Analysis of An Online Health Social Network

Xiaoxiao Ma
Department of Computer Science, University of Massachusetts Lowell
xma@cs.uml.edu

Guanling Chen
Department of Computer Science, University of Massachusetts Lowell
glchen@cs.uml.edu

Juntao Xiao
Jiangsu HanWin Technology Company LTD.
Suzhou, China
xiaojt@hanwintech.com

ABSTRACT

With the continued advances of Web 2.0, health-centered Online Social Networks (OSNs) are emerging to provide knowledge and support for those interested in managing their own health. Despite the success of the OSNs for better connecting the users through sharing statuses, photos, blogs, and so on, it is unclear how the users are willing to share health related information and whether these special-purpose OSNs can actually change the users' health behaviors to become more healthy.

This paper provides an empirical analysis of a health OSN, which allows its users to record their foods and exercises, to track their diet progress towards weight-change goals, and to socialize and group with each other for community support. Based on about five month data collected from more than 107,000 users, we studied their weigh-in behaviors and tracked their weight-change progress. We found that the users' weight changes correlated positively with the number of their friends and their friends' weight-change performance. We also show that the users' weight changes have rippling effects in the OSN due to the social influence. The strength of such online influence and its propagation distance appear to be greater than those in the real-world social network. To the best of our knowledge, this is the first detailed study of a large-scale modern health OSN.

Categories and Subject Descriptors

J.4 [**SOCIAL AND BEHAVIORAL SCIENCES**]: Sociology; H.4.0 [**INFORMATION SYSTEMS APPLICATIONS**]: General

General Terms

Measurement, Experimentation, Algorithms

Keywords

Health OSN, Social Influence, Correlation Analysis

1. INTRODUCTION

Patient-centered health management is an important direction that promotes preventative care to reduce health risks, to reduce hospital visits, and ultimately to reduce the overall healthcare cost. In particular, overweight and obesity may result in premature death and possibly lead to diabetes, heart diseases, and cancer. Studies show that about 64% of adult Americans were either overweight or obese, and 60% of U.S. adults do not exercise enough and 25% of adults do not exercise at all. It is estimated that the obesity healthcare costs US 147 billion dollars a year, doubled in less than a decade, and the cost will rise to 344 billion dollars by 2018 [5].

Weight management, however, requires that the participants be aware of diet knowledge, change health behaviors, and be persistent. Building a group of supporting family members, friends, and others with similar weight change goals is also an important factor to keep the participants motivated. For example, WeightWatchers is a program that works for many people (though not for all), aiming for steady and long-term weight loss by providing ongoing support and advice, education of healthier eating habits, plenty of tools and resources, and group meetings with other dieters to discuss problems.

With the continued advances of Web 2.0, health-centered Online Social Networks (OSNs) are emerging to provide knowledge and support for those interested in managing their own health. These health OSNs aim to empower the users with modern technologies, such as smartphone applications and sophisticated websites, that provide easier access to health knowledge, increase health awareness, motivate better health behaviors, and track weight loss progress. These emerging health communities are often formed around new devices and applications, such as wearable FitBit activity tracker,[1] Internet connected Withings body scale,[2] and smartphone based RunKeeper mobile application.[3]

Despite the tremendous success of the general-purpose OSNs, such as Facebook and Twitter, for better connecting the users through sharing statuses, photos, blogs, and so on, it is unclear how well the users are willing to share health related information, which is often considered personal and quite sensitive, and whether these special-purpose health OSNs can actually change the users' health behaviors to become more healthy.

This paper provides an empirical analysis of a popular

[1]http://www.fitbit.com/
[2]http://www.withings.com/
[3]http://www.runkeeper.com/

health OSN, FatSecret,[4] that allows its users to record their foods and exercises, to track their diet progress towards weight-change goals, and to socialize and group with each other for community support. Based on about five month data collected from more than 107,000 users, we studied their weigh-in behaviors and tracked their weight-change progress. We found that the users' weight changes correlated positively with the number of their friends and their friends' weight-change performance. We also show that the users' weight changes have rippling effects in the OSN due to the social influence. The strength of such online influence and its propagation distance appear to be greater than those in the real-world social network. To the best of our knowledge, this is the first detailed study of a large-scale modern health OSN.

