

I'll Be Back: On the Multiple Lives of Users of a Mobile Activity Tracking Application

Zhiyuan Lin
Stanford University
zylin@cs.stanford.edu

Tim Althoff
Stanford University
althoff@cs.stanford.edu

Jure Leskovec
Stanford University
jure@cs.stanford.edu

ABSTRACT

Mobile health applications that track activities, such as exercise, sleep, and diet, are becoming widely used. While these activity tracking applications have the potential to improve our health, user engagement and retention are critical factors for their success. However, long-term user engagement patterns in real-world activity tracking applications are not yet well understood.

Here we study user engagement patterns within a mobile physical activity tracking application consisting of 115 million logged activities taken by over a million users over 31 months. Specifically, we show that over 75% of users return and re-engage with the application after prolonged periods of inactivity, no matter the duration of the inactivity.

We find a surprising result that the re-engagement usage patterns resemble those of the *start* of the initial engagement period, rather than being a simple continuation of the *end* of the initial engagement period. This evidence points to a conceptual model of *multiple lives* of user engagement, extending the prevalent *single life* view of user activity. We demonstrate that these multiple lives occur because the users have a variety of different *primary intents or goals* for using the app. These primary intents are associated with how long each life lasts and how likely the user is to re-engage for a new life. We find evidence for users being more likely to stop using the app once they achieved their primary intent or goal (e.g., weight loss). However, these users might return once their original intent resurfaces (e.g., wanting to lose newly gained weight). We discuss implications of the *multiple life paradigm* and propose a novel prediction task of predicting the number of lives of a user. Based on insights developed in this work, including a marker of improved primary intent performance, our prediction models achieve 71% ROC AUC. Overall, our research has implications for modeling user re-engagement in health activity tracking applications and has consequences for how notifications, recommendations as well as gamification can be used to increase engagement.

ACM Reference Format:

Zhiyuan Lin, Tim Althoff, and Jure Leskovec. 2018. I'll Be Back: On the Multiple Lives of Users of a Mobile Activity Tracking Application. In *WWW 2018: The 2018 Web Conference, April 23–27, 2018, Lyon, France*. ACM, New York, NY, USA, 11 pages. <https://doi.org/10.1145/3178876.3186062>

1 INTRODUCTION

Activity tracking applications for mobile health have become an important part of people's daily lives. A US-nationwide study in

2013 found that 69% of US adults keep track of one or more health indicators including weight, diet, exercise, or symptoms, and 21% among them used a mobile app or device to do so [17]. Globally, the mobile health market is projected market to grow to over \$500 billion by 2025 [44]. Understanding how users engage with mobile activity tracking applications has the potential to significantly improve people's health, for instance by preventing negative health outcomes and promoting the adoption and maintenance of healthy behaviors [4, 30, 43].

However, user engagement patterns in these activity tracking applications, especially long-term and at scale, are not yet well understood. While there has been a wealth of research on user engagement in various *online* settings (e.g., [12, 24, 31, 32, 36, 37]), it is unclear how users engage and re-engage with activity tracking applications that capture their *offline* lives in the real world. Understanding user engagement in these real-world contexts is particularly important given that we could help people improve their health. As mobile health applications that track a variety of health metrics and daily activities are becoming more popular, it is important to model and understand user engagement in these applications.

Typical modeling of user engagement considers users to have a single "lifetime" (or "lifespan") during which the user typically becomes less and less engaged on the platform [11, 13, 15, 22, 45, 48]. This conceptual model of a *single lifetime* has been widely used in user modeling and engagement research, as well as interventions aimed at increasing user engagement [5, 9–11, 13, 22, 45]. For example, existing work has focused on predicting the duration of user's single lifetime [15, 21, 23, 46] and attempted to extend a user's lifetime to *retain* them [6, 48]. Usually, once a user has been absent for a long time, they are very unlikely to return and thus pronounced "dead" [25]. Survival modeling techniques, which have been applied to user modeling [16, 21, 23], also assume that once a user has "died", they are not coming back to re-engage with the app. However, in the context of activity tracking applications that track real-world behaviors, user engagement patterns may not follow this single lifetime model. Users may return to re-engage with applications after long periods of inactivity and, this way, start a new life. There is a limited understanding of how users re-engage with activity tracking apps after long periods of inactivity, the mechanisms behind such behavior, and whether this behavior may be predictable ahead of time.

This work. Here, we conduct a large-scale observational study of user re-engagement patterns within a mobile activity tracking application. We demonstrate empirically, across 115 million logged activities taken by over a million users over 31 months, that over 75% of active users do return to the application after a prolonged period of inactivity. This behavior is independent of the duration of inactivity and often users return to the application multiple times. While many applications use notifications and e-mails to regain

This paper is published under the Creative Commons Attribution 4.0 International (CC BY 4.0) license. Authors reserve their rights to disseminate the work on their personal and corporate Web sites with the appropriate attribution.

WWW 2018, April 23–27, 2018, Lyon, France

© 2018 IW3C2 (International World Wide Web Conference Committee), published under Creative Commons CC BY 4.0 License.

ACM ISBN 978-1-4503-5639-8/18/04.

<https://doi.org/10.1145/3178876.3186062>

user attention after brief periods of inactivity (e.g., a few days), we observe a large fraction of users returning after much longer periods of time (e.g., 90 days). Importantly, the usage patterns when re-engaging with the app mimic those of the *start* of the user’s initial engagement period and cannot be explained as a simple continuation of the end of the initial engagement period. These observations suggest a conceptual model of user engagement exhibiting *multiple lives*, where user activity after re-engagement is more similar to the beginning of the previous life than the end of it. Users re-engage with this application after prolonged inactivity to begin a *new life*, rather than continuing patterns of their previous life—often again and again. This multiple life paradigm stands in contrast to the single life paradigm prevalent in existing work on user lifetime modeling.

Despite a variety of available activities in the app, we find that users generally focus on logging a single primary activity, reflecting the user’s goal or *primary intent*. We show that multiple lives occur because users have a variety of different primary intents for using the mobile health app. While primary intents are different from user to user, the intents stay relatively constant over multiple lives of the same user. Primary intents shape how long each life lasts and how likely the user is to return for another new life. Associated with the user’s intent are performance goals such as losing body weight, walking or running more, or sleeping longer. We demonstrate that users getting closer or reaching such goals are more likely to stop using the app; that is, the app has fulfilled its purpose for them. However, users regularly return after prolonged inactivity with a lower level of performance (e.g., body weight increased again), perhaps after a setback or after setting a new goal; that is, the app may become useful again to pursue an old or new goal. We further demonstrate that user re-engagement may be driven by external and seasonal factors such as New Year’s resolutions in January, and summer months during which users may be more interested in physical activity and losing weight. Multiple life patterns also vary with user demographics. For example, we find that young users have much shorter lifetimes and are less likely to return for another life than older users.

To demonstrate the predictive power of our observations, we also formulate a novel prediction task of predicting how many lives a user will have. We demonstrate that the insights developed in this work allow us to predict whether a user re-engages for another life with 71% ROC AUC. We also show that improving performance related to the user’s primary intent is a marker of likely leaving the app soon, because their primary goal may have been fulfilled.

