ exercising habits follow meaningful behaviors that match different motivations and which have strong demographic biases toward gender and even age. Details of additional behavior models and an interactive demonstration are available at <http://networkdynamics.org/resources/exercise>. In future work, we plan to study the social phenomena on the platform including measuring social influence and the roles of community and group behavior in exercise motivation.

References

- Blei, D. M.; Ng, A. Y.; and Jordan, M. I. 2003. Latent dirichlet allocation. *The Journal of Machine Learning Research* 3:993–1022.
- Kendall, L.; Hartzler, A.; Klasnja, P.; and Pratt, W. 2011. Descriptive analysis of physical activity conversations on twitter. In *Proceedings of CHI*, 1555–1560. ACM.
- Knapp, B. A. 2014. Gender representation in the crossfit journal: a content analysis. *Sport in Society* 0(0):1–16.
- Ma, X.; Chen, G.; and Xiao, J. 2010. Analysis of an online health social network. In *Proceedings of the 1st ACM international health informatics symposium*, 297–306. ACM.
- McDowell, M. A.; Fryar, C. D.; Ogden, C. L.; and Flegal, K. M. 2008. *Anthropometric reference data for children and adults: United States, 2003–2006*. US Department of Health and Human Services, Centers for Disease Control and Prevention, National Center for Health Statistics.
- Merrill, R. M., and Richardson, J. S. 2009. Validity of self-reported height, weight, and body mass index: findings from the national health and nutrition examination survey, 2001–2006. *Prev Chronic Dis* 6(4).
- Otoni, R.; Pesce, J. P.; and Las Casas, D. 2013. Ladies first: Analyzing gender roles and behaviors in pinterest. In *Proceedings of ICWSM*.
- Schoenebeck, S. Y. 2013. The secret life of online moms: Anonymity and disinhibition on youbemom.com. In *Proceedings of ICWSM*.
- Teodoro, R., and Naaman, M. 2013. Fitter with twitter: Understanding personal health and fitness activity in social media. In *Proceedings of ICWSM*.
- Walsh, G., and Golbeck, J. 2014. StepCity: A Preliminary Investigation of a Personal Informatics-based Social Game on Behavior Change. In *Proceedings of CHI*, 2371–2376.
- Xu, Y.; Poole, E. S.; Miller, A. D.; Eiriksdottir, E.; Kestranek, D.; Catrambone, R.; and Mynatt, E. D. 2012. This is not a one-horse race: understanding player types in multiplayer pervasive health games for youth. In *Proceedings of CSCW*, 843–852. ACM.