The rest of this paper is organized as follows. In Section 2 we discuss related work. Section 3 introduces the FatSecret service. We describe our data collection process and present the characteristics of the collected data in Section 4 and 5, respectively. In Section 6, we present the correlation results to show what factors may impact the users' weight changes and quantify the social influence in the observed social network. Finally we discuss how a health OSN may improve its users' effectiveness of managing their health in Section 7, and conclude in Section 8.

2. RELATED WORK

Christakis and Fowler studied the nature and the extent of the social influence on obesity from family members and friends in a real social network, using data collected from 12,067 people over 32 years [3]. They showed that there is strong influence on one another's weight if two persons are relatives or friends. If one of a person's friends became obese, the risk of obesity for that person will increase significantly. The difference of our study and theirs lies in that we applied similar methodology for an *online* health social network, and we found that the social influence on weight changes is much stronger and it travels further distance in this particular OSN than what Christakis and Fowler have shown for a real social network.

There have been few studies of online health support communities. Maloney-Kichmar and Preece studied for two years of a thriving health bulletin board and how it played in the lives of its members and the relation between its members' online participation and their offline lives [9, 10]. They discussed the sociability issues on group process, the membership roles and patterns of social interaction. It is shown from this study that the members' participation in the community positively influenced their offline lives. The bulletin board, however, mainly focuses on facilitating group communications. Thus the authors had no quantitative and structural analysis of friend relationships and weight-change performances.

3. FATSECRET SERVICE

FatSecret is a free online social network service for people who are interested in food and diet. It has a comprehensive database of food nutrition facts, and allows its users to easily record food diaries to track their calorie intakes. Similarly it has estimations of the energy burned for common exercises, and allows its users to record their daily exercises.

[4]http://www.fatsecret.com/

In addition to recording diaries of food and exercises, its members are also encouraged to record their weight to track the progress towards their weight loss (or gain) goals. FatSecret engages its users by sending out emails periodically (every two weeks, but adjustable) to remind them to weigh-in (updating their weight on the website). The reminder email also informs the users about their friends' recent performance, like how much weight their friends have lost (or gained) every week. It will also show some recent recipes and diet tips to attract the users to come back visiting the website.

As a community, FatSecret enables its users to share cooking recipes, follow popular diets, and take up challenges together with other users (such as "Do not eat after 8pm"). FatSecret users can also organize themselves into groups with similar interests or goals, such as groups for "30's with 25-50 lbs to Lose" and "Google G1 Calorie Counter APP users." Within a group, its members can submit posts and share photos, and the group page shows all its members' recent activities, such as recording weight and updating the food and exercise diaries. There is also a bulletin board style forum that all FatSecret members can participate.

FatSecret members can add each other as friends by following a request and accept process. A user sends out a friendship request to another user, and they become friends if the request is accepted. Like many other online social networks, such as Facebook and Myspace, the friendships in FatSecret are symmetric. FatSecret users can adjust their privacy settings to share their weight history and diet diaries (food and exercise) with all other users, with their friends only, or with nobody (private mode).

Once the user logs onto the FatSecret website, he can see an activity "timeline," which includes all public and that his friends' weigh-in updates, food and exercise diary updates, group and forum updates, and so on. The user can filter this timeline stream by friends, groups, and types of the activities. For example, if a user chooses to view diet tips only, only the most recently submitted diet tips by others will be shown in the timeline. This is different from some social networking sites, such as Facebook, where only the friends' activities are shown in the timeline. By viewing all public activities, the user has access to much more user-generated information and may be more engaged and motivated. FatSecret also promotes a section on the main page to feature the members who are "doing well," making good progress towards their weight-change goals.