While we demonstrate the *multiple lives paradigm* within a single mobile activity tracking application, our conceptual model may be applicable to other mobile health applications, in particular when users may be pursuing, reaching, and resetting goals. Overall, our findings around the multiple lives paradigm have important implications for increasing user engagement in activity tracking applications. For example, early recognition of the primary intent could enable a more engaging personalized user experience and better notification or gamification experience. Further, recognizing that certain usage intents lead to few and short lifetimes can highlight which experiences in the app are currently not well supported. In addition, intents associated with many but very short lives could be a prime target to win users back.

| Dataset Statistics | |
|---|-----------------------------------|
| Observation period | 31 months July 2013 - Jan 2016 |
| # total users | 1,329,767 |
| # total check-ins | 114,947,892 |
| Avg. # of active periods ($\delta = 30$ days) | 1.7 |
| Avg. inactive period duration ($\delta = 30$ days) | 102.3 days |
| Avg. active period duration ($\delta = 30$ days) | 22.4 days |
| Median age | 32 years |
| % users female | 48.5% |
| % underweight (BMI < 18.5) | 4.2% |
| % normal weight (18.5 \leq BMI < 25) | 43.6% |
| % overweight (25 \leq BMI < 30) | 31.0% |
| % obese (30 \leq BMI) | 21.3% |

Table 1: Dataset statistics.

2 EXISTENCE OF MULTIPLE LIVES

In this section, we describe our dataset from an activity tracking application and demonstrate that user engagement with the application is segmented into multiple disjoint active periods with distinct characteristics.

2.1 Dataset

We conduct an observational study using data from a mobile activity tracking application, Argus by Azumio [5, 7, 38] (Table 1). This smartphone app allows users to track various daily activities including running, walking, cardio, heart rate, weight, sleep, drink, and food logging activities (the app also supports other more rarely used activities such as measuring stress or logging yoga which we aggregate and call “other” activities; less than 2% of total logging). For example, the drink activity is used to keep track of the user’s daily fluid intake and the workout activity is used to log various indoor exercises such as weightlifting or indoor-cycling. In addition, the app passively logs steps through the phone’s accelerometers and automatically infers calories consumption for users. However, we do not consider passively logged activity as active engagement and thus filter out such data. We call the user action of logging a particular activity at a particular time a *check-in*. We focus on users who use the app for at least a week. Our final dataset ¹ includes over one million users actively logging 115 million check-ins over the course of 31 months (Table 1). This long observation period and large scale uniquely enables us to study re-engagement patterns after prolonged user absence. Due to the popularity of the app, its users are relatively diverse in terms of age, gender, weight status, country of origin, and other features [7].

2.2 Multiple lives

First, we demonstrate that user engagement with the application is segmented into multiple disjoint *active periods*. We define an active period as a maximal segment with at most δ days in between consecutive user’s check-ins. If a user has not logged an action for at least δ days, we refer to it as an *inactive period*. An inactive period between two consecutive active periods is longer than δ days by definition. We use “(in-)activity index” to refer to a user’s n -th (in-)active period. We calculate the duration of an active period as the number of days between the first and last check-in in the period. We consider only users with λ days of total lifetime, where $\lambda > \delta$, such that it would be at least theoretically possible for a user to have more than one active period at a given threshold δ . Otherwise,

¹All data analyzed is pre-existing and de-identified. We have also obtained necessary IRB approvals.

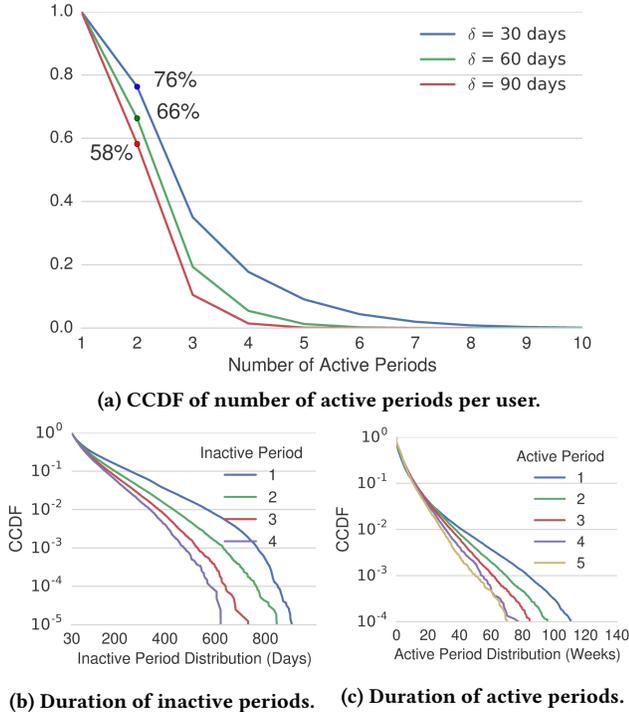


Figure 1: Users have multiple active periods. (a) 76% of users re-engage with the app for a second active period or more (after being absent for more than $\delta = 30$ days). (b) The gap between consecutive active periods (inactive period) becomes shorter over time. (c) Active periods also become shorter over time. CCDF refers to Complementary Cumulative Distribution Function.

we would, by design, obtain a downwardly biased estimate of the number of active periods of a user.

Results. We find that over 75% of users (total lifetime $\lambda > 30$ days) become active again after a prolonged $\delta = 30$ days of inactivity (Figure 1a). And these users often return more than once to the app. For example, 59% of users have at least three active periods where each time they were absent from the app for more $\delta = 30$ days. Importantly, this dynamics does not depend on the definition of the inactivity gap δ . Even when users are gone for $\delta = 90$ days, we find that 58% of them return for at least one more active period. Thus, even for very large inactivity thresholds δ users have *multiple* active periods, and this not simply an artifact of a small value of δ . We emphasize that, no matter the particular inactivity threshold, *most* users return to have multiple active periods. Therefore, studying multiple active periods does not restrict us to a small subsample of the population but, in fact, is relevant to the vast majority of users. We note that many applications use notifications and e-mails to regain user attention after brief periods of inactivity (e.g., a few days). However, we observe a large fraction of users returning after much longer periods of time (e.g., 90 days), when these users would typically have been considered “dead” for all practical purposes. In the following, we use $\delta = 30$ days unless specified otherwise.

We find that duration of both active as well as inactive period decrease with each additional active period as shown in Figures 1b and 1c, respectively. The average duration of the *first* active period is 24 days and the average inactive period duration between the first

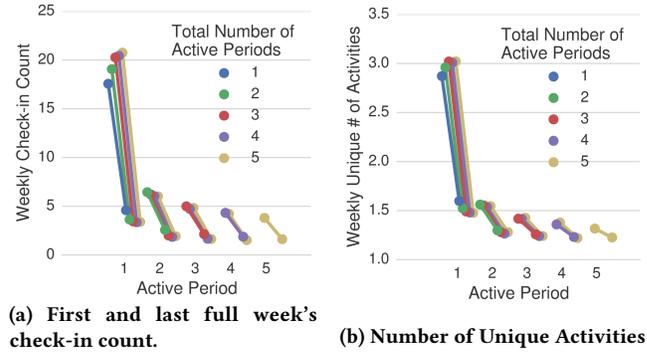


Figure 2: Users start active periods with frequent and diverse check-ins but end them with fewer and less diverse check-ins. When they re-engage with the app in another active period, check-ins are again more frequent and diverse again. Error bars in all plots correspond to bootstrapped 95% confidence intervals. Here they are too small to be visible.

and second active period is 114 days long. Note the log scale of the y -axis and the partially linear behavior of all curves. This suggests that both inactive as well as active period duration distributions can be approximated by exponential distributions.