Besides the main website, FatSecret also provides a variety of methods for its users to get connected. It has mobile applications that run on iPhone, iPod touch, Android phones, BlackBerry phones, and iPad. The mobile applications allow the users to easily look up nutrition facts of different food and to record their weight, food, and exercise diaries. There is also a Facebook application that the users can use to compare their diet progress with their friends on Facebook.

4. DATASET COLLECTION

In this section we describe the activity timeline, the social graph, and the users' profile data that we have been collecting since January 5, 2010.

4.1 Timeline

Once the user logs onto the FatSecret website, the main page shows all public and that user's friends activities, include recording weigh in, recording food and exercise diary entries, recording journal notes, becoming friends of someone, joining user groups and challenges, posting comments in groups and forums, and so on. We created a FatSecret user, without adding any friend, and used a script that automatically logs onto the website and fetches the timeline data every 30 seconds. If some user marks his weight history and diet dairies to be private or friends only, such information will not be collected in our timeline data.

4.2 Social graph

FatSecret users are connected with each other through their friendship relations, forming a social graph with the vertices representing the users and the edges representing their friendships. A user's profile page on the FatSecret website also lists his friends, so we can use a script to parse and retrieve the names of those friends. Every day our script tries to collect the social graph based on a daily seed list, which contains a list of users we will query and retrieve their friends. A seed list contains all the seed users from yesterday and unseen users that appeared in today's timeline data. The seed list on the first day of our collection contains all the users appeared in that day's timeline.

Our script queries these seed users' friends lists, and if there are some users not in the seeds are found, the query continues and those users are also added to the seed list. The process repeats until no more unseen users are found, resulting in a snapshot of the daily social graph. Note that the daily social graph we collected is not necessarily complete, as we missed the user who joined before our collection started and none of her connected users (directly or indirectly) appeared in the public timeline during our study (no public activity). We do believe, however, that we have extracted the majority and particularly the complete largest connected component, given the number of users we found and the length of our data collection.

4.3 User profiles

When a user signs up with FatSecret, he needs to specify his weight at the time of registration, as the start weight. He also needs to specify the goal weight, as the target he intends to reach. Unlike other online social networks where the user profiles may contain gender and age information, FatSecret website keeps such information private and not visible to the public. For each user, we can also parse her profile page to get her number of friends (see above), registration time (the date when she signed up) and her current weight (from the last recorded weigh-in). In addition, the user's profile page contains a weight history table, from which we extracted the complete history of the time and the weight of every weigh-in recorded by that user. This allows us to progressively calculate how well the user followed her diet and performed against her weight goal.

5. DATA CHARACTERISTICS

In this section we describe basic characteristics of user categories, weigh-in behaviors, user friendships, and weight changes.

5.1 User categories

Until May 20th, we have collected data of 107,907 users from the FatSecret website. These users can be divided into three groups based on their sharing behaviors: public users, friends-only users, and private users. Public users publish their activities to all users, friends-only users only share their activities with their friends, and private users keep the data to themselves. We can differentiate friends-only and private users because their profile pages will state either "information only available to buddies" or "the user is not sharing her weight history." Among all the users, we found 91,047 public users (84.4%), 5641 friends-only users (5.2%), and 11,219 private users (10.4%). While intuitively the weight information tends to be personal, most FatSecret users created anonymous accounts and may feel more comfortable to share online. Since we cannot see the activities from the friends-only and private users, our study focused on the public users.

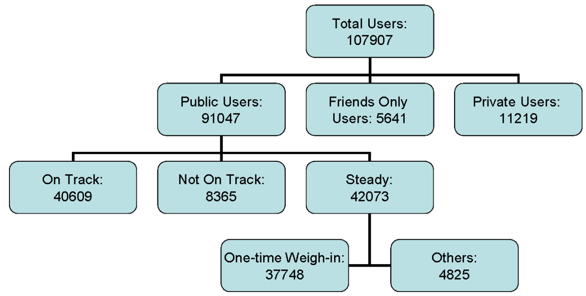


Figure 1: FatSecret user classification groups.

By comparing the users' start weight and goal weight, we can divide the 91,407 public users into three groups based on their objective: to lose weight, to gain weight, and to keep current weight. There are 88,613 users (97.3%) who want to lose weight, 1525 users (1.7%) who want to gain weight, and 909 users (1%) who want to keep their weight. As expected, the overwhelming majority of the FatSecret users want to lose weight. We can also group the users based on their weight-change progress by comparing their start weight, goal weight, and their current weight (from latest weigh-in). We call a user *on track* if his weight change is towards the direction he desires. For example, if a user wants to lose weight and his current weight is less than his start weight, he is on track. Similarly, if a user's weight change is on the direction opposite from her goal, he is *not on track*. For those users whose weight has not changed, we call them *steady* users. Among all the public users, we found 40,622 users (44.4%) on track, 8378 users (9.2%) not on track, and the rest of 42,047 (46.4%) users steady. Among the steady users, 37,722 of them only recorded one weigh-in, 3349 of them made two weigh-ins, while 1476 of them recorded more than two weight-ins. All these steady users' last weigh-in is the same as the start weight.

Figure 1 summarizes the categories of all the 107,907 users based on their respective groups.

5.2 User weigh-ins

Keeping the users engaged with the website to update their diet and weight is important, as it will increase the users' awareness of their health behaviors and results. A recent study shows that those who kept food records six days a week lost about twice as much weight as those who kept food records one day a week or less [6]. In this section we focus on the users' weigh-in activities. Note that the

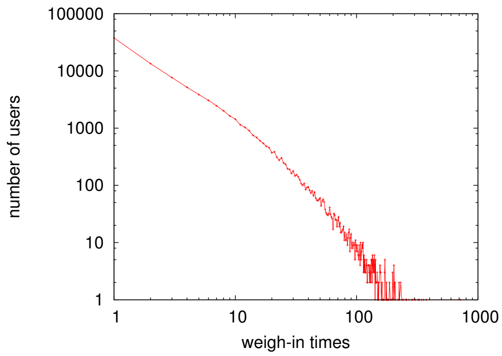


Figure 2: Distribution of weigh-in times.

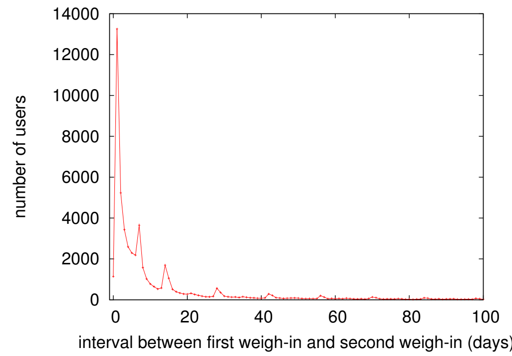


Figure 3: Intervals between first and second weigh-ins.

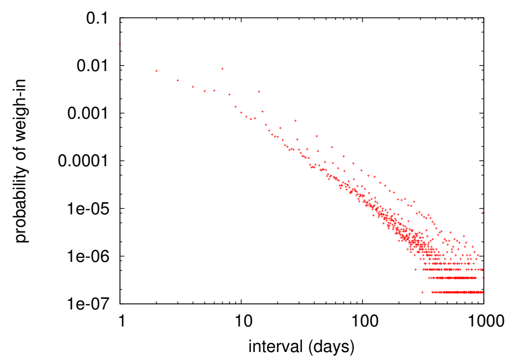


Figure 4: Distribution of intervals between two consecutive weigh-ins.

users by default receive periodic reminders biweekly, asking them to weigh-in on the FatSecret website.

On average, we saw about 2925 daily weigh-ins during our study period. Until May 20, all public users have recorded 555,706 weigh-ins, including those recorded before the starting date of our study as the users' profile page contains full weigh-in history. Thus on average each user has recorded 6.1 weigh-ins. If we remove those users who only made one weigh-in, the average number of weigh-ins per user is 9.7. On February 20, we had 44,012 seed users and we tracked their weigh-in activities during the next three-month period. Interestingly we found that 283 users restored their diet plan during this period, as they cleared the weight history preceding the reset date. For the rest of 43,729 users, they recorded 76,824 total weigh-ins. Thus on average a user only recorded 1.76 weigh-ins during these three months. Most users recorded less than 15 weigh-ins except that one person made 106 weigh-ins.