Are multiple active periods simply segments of a single, long, fragmented life? We consider two simple measures of user activity: The number of check-ins per week and the number of distinct activities per week as a measure of diversity. (We find similar results when using entropy of activity distribution or fraction of check-ins of most frequent activity instead.) In Figure 2a, we plot the number of check-ins per week for every active period’s first and last week, connected by a line. Figure 2b shows the same for the number of distinct activities per week.

In both plots we observe a “zig-zag” pattern demonstrating that users start each active period with a larger number of check-ins compared to when they ended their previous active period, introducing a behavior discontinuity. Furthermore, these check-ins are also more diverse. Users also end each active period being less engaged and focusing on a less diverse set of activities than at the start of their active period. Importantly, note that the usage patterns at the *start* of a new active period mimic those of the *start* of the previous active period and are not simply a continuation of the *end* of the previous active periods. The observation suggests that a user’s active periods should in fact be considered separately, instead of considering them to be a single, long, fragmented life of a user. We observe this pattern for a wide range of potential δ values, which further supports this claim.

Our findings point to a very interesting pattern of user engagement with a health activity tracking app. We observe that most users’ lifetimes can be segmented into multiple active periods (average duration of 24 days) separated by long periods of inactivity (average duration 114 days). Furthermore, we observe that users often return after a long absence and that their activity is not a continuation of the previous usage patterns but it looks like as if they are using the app for the first time (Figure 2).

Conceptually, this is consistent with a *multiple lives* metaphor, where user’s leave the app, are absent for a long time, and then return as if they are a new user. Thus, we will use “life” to refer to an individual active period, use “lifetime” to refer to the duration of an active period, and use “life index” to refer to a user’s n -th active

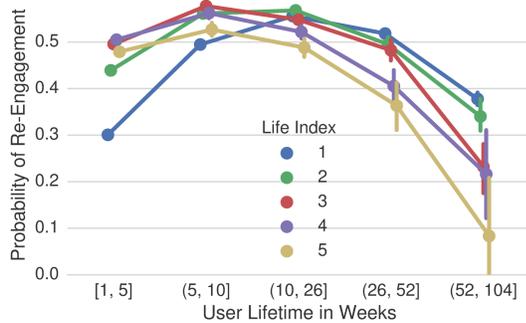


Figure 3: Probability of users re-engaging in new lives across different user groups of different lifetime.

period. In the rest of the paper, we show how this multiple lives paradigm affects health tracking app usage².

Relationship between total user lifetime and the number of lives. We observe a relationship between the total user lifetime and the number of lives (active periods). Figure 3 shows the probability of re-engaging for another life as a function of the duration of the previous life. In the first life (blue curve) we observe a clear U-shape with both very short and very long first lives being associated with a smaller probability of multiple lives. In later lives, this U-shape relationship attenuates and short lives are associated with the highest rates of return. Users with a very short *first* lives are unlikely returners because they might not have found the app very useful. However, in later lives, a short life does not mean the same thing—users return more often and have likely found value in the application before.

The existence of multiple life patterns for the vast majority of users shows that while users stop using the app regularly, they do frequently return as well, suggesting that there are times when application is valuable to its users.

3 WHY MULTIPLE LIVES?

Next, we illuminate why there are multiple lives and what are the mechanisms behind users leaving and returning multiple times? We demonstrate that engagement patterns across multiple lives vary based on the user’s primary intent as well as external influences.

3.1 Multiple User Intents

Users of an activity tracking application, such as the one studied in this work, use the app with a wide variety of intentions. For example, some users might want to lose weight and use the app for regularly tracking their weight changes. Others might want to be more physically active or sleep better. We first formalize this notion of user intent and will later show that it helps explain how users are using the app and why their engagement patterns follow multiple lives.

Empirically, we find that most users use the app in a very focused way, only using a small number of different activities. Specifically, just one activity is enough to cover 50% of all check-ins for 87% of all users (Figure 4; green line). To cover 90% of all check-ins, two activities are enough for half the user population (blue line).

²Of course, one could also view the active periods simply as segments of a single, long, fragmented user life. However, we would then expect usage patterns to be continuous between active periods, which is not the case. Instead, we observe that the overall user life is segmented into periods of activity with long inactive periods in between.

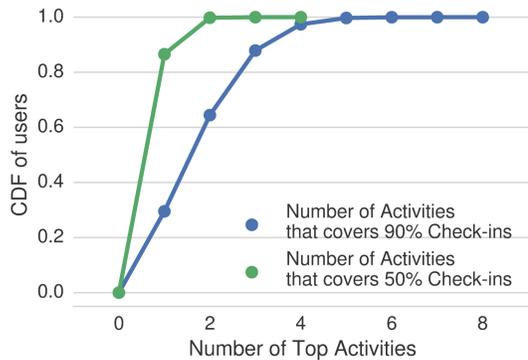


Figure 4: Number of activities required to cover a specific fraction of a user’s check-ins. Users tend to focus on very few activities. For example, one activity already covers 50% of all check-ins for 87% of all users (not necessarily the same activity for each user).

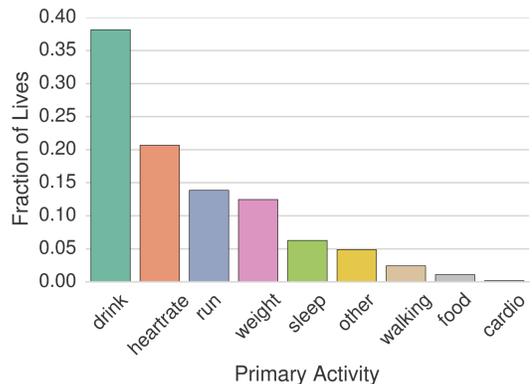


Figure 5: Fraction of lives focusing on each primary activity. Note that users pursue a wide variety of primary activities.

Because usage of the app is concentrated on very few activities, we can use the *primary activity* of each user (*i.e.*, the most commonly used one) as a proxy for the user’s *primary intent* of using the app.

The primary activity varies from user to user. As shown in Figure 5, about one third of users uses the app to track their drinking (*i.e.*, monitor their water, caffeine, or alcohol intake). One fifth of the users primarily track their heart rate (*e.g.*, immediately after workouts or in a resting condition; both are indicators of cardiovascular health and fitness). We observe that other users primarily track their weight (typically with the intent to lose weight), runs, walks, or sleeping patterns. A smaller number of users primarily focuses on food logging and cardio activities.

3.2 User Intent & Multiple Lives

Do users keep their primary intent in future lives? We compute for every user and life which activity a user logged most often (*i.e.*, the primary activity), and how likely a user is to keep their primary activity in consecutive lives. As shown in Figure 6, we find that users keep their primary activity from life to life most of the time, ranging from about 50% after the first life to almost 70% after the fifth life (blue curve). Users’ primary activity matches their first primary activity about 50% of the time (green curve). These probabilities are substantially higher than consecutive agreement

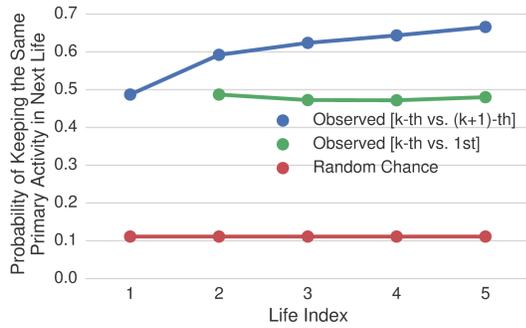


Figure 6: The probability of keeping the same primary activity in the next life ranges from 49% for the first life to 67% for the fifth life (blue curve). Thus, most of the time, users keep the same primary activity across lives. The probability of having the same primary activity in the first life ranges from 47% to 49% (green curve).

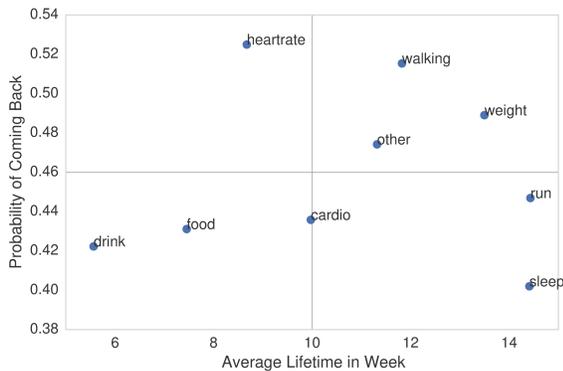
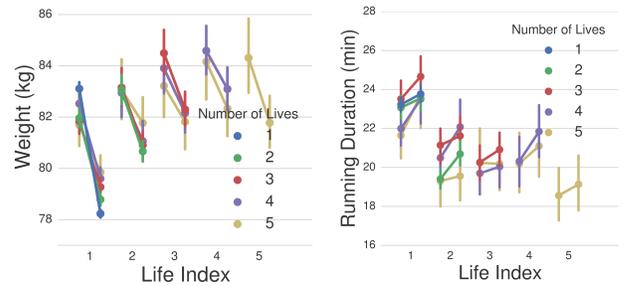


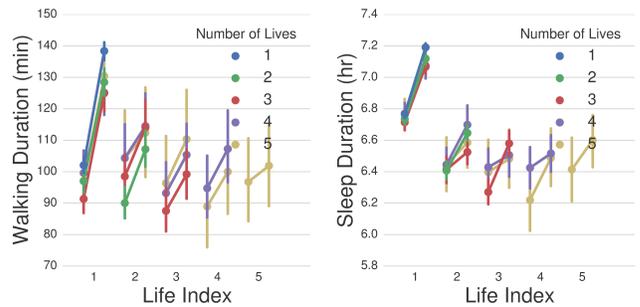
Figure 7: Average lifetime and return probability vary depending on the user’s primary activities.

based on picking one of the nine possible activities uniformly at random (red curve). These findings show that users are typically “resurrected” with the same primary intent as before, rather than changing their intentions between consecutive lives. Note, however, that usage patterns in the next life are not simply a continuation of the previous life but show distinctive signs of a new life as discussed in Section 2.2.

Multiple life patterns vary with primary intent. We find that the way users use the app varies based on the user’s primary intent. Figure 7 shows the average lifetime in weeks on the x-axis and the probability of returning for another life on the y-axis across primary intents of the user (here, we only consider user lives with at least 10 check-ins to reduce noise). Empirically, we find that users primarily logging their heart rate are most likely to re-engage for another life (52%). Users who primarily log runs and sleep are less likely to return, but their average lifetime is almost twice as long compared to primary heart rate users (14.4, 14.4 vs 8.7 weeks). This suggests that heart rate measurements are used every now and then, but that most users do not monitor them over long periods of time. In contrast, run- and sleep-loggers appear to find the associated features (GPS trail, timing and pace statistics; sleep timing, duration and quality) useful over longer periods of time. Activities such as drink- and food-logging are associated with both few and short lives. On the other hand, walking- and weight-logging activities



(a) First and last logged weight of each life **(b) First and last logged running duration of each life.**



(c) First and last logged walking duration of each life. **(d) First and last logged sleep duration of each life.**

Figure 8: Average first and last logged value of weight (a), running (b), walking (c), and sleep duration (d). We find that users have lower weight, and longer running, walking, and sleep durations when they are about to stop using the app. However, when they re-engage, they log a higher weight and shorter running, walking, and sleep durations.

are linked to more and longer lives, suggesting that users regularly find value in this functionality and for long periods of time.

We discuss design implications for increasing user engagement based on this heterogeneity in multiple life patterns based on primary intent in Section 6.

3.3 Intent-oriented Performance Driving Multiple Lives

Associated with the user’s primary intent are specific performance goals. For example, a user logging weight would often be interested in losing weight, a user logging running or walking would typically be interested in doing so longer or faster, and users logging their sleep would often be interested in sleeping longer or more consistently. We hypothesize that users may stop using the application once they significantly improved towards a goal associated with their primary intent. Furthermore, these users may return to the app once they have a new goal to attain. In order to test this hypothesis, we choose four quantities for which we have associated performance outcomes that are reported through the application (Figure 8): weight activity using weight loss in kilograms, running activity using run duration in minutes, walking activity using walk duration in minutes, and sleep activity using sleep duration in hours (higher is better, except for weight as most users are overweight or obese; Table 1). We measure these activity-associated outcomes for both the first and last logged activity in each life. Furthermore, we consider user populations with a different number of lives separately in order to compare unchanging populations over time.

| Outcome\Prim. Activity | Weight | Running | Walking | Sleep |
|------------------------|---------|---------|---------|---------|
| Weight Change (kg) | -7.5*** | -3.9*** | -5.2*** | -1.4*** |
| Running Duration (min) | 1.3 | 0.3*** | 2.00* | 0.0 |
| Walking Duration (min) | 23.7*** | 21.8*** | 42.0*** | 22.8*** |
| Sleep Duration (hr) | 0.4*** | 0.7*** | 0.5** | 0.2*** |

Table 2: Users that focus on a particular activity, tend to improve outcomes associated with that activity (e.g., primarily-walkers increase their walk durations but also tend to lose weight). Every column is a different user population with corresponding primary activity. Rows correspond to different outcome measures. Wilcoxon signed-rank test is used to test significance (* $p < 0.05$, ** $p < 0.01$, * $p < 0.001$).**

Results. We show average activity-associated outcomes for weight (Figure 8a), running (Figure 8b), walking (Figure 8c), and sleep activities (Figure 8d) with user groups of different number of total lives in different colors. Across all four plots we observe a “zig-zag” pattern, where users start their new life with a lower level of performance than they end it with; that is, weight decreases, runs duration increases, walk duration increases, and sleep duration increases within each life. Furthermore, we find that in all cases when users are returning to the app, they have a lower performance level again (e.g., regained weight, or shorter runs/walks/sleeping). We find very similar result when focusing on primary users of each activity only.

This suggests that making significant progress towards these performance goals is associated with leaving the app, in which case the app might have fulfilled its purpose for that particular user. However, users may set themselves new goals or experience setbacks (e.g., increased weight or decreased fitness levels) giving them a reason to return to the activity tracking app, which may have helped them previously to realize their goals. Then, the app becomes useful again to improve performance in some domain. We note that the “zig-zag” patterns in Figure 8, where users leave the app with improved performance and return with decreased performance, provide additional evidence that user engagement in this activity tracking application is indeed segmented into multiple lives.