For all the 91,047 public users, we retrieved their complete weigh-in histories and counted the number of their total weigh-ins. The log-log distribution plot in Figure 2 shows that the number of weigh-ins for these users follows a Pareto distribution.

For all the 53,299 public users who made at least two weigh-ins after their registration, we retrieved 555,706 total weigh-ins from their histories. We look at how long it took for a user to record the second weigh-in (the start weight entered at the registration is the first weigh-in). Figure 3 shows that 13,252 of these users (24.9%) made their second weigh-in on the next day of registration. There were 1,136 users recorded the second weigh-in on the same day of their registration. There were 33,764 users (63.3%) recorded the second weigh-in within one week. Note that there are spikes of interval length of 7 days, 14 days, and so on, which were caused by the users responding to the periodic weigh-in email reminders.

Next we study the intervals between two consecutive weigh-ins for the same user. The maximum interval for a user is 1253 days between two weigh-ins, while the median is just 2.3 days. About 80% of the weigh-in intervals are shorter than 11 days and 90% of the intervals are shorter than 21 days. An interesting question is whether the users' weigh-in behaviors can be modeled collectively. Often human activities are highly dynamic and difficult to predict. While many have assumed that the users' inter-activity intervals follow an exponential distribution, some researchers have found that heavy-tailed or Pareto distributions are better to describe the user activities where very long periods of inactivity separate bursts of intensive activity [2]. In our case, Figure 4 shows the distribution of the weigh-in intervals on a log-log plot, which fits better with an exponential distribution. This suggests that the users' weigh-in activities, at least for the FatSecret users during our study, follow a Poisson process so that the consecutive weigh-ins follow each other at relatively regular time intervals without many very long waiting times.

5.3 User friendships

As the profile has the user's registration date, we can trace back to see how FatSecret user base grew over time. Figure 5 shows the number of new users weekly registered with FatSecret, which released a public API for third-party application developers in mid-August 2009 (about 38 weeks before May 20, 2010). Since then the website grew considerably. For clarity, the Figure starts from June 27, 2008, as there were not many weekly new users before that. During the week of the beginning of 2010, the website exploded with 5000 newly registered users in that week, probably as people made new year resolutions to lose weight. Now the website attracts 3000 to 4000 new users per week.

On May 20th, there were total 107,907 users and 78,188 friendship connections among them, thus on average each user has about 1.45 friends. Note that even for friends-only and private users who do not share their diet activities and weight history with the public, their friend lists are still vis-

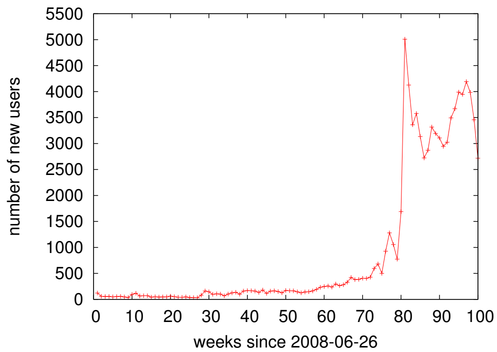


Figure 5: Weekly new users.

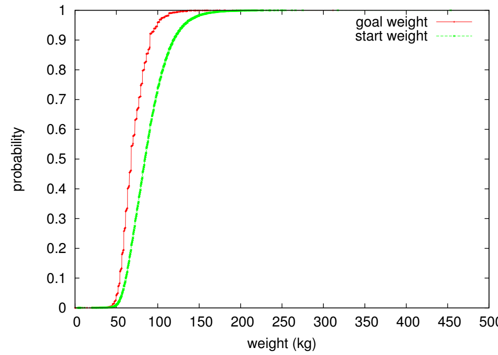


Figure 7: CDF of start weight and goal weight.