Users typically improve performance of primary activity.

Here we provide additional evidence that users focusing on a certain activity tend to improve performance outcomes associated with that activity (Table 2). Specifically, users tend to improve the most on their primary activities. For example, primary-walkers improve the most (42 minutes) on walking durations. For users primarily logging weight, we find that the last weight logged by these users is on average 7.5 kilograms lower than their first logged weight. These users lost the most weight compared to other groups focusing on different primary activities. We note that activities are also intimately related. For example, primarily logging weight is not the only way app users can lose weight. In fact, we find that primarily-walkers and primarily-runners also log significant weight loss (-5.2kg and -3.9kg, respectively) and increased sleep duration (0.5 and 0.7 hours, respectively). This highlights the complexity of interrelated activities and health behaviors, where some activities may replace another, while other activities may even further support each other and the goals associated with them (e.g., increased physical activity from running and walking may support weight loss). Overall, we find that users tend to improve performance outcomes associated with their primary intent, and that users may

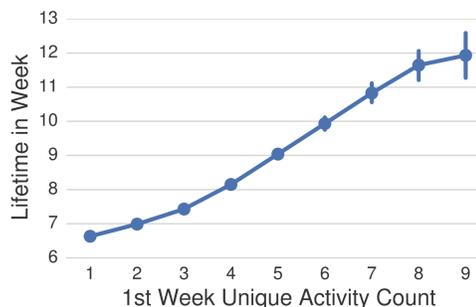


Figure 9: Users with higher number of distinct activities during the first week experience longer lifetimes.

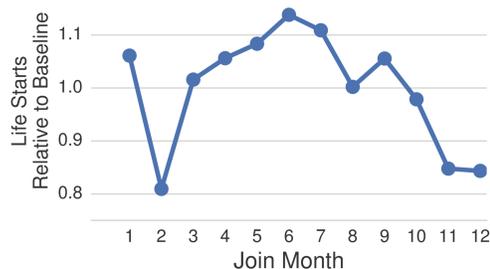


Figure 10: Users start new lives predominantly in January (6% over average rate) and summer months (14% over average rate), possibly related to New Year’s resolutions and summer related activity and weight loss goals.

leave the app afterwards.

Users with more diverse activities live longer. We just observed that, for example, users primarily focusing on running or walking also logged large weight loss. Intuitively, this makes sense. In order to lose weight, one needs to do more than just log weight. Instead, one needs to increase energy expenditure through exercise or reduce energy intake from food and drinks. Logging multiple activities could be a marker of increased interest in one’s health or a marker of intended behavioral changes that would support specific health goals. Here, we test the hypothesis whether users with more diverse sets of activities (e.g., running and weight instead of just weight logging activities) would be different from less diverse users in terms of their lifetime duration. Empirically, we find that users with more diverse initial usage patterns live longer (Figure 9).

3.4 External Influences

The activities tracked by activity tracking applications take place in the real, physical world. As such, they are influenced by the practices and demands of the real world. For example, there may be seasonal influences driving user intents and goals. To investigate external, seasonal influences, we consider each new user life and count the number of lives that are started in each month of the year (i.e., users either start using the app for the first time or return to the app after a prolonged inactivity). We normalize with respect to the average monthly rate. We only consider observations between January 2014 and December 2015 to ensure that our observation period includes exactly two instances of each month.

Results. We observe that the distribution of starting time of new lives is clearly non-uniform (Figure 10). We find the highest number of new lives starting in January (6% over average rate) and June

(14% over average rate). These findings can be explained by external influences driving user intentions. For example, most users come from the United States and other western, developed countries in the Northern Hemisphere. In these countries, New Year’s resolutions are commonly expressed on January 1; a tradition in which a person resolves to change an undesired trait or behavior, to accomplish a personal goal, or otherwise improve their life. Many New Year’s resolutions revolve around being more physically active or losing weight and users with these resolutions may seek out activity tracking applications to support them. Furthermore, people often seek to lose weight or be more active during the summer months around June. During this time weather is often more favorable to physical activity and users may attempt to lose weight for the “swimsuit season”. The patterns of when new lives are started suggest that users of activity tracking applications are externally influenced by real-world practices and demands. Particularly in this example, user (re-)joining time may be influenced by seasonal effects.

We note that multiple lives patterns also vary with demographic factors such as age, gender, and weight status. For example, we find that young users have much shorter lifetimes and are less likely to re-engage after long inactivity than older users. We will exploit these correlations in Section 4 in order to predict whether a user will re-engage in a new life and to predict their lifetime.

4 PREDICTING MULTIPLE LIVES

We formulate a novel prediction task of predicting how many lives a user will have. This section leverages previously described insights, including a marker of improved primary intent performance, in order to predict whether a user comes back for another life. We also predict how long a given life will last. We demonstrate that while there is large variability in user re-engagement patterns, the factors studied in this work allow us to successfully predict whether a user re-engages for another life (71% ROC AUC) and how long this life will last (82% ROC AUC). We note that the prediction tasks are designed to validate our empirical findings. Features derived from our findings are potentially generalizable to capture multiple lives in other similar applications.

4.1 Task Description

Future Life Prediction. For any (completed) user life, we predict whether this user will have additional future lives; that is, whether they will re-engage with the app again after prolonged inactivity. This is a new prediction task based on the multiple life paradigm introduced in this work.

Lifetime Prediction. For any user life, after observing the first four weeks ($w = 28$ days), we predict if a user is going to be a *short-term user* (leaving within the next $m = 30$ days) or a *long-term user* (living longer than $n = 183$ days). This follows the setting proposed by Danescu-Niculescu-Mizil et al. [15]. In this setting, we drop users with lifetime in $[w + m, n) = [58, 183)$ days to increase contrast between the classes. We also evaluated our prediction models for different parameter choices and found that our results, including the relative predictive power of individual features, were robust across a wide range of parameters.

4.2 Experimental Setup

Note that both tasks are formulated as binary prediction tasks. To define the user lives, we use $\delta = 30$ as before (again, we find very similar results for other choices of δ). In order to avoid premature classification of whether users will re-engage or not, we only consider user lives that are complete well before the end of the observation period. Specifically, we ensure that all considered users have no check-ins within the last 180 days of the observation period (this is larger than the average inter-life gap which is 114 days after the first life and shorter for following lives; see Figure 1b). Note that by the very nature of the re-engagement patterns described in this work, one can never be sure whether a particular users may still re-engage after 180 days. However, we tested various other thresholds and found similar results. In total, this leaves us with 1,267,897 user lives of 851,582 distinct users for prediction. We use the area under the ROC curve (ROC AUC) as our evaluation metric and use 10-fold cross validation for estimation. We report performance for Gradient Boosted Tree models and optimize number of trees, tree depth, and learning rate through cross-validation on the training data. We also experimented with Logistic Regression and linear SVM models, which consistently gave lower performance due to prominent nonlinear relationships (e.g., users with medium-length lifetime are most likely to return to the app; Figure 3).

Models. We define a series of models with different feature sets in order to learn what features are most predictive of future lives and lifetime. We use the same features for both prediction tasks. However, in lifetime prediction, we compute these features from only the first four weeks of each life. If features are missing we impute zero and include a binary variable indicating missingness.