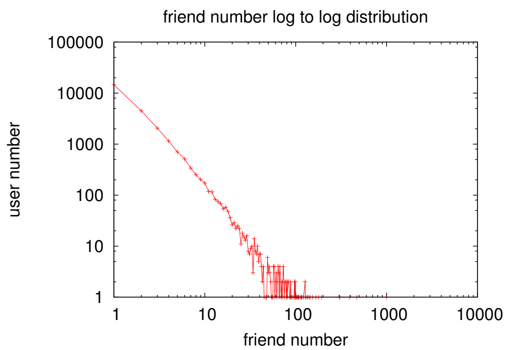


Figure 6: Distribution of friends number.

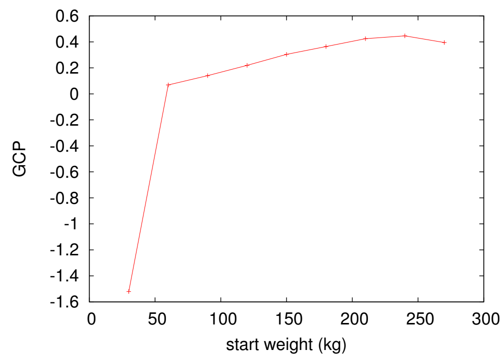


Figure 8: Desired weight-change percentage.

ible. As we had taken snapshots of the social graph (Section 4.2), we can track the changes of the friendship network. We observed that average number of friends or the network density of the social graph, decreased over time, from February 20 where it was 2.26, to 1.82 on March 25, then to 1.58 on April 23, and again to 1.45 on May 20. The reason could be: 1) the user base grew quickly during our relatively short study period, and these new users have not made many online friends yet; or 2) the users can already socialize through groups, challenges, and forums, thus they may not feel making online friends was necessary.

Figure 6 shows the distribution of the number of friends for all users using a log-log plot. Like many other online social networks [12, 8], the friendship distribution follows roughly the power law, though the tail is not quite smooth. The maximum number of friends a user had was 1002, and the second largest number of friends was 471. There were only 24 users who had more than 100 friends on FatSecret.

5.4 Weight changes

We found that the distribution of the users' start weight is a normal distribution with the mean value at about 82 kg. The distribution of the users' goal weight, however, is not the same. The users seem to choose their goal weight to be several particular values. There were 17 popular goal weight values, ranging from 52.2 kg to 90.7 kg, each was chosen by more than 2000 users. Of all the 91,407 public users, the three most popular goal weights were 68 kg, 63.5 kg, and 59 kg, which were chosen as the goal weight by 7412, 5963, and 5752 users, respectively. These values correspond to about 150, 140, and 130 pounds, respectively.

Figure 7 shows the cumulative distributions of the users' start and goal weights. The start weight curve is smoother than the goal weight curve, as the values of the start weight spread more evenly than the values of the goal weight. Also as most of the users intend to lose weight, the goal weight curve is on the left of the start weight curve. There are some abnormal values, such as two users set their goal weight to be 0, which probably means that they do not have a particular goal weight in mind. It is also interesting to see some users registered their pets for weight loss. For instance, a user registered her cat whose start weight is 5.6 kg and goal weight is 4.5 kg. Though there are a whole range of people from 50 kg to 150 kg who want to lose weight, it is unclear how obese they really are as the website do not provide their BMI data.

Figure 8 shows the correlation between the users' start weight and the percentage of the weight they want to change, with y as the average of the points in the x buckets with size of 30. A positive y value means that the user wants to lose weight, while a negative value means to gain weight. It is clear that the larger the start weight is, the more likely the user wants to lose weight, and the larger the percentage the user wants to lose. The median goal of all the public users was to lose 15% of the start weight. There were 32 users whose start weight was less than 30 kg, and 11 of them wanted to increase weight. Several of these users wanted to increase their weight by several folds, which seems like they were experimenting with the website.