- (1) **Lifetime:** Lifetime of current life in days (Section 2.2). (Of course, we exclude this feature when predicting lifetime.)
- (2) **Usage Pattern:** Weekly numbers of check-ins, number of distinct activities, entropy of activity distribution, and fraction of check-ins from most frequent activity (Section 2.2). We also include the week-to-week changes in these metrics as features.
- (3) **Primary Activity:** Categorical variable indicating the user’s most frequent activity (Section 3.2).
- (4) **Performance Change:** As a marker of improved primary intent performance, we include the change in weight, and running, walking, and sleep durations as well as the number of check-ins for each activity (Section 3.3).
- (5) **Demographics:** Three categorical variables for age, gender, and body-mass index (Section 3.4).
- (6) **Join Time:** Number of days between the user’s first activity and the launch of Argus app in July 2013 (Section 3.4).
- (7) **All:** A combination of all features.

4.3 Results

Future Life Prediction. The prediction accuracies for predicting whether a user will re-engage with the app in the future are shown in Figure 11, separately for the first to fifth life of each user. For predicting whether a user will re-engage after their first life, our full model using all features achieves 71% ROC AUC. Note that a random baseline would achieve 0.50 ROC AUC. We observe that all feature groups carry significant predictive performance. In addition, we find that the performance change marker (Section 3.3) predicts user re-engagement with up to 58% ROC AUC. Further, prediction performance drops with each life.

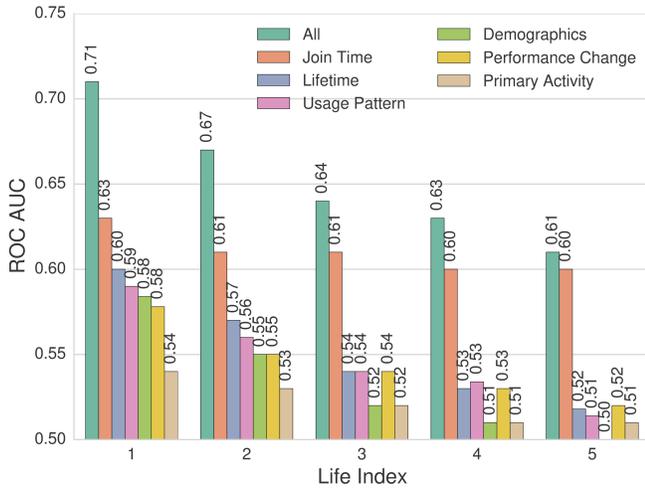


Figure 11: Future life prediction performance. We predict at the end of each life whether a user will re-engage eventually.

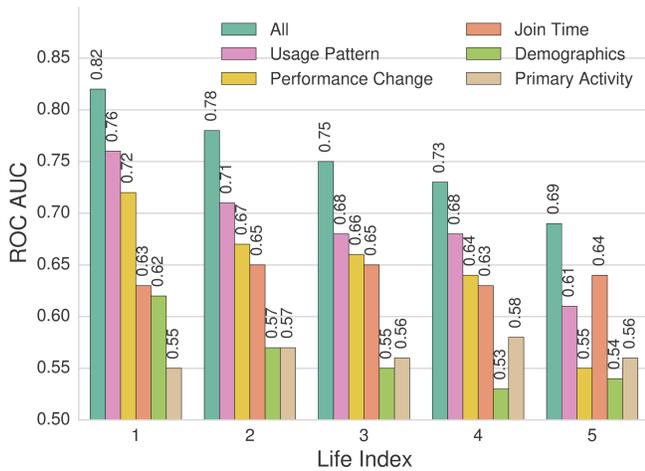


Figure 12: Lifetime prediction performance. We predict whether a user will leave shortly or stay long-term.

Lifetime Prediction. The prediction accuracies for lifetime prediction are shown in Figure 12, again separately for the first to fifth life of each user. Our full model achieves 82% ROC AUC for the first life. We again observe that all feature groups carry significant predictive performance. We find that the performance change marker is more predictive of lifetime than re-engagement with up to 72% ROC AUC (compared to 58% ROC AUC before). Again, we observe that the task becomes harder for later lives with full model accuracy ranging from 82% to 69% ROC AUC for the fifth life.

Discussion. In the lifetime prediction task, usage patterns and performance changes are the most predictive features to distinguish long-term users from short-term users, as highly engaged users with frequent check-ins also tend to stay longer using the app (Section 3.3). In the future life prediction task, the join time is highly predictive of re-engagement since it differentiates early adopters from late-joining users as well as identifying users joining in specific months including those with New Year’s resolutions and summer weight loss goals (Section 3.4). Prediction performance drops in later lives is potentially due to both fewer data points in later lives (Figure 1a) and more noise due to shorter lifetime

(Figure 1c). Users in earlier lives display greater changes in usage pattern (Figure 2) and performance (Figure 8), which may also contribute to better prediction accuracy in earlier lives as they are more distinguishable from noise in data. Our results demonstrate that the insights described in this work allow to successfully predict whether a user re-engages for another life and how long this life will last. These models could be used to identify well-suited user populations that could be targeted with additional notifications, e-mails and incentives, with important implications for increasing user engagement and retention.

5 RELATED WORK

User engagement. User engagement has been defined as the “quality of the user experience that emphasizes the positive aspects of interacting with an online application and, in particular, the desire to use that application longer and repeatedly [24].” An extensive literature has studied how to measure user engagement through subjective (e.g., self-report) and objective measures (e.g., eye tracking, mouse movements) [12]. In the context of mobile applications, engagement has been measured via clicks on notifications, time past before seeing the notification, and usage time [29]. In this work, we used simple, objective statistics to operationalize user engagement based on the frequency of interactions with the application. Previous work on user engagement has discussed the concept of *re-engagement* where users interrupt usage for a few minutes or hours [31, 32]. Many studies were conducted over a brief period of time which made studying re-engagement patterns impossible. In contrast to brief usage interruptions, in this work we study re-engagement patterns after extensive inactivity (e.g., 30-90 days).

Modeling & predicting user engagement. User engagement has previously been modeled based on repeat consumption patterns [11, 13], binge watching [45], and patterns of switching between boredom and sensitization [22]. A user’s intent in using a particular application can be predicted from the user’s behavior [14, 28] since it is strongly influenced by their intent [2]. Typical modeling of user engagement has considered users to have a single “lifetime” during which the user typically becomes less and less engaged on the platform [15, 48]. Much research has been devoted to predicting time of next user activity [6, 21, 23], total lifetime [15, 46], often using survival modeling techniques [16, 21, 23]. In the context of wearable and mobile device data, user engagement research has attempted to proactively engage users [33], infer the mood of users [26], and suggest healthy behavior changes [19, 20, 34]. All these studies assumed that user engagement follows a single life paradigm.

Increasing user engagement. Previous studies have shown that user engagement can be increased through notifications [3], incentives including badges [9, 10], gamification [8, 18, 20], and social network features [5, 20]. Furthermore, understanding the user’s intent [14, 28] can help designers surface different interaction modes [39], provide better contextual help [40], and personalize search results and recommendations [41, 42].

This work. This work studies user engagement and re-engagement patterns after long periods of inactivity in the context of activity tracking applications. We extend previous work (e.g., [1, 5, 7, 27, 35, 38, 47]) by discovering that users regularly re-engage after long periods of inactivity structuring user engagement into *multiple lives*

with distinct characteristics. We demonstrate that these multiple lives are driven by user intent and external influences. Further, we propose a novel prediction task of predicting the number of lives of a user.

6 DISCUSSION & CONCLUSION

Finally, we summarize our findings in the form of a conceptual model of user engagement across multiple lives. We also discuss design implications for increasing user engagement in this setting.

6.1 Conceptual Model of Multiple Life User Engagement in Activity Tracking Apps

The presented empirical evidence can be explained by the following model of user engagement across multiple lives:

- (1) Users join the application with a specific, primary intent and associated goals, starting a new life (Section 3.1).
- (2) The user uses the app in accordance with their primary intent (Section 3.1). Users typically improve in performance metrics associated with their primary intent during this time (Section 3.3).
- (3) Once the user has made significant improvements towards their goal they may leave the app, corresponding to the end of their lifetime (Section 3.3).
- (4) When the user forms a new goal (often based on external, seasonal influences; Section 3.4), or experiences a setback (e.g., regaining lost weight), they may join the app again starting a new life (Section 3.3). In many cases, the primary intent of their new life remains the same as before (Section 3.2).

6.2 Implications for Increasing Engagement

Our findings around the novel multiple life paradigm have implications for increasing user engagement in activity tracking applications. For example, early recognition of the primary user intent could enable a more engaging personalized user experience. We find that it only takes 6 days from when a user starts using the app to predict a user's eventual primary activity with 80% accuracy. Once the user intent is inferred, one could cross-promote app features that support the user's primary intent (e.g., walking, running, or cardio for users attempting to lose weight).

Furthermore, observing how different user intents lead to user engagement with varying number of lives and lifetime can shine a light on how different features are used and where users see or do not see value. Consider the four quadrants of Figure 7. The bottom left quadrant, containing the primary intents drink and food, corresponds to few and short lifetimes. This suggests that these experiences in the app are currently not well supported. The top left quadrant, containing primarily heart rate-logging users, shows that these users tend to log heart rate across multiple lives, but only for a relatively short time each. These may be activities that are useful to track, but do not necessarily need to be tracked regularly over long periods of time since they do not vary as dynamically and quickly (e.g., resting and maximum heart rate changes rather slowly). Here, one could consider interventions targeted at getting previous users with such intents to re-engage (e.g., notifications or e-mails such as "We saw you have not checked your resting heart rate in a while. It might have changed!"). The top right quadrant contains walking, weight, and logging of other activities which are associated with both more and longer lives. Likely, users find value in these features as they use them for a long, consecutive time,

and often re-engage after prolonged inactivity. User intents in the bottom right quadrant (run, sleep, and cardio) correspond to intents associated with long but few lives. In these cases, users find value in tracking these behaviors for a long, consecutive time, but once they have completed their life, they are less likely to re-engage again. Here, one can consider, for example, performance/goal-driven interventions such as "Can you still run a 10k in 53 minutes?" or "Can we help you manage your weight?"

Lastly, gamification techniques such as badges and rewards could be used to incentivize specific multiple-life behaviors. For example, they could be awarded for each re-engagement after prolonged inactivity. Importantly, this would need to be complemented with rewards for long-time use. Otherwise, users may feel incentivized to leave the app when they would not have done so otherwise.

In summary, modeling usage periods as multiple lives can enable a better understanding of the user engagement mechanisms in mobile activity tracking applications and support of the user's motivations and intents.

Limitations. This study is a large-scale case study of the mobile activity tracking application Argus. Therefore, we are limited in our ability to generalize our findings and the multiple life paradigm to other mobile health applications. However, Argus is one of the most popular and general activity tracking apps on the market. Further, our results show that multiple lives are likely due to real-world intentions and goals, and therefore are influenced by external factors. Therefore, our discoveries may generalize to other mobile health applications, where engagement is also driven by real-world intentions and factors. In fact, we are currently working with another dataset from the mobile app MyFitnessPal by Under Armour, used for calorie counting and weight loss. In this dataset we also observe that user engagement is clearly segmented into multiple lives.

6.3 Conclusion

User engagement is important to the success of mobile health applications and has historically been modeled as a single lifetime of user engagement. In this paper, we study user engagement patterns of over a million users over 31 months in a mobile activity tracking application. In contrast to previous work on modeling user engagement in online contexts, we discover that user engagement in activity tracking applications exhibits *multiple lives*. We demonstrate that multiple lives occur because users set new goals, experience setback from previous progress, or are influenced by external factors (e.g., seasonal effects). We discuss implications of the multiple life paradigm on increasing user engagement within activity tracking applications and propose a novel prediction task of predicting the number of lives of a user as well as their lifetime. We show that predictive models based on the insights developed in this work can successfully predict whether a user will re-engage with the app and the length of their lifetime. Our work has implications for modeling user re-engagement in mobile health applications, and consequences for how notifications, recommendations as well as gamification can be used to increase engagement.

ACKNOWLEDGMENTS

This research has been supported in part by NIH BD2K, DARPA NGS2, ARO MURI, IARPA HFC, Stanford Data Science Initiative, and Chan Zuckerberg Biohub.