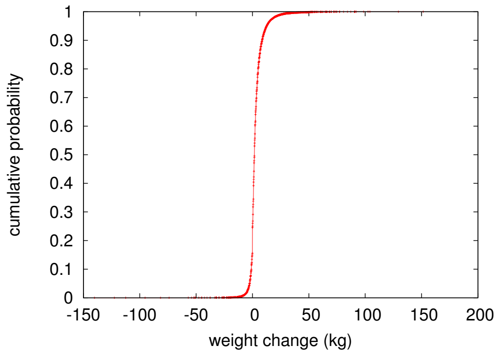


Figure 9: CDF of weight changes.

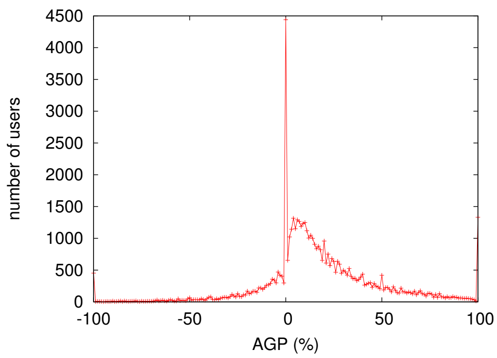


Figure 10: Users reaching different AGP.

We define the user's weight change (WC) as follows. The absolute value of WC is the user's current weight minus start weight, and the sign of the WC is determined by whether this user is on track (Section 5.1). Namely, if the user wants to lose weight and his current weight is less than his start weight, then his WC is positive (meaning that he is making desired progress). Otherwise, the sign of his WC is negative, meaning that he is making negative progress. Figure 9 shows the cumulative distribution of the users' WC, in which we removed the users who only recorded a single weigh-in. The distribution follows a normal distribution, and the median WC was only 1.75 kg. Clearly losing significant weight remains a great challenge.

We define a user's Achieved Goal Percentage (AGP) as her WC divided by the absolute difference between her start and goal weights. This attribute reflects how much the user has achieved her goal. We excluded one-time weigh-in users and steady users. The distribution of AGP is shown in Figure 10. Here in the plot the points are grouped together for x values rounded to the closest integers, and points with x less than -100% and greater than 100% are grouped together, respectively. There are about 4500 users who made marginal progress towards their goals. The mean was 18.1% and the median is 13%, reflecting current FatSecret users' overall weight-change performance. Note that there were 1331 users who successfully reached their goal (AGP is 1 or greater).

We also calculated the maximum AGP a user had ever achieved and compared it with her current AGP, and the difference reflects the setback towards her weight goal. The

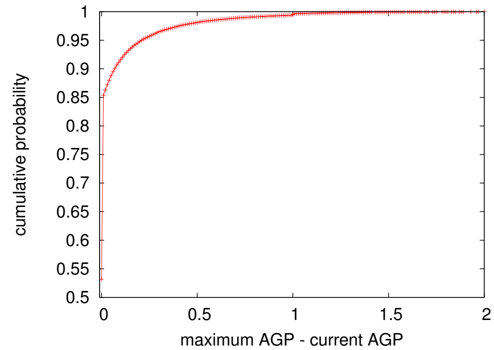


Figure 11: CDF of weight-change setbacks.

Figure 11 shows the cumulative distribution of the setbacks, suggesting that more than half of the users had no setbacks (maximum AGP was the same as current AGP), or they did not want to report. Interestingly there were 99 users who once achieved their goal (maximum AGP was 1), but then completely reversed course and their current weight was even worse than their start weights towards their goal.

6. CORRELATED ANALYSIS

In this section, we study how the users' weight changes correlate with other attributes. With this knowledge, we may be able to better understand and even proactively manipulate health OSN services to influence the users' health behaviors for better results.