REFERENCES

- [1] Aino Ahtinen, Minna Isomursu, Ykä Huhtala, Jussi Kaasinen, Jukka Salminen, and Jonna Häkkinen. 2008. Tracking outdoor sports—user experience perspective. *Ambient Intelligence* (2008).
- [2] Icek Ajzen. 1985. From intentions to actions: A theory of planned behavior. In *Action Control*.
- [3] Ghadah Alkhalidi, Fiona L Hamilton, Rosa Lau, Rosie Webster, Susan Michie, and Elizabeth Murray. 2016. The effectiveness of prompts to promote engagement with digital interventions: a systematic review. *JMIR* (2016).
- [4] Tim Althoff. 2017. Population-Scale Pervasive Health. *IEEE Pervasive Computing* 16, 4 (2017), 75–79.
- [5] Tim Althoff, Pranav Jindal, and Jure Leskovec. 2017. Online Actions with Offline Impact: How Online Social Networks Influence Online and Offline User Behavior. In *WSDM*.
- [6] Tim Althoff and Jure Leskovec. 2015. Donor retention in online crowdfunding communities: A case study of DonorsChoose.org. In *WWW*.
- [7] Tim Althoff, Rok Susic, Jennifer L Hicks, Abby C King, Scott L Delp, and Jure Leskovec. 2017. Large-scale physical activity data reveal worldwide activity inequality. *Nature* (2017).
- [8] Tim Althoff, Ryen W White, and Eric Horvitz. 2016. Influence of Pokémon Go on physical activity: Study and implications. *JMIR* (2016).
- [9] Ashton Anderson, Daniel Huttenlocher, Jon Kleinberg, and Jure Leskovec. 2013. Steering user behavior with badges. In *WWW*.
- [10] Ashton Anderson, Daniel Huttenlocher, Jon Kleinberg, and Jure Leskovec. 2014. Engaging with massive online courses. In *WWW*.
- [11] Ashton Anderson, Ravi Kumar, Andrew Tomkins, and Sergei Vassilvitskii. 2014. The dynamics of repeat consumption. In *WWW*.
- [12] Simon Attfield, Gabriella Kazai, Mounia Lalmas, and Benjamin Piwowarski. 2011. Towards a science of user engagement (position paper). In *WSDM Workshop on User Modelling for Web Applications*.
- [13] Austin R Benson, Ravi Kumar, and Andrew Tomkins. 2016. Modeling User Consumption Sequences. In *WWW*.
- [14] Justin Cheng, Caroline Lo, and Jure Leskovec. 2017. Predicting Intent Using Activity Logs: How Goal Specificity and Temporal Range Affect User Behavior. In *WWW Companion*.
- [15] Cristian Danescu-Niculescu-Mizil, Robert West, Dan Jurafsky, Jure Leskovec, and Christopher Potts. 2013. No country for old members: User lifecycle and linguistic change in online communities. In *WWW*.
- [16] Vachik S Dave, Mohammad Al Hasan, and Chandan K Reddy. 2017. How Fast Will You Get a Response? Predicting Interval Time for Reciprocal Link Creation. In *ICWSM*.
- [17] Susannah Fox and Maeve Duggan. 2013. *Tracking for health*. Pew Research Center’s Internet & American Life Project. <http://www.pewinternet.org/2013/01/28/tracking-for-health/>
- [18] Juho Hamari, Jonna Koivisto, and Harri Sarsa. 2014. Does gamification work?—A literature review of empirical studies on gamification. In *HICSS*.
- [19] Katrin Hänsel, Natalie Wilde, Hamed Haddadi, and Akram Alomainy. 2015. Challenges with current wearable technology in monitoring health data and providing positive behavioural support. In *MobiHealth*.
- [20] Katrin Hänsel, Natalie Wilde, Hamed Haddadi, and Akram Alomainy. 2015. Wearable Computing for Health and Fitness: Exploring the Relationship between Data and Human Behaviour. *arXiv preprint arXiv:1509.05238* (2015).
- [21] How Jing and Alexander J Smola. 2017. Neural survival recommender. In *WSDM*.
- [22] Komal Kapoor et al. 2015. Just in time recommendations: Modeling the dynamics of boredom in activity streams. In *WSDM*.
- [23] Komal Kapoor, Mingxuan Sun, Jaideep Srivastava, and Tao Ye. 2014. A hazard based approach to user return time prediction. In *KDD*.
- [24] Mounia Lalmas, Heather O’Brien, and Elad Yom-Tov. 2014. Measuring user engagement. *Synthesis Lectures on Information Concepts, Retrieval, and Services* (2014).
- [25] Jure Leskovec, Lars Backstrom, Ravi Kumar, and Andrew Tomkins. 2008. Microscopic evolution of social networks. In *KDD*.
- [26] Robert LiKamWa, Yunxin Liu, Nicholas D Lane, and Lin Zhong. 2013. Moodscope: Building a mood sensor from smartphone usage patterns. In *MobiSys*.
- [27] Cameron Lister, Joshua H West, Ben Cannon, Tyler Sax, and David Brodegard. 2014. Just a fad? Gamification in health and fitness apps. *JMIR Serious Games* (2014).
- [28] Caroline Lo, Dan Frankowski, and Jure Leskovec. 2016. Understanding Behaviors that Lead to Purchasing: A Case Study of Pinterest. In *KDD*.
- [29] Abhinav Mehrotra, Sandrine R Müller, Gabriella M Harari, Samuel D Gosling, Cecilia Mascolo, Mirco Musolesi, and Peter J Rentfrow. 2017. Understanding the Role of Places and Activities on Mobile Phone Interaction and Usage Patterns. *IMWUT* (2017).
- [30] Inbal Nahum-Shani, Shawna N Smith, Bonnie J Spring, Linda M Collins, Katie Witkiewitz, Ambuj Tewari, and Susan A Murphy. 2016. Just-in-Time Adaptive Interventions (JITAs) in mobile health: Key components and design principles for ongoing health behavior support. *Ann. Behav. Med.* (2016).
- [31] Heather L O’Brien. 2011. Exploring user engagement in online news interactions. *ASIS&T* (2011).
- [32] Heather L O’Brien and Elaine G Toms. 2008. What is user engagement? A conceptual framework for defining user engagement with technology. *ASIS&T* (2008).
- [33] Martin Pielot, Bruno Cardoso, Kleomenis Katevas, Joan Serrà, Aleksandar Matic, and Nuria Oliver. 2017. Beyond Interuptibility: Predicting Opportune Moments to Engage Mobile Phone Users. *IMWUT* (2017).
- [34] Mashfiqui Rabbi, Min Hane Aung, Mi Zhang, and Tanzeem Choudhury. 2015. MyBehavior: Automatic personalized health feedback from user behaviors and preferences using smartphones. In *UbiComp*.
- [35] John Rooksby, Mattias Rost, Alistair Morrison, and Matthew Chalmers Chalmers. 2014. Personal tracking as lived informatics. In *SIGCHI*.
- [36] Martin Saveski, Daniele Quercia, and Amin Mantrach. 2014. The geography of online news engagement. In *SocInfo*.
- [37] Muhammad Zubair Shafiq, Jeffrey Erman, Lusheng Ji, Alex X Liu, Jeffrey Pang, and Jia Wang. 2014. Understanding the

- impact of network dynamics on mobile video user engagement. In *SIGMETRICS Performance Evaluation Review*.
- [38] Ali Shameli, Tim Althoff, Amin Saberi, and Jure Leskovec. 2017. How gamification affects physical activity: Large-scale analysis of walking challenges in a mobile application. In *WWW*.
- [39] Katie Sherwin. 2016. User intent affects filter design. *Nielsen Norman Group* (2016). <http://bit.ly/2evdOjs>.
- [40] Simone Stumpf, Xinlong Bao, Anton Dragunov, Thomas G Dietterich, Jon Herlocker, Kevin Johnsrude, Lida Li, and J Shen. 2005. Predicting user tasks: I know what you're doing. In *AAAI*.
- [41] Jaime Teevan, Susan T Dumais, and Eric Horvitz. 2005. Personalizing search via automated analysis of interests and activities. In *SIGIR*.
- [42] Jaime Teevan, Susan T Dumais, and Daniel J Liebling. 2008. To personalize or not to personalize: Modeling queries with variation in user intent. In *SIGIR*.
- [43] J Graham Thomas and Dale S Bond. 2015. Behavioral response to a just-in-time adaptive intervention (JITAI) to reduce sedentary behavior in obese adults: Implications for JITAI optimization. *Health Psychology* (2015).
- [44] Transparency Market Research. 2017. Digital Health Market - Global Industry Analysis, Size, Share, Growth, Trends and Forecast, 2017 - 2025. (2017). <https://www.transparencymarketresearch.com/digital-health-market.html>.
- [45] William Trouleau, Azin Ashkan, Weicong Ding, and Brian Eriksson. 2016. Just One More: Modeling Binge Watching Behavior. In *KDD*.
- [46] Yi Wang, Yuchun Guo, and Yishuai Chen. 2016. Accurate and early prediction of user lifespan in an online video-on-demand system. In *ICSP*.
- [47] Joshua H West, P Cougar Hall, Carl L Hanson, Michael D Barnes, Christophe Giraud-Carrier, and James Barrett. 2012. There's an app for that: Content analysis of paid health and fitness apps. *JMIR* (2012).
- [48] Jiang Yang, Xiao Wei, Mark S Ackerman, and Lada A Adamic. 2010. Activity Lifespan: An Analysis of User Survival Patterns in Online Knowledge Sharing Communities. In *ICWSM*.