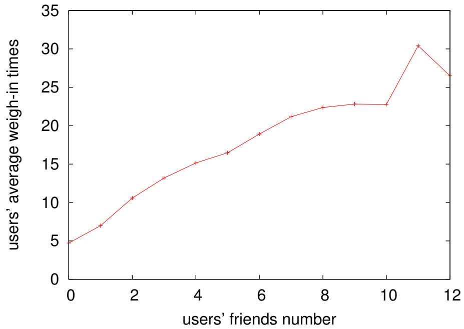
The user attributes we study include the number of friends, start weight, current weight, goal weight, weight change (WC), weight change percentage (WCP), goal change percentage (GCP), achieved goal percentage (AGP), and the number of weigh-in times since registration. Here we define GCP to be the percentage a user wants to change the weight, as the difference of start and goal weights divided by the start weight. WCP is the percentage of the weight change the user has already made, as the difference of the user's current and start weights divided by her start weight. AGP was defined before as the current weight change divided by intended weight change.

6.1 Correlation with own attributes

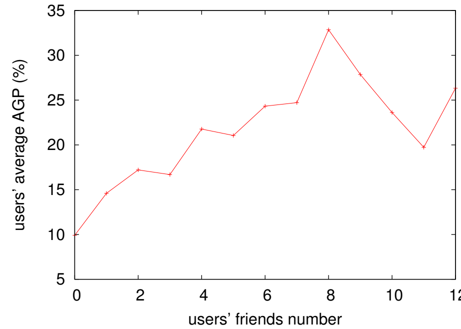
We grouped all the public users based on their number of friends and calculated the averages of the attributes to study in each group. We found that most of the attributes increase proportionally with the friends number. For example, Figure 12(a) shows that the more friends a user had, the more weigh-ins he recorded. This is a positive behavior as weigh-ins increase the user's own awareness and accountability, and will likely keep the user motivated for weight change.

Figure 12(b) shows that the more friends a user has, the more progress she has made towards her goal. This positive correlation suggests that more social support, even just in cyberspace, may encourage the users to get more aggressive on their weight-change progress. Similarly, Figure 12(c) shows that the more friends a user had, the more likely she had made a larger weight change percentage.

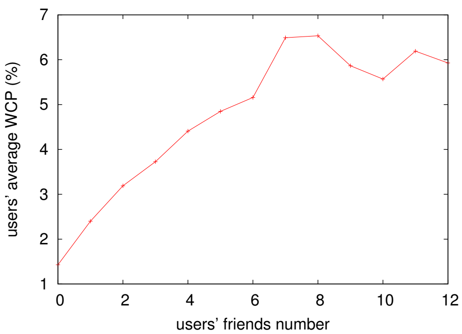
On the other hand, Figure 12(b) and 12(c) suggest that the weight-change benefits of having more online friends



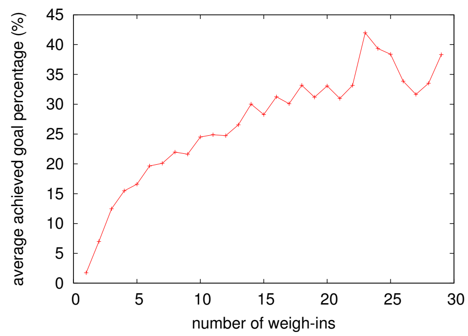
(a) Friends Number and Weigh-Ins



(b) Friends Number and AGP



(c) Friends Number and WCP



(d) Weigh-Ins and AGP

Figure 12: Correlation with the users' own attributes.

become less obvious when the user has already had eight friends. AGP even decreases when the users have more than eight friends, while WCP flattens. This may be because a user is likely to have some negatively influencing friends when he adds more users as his friends.

We then grouped the users based on the number of weigh-ins they have recorded, and calculated the average of the attributes to study in each group. We found that the values of WC, GCP, WCP, AGP all increase with the number of weigh-ins, as shown in Figure 12(d) for AGP (other results are similar and not shown). This confirms with our intuition that the more weigh-ins a user makes, the more effective he is achieving the weight-loss goal. Such positive correlation between the user awareness of her own progress and the improved results has also been documented in other studies [6, 11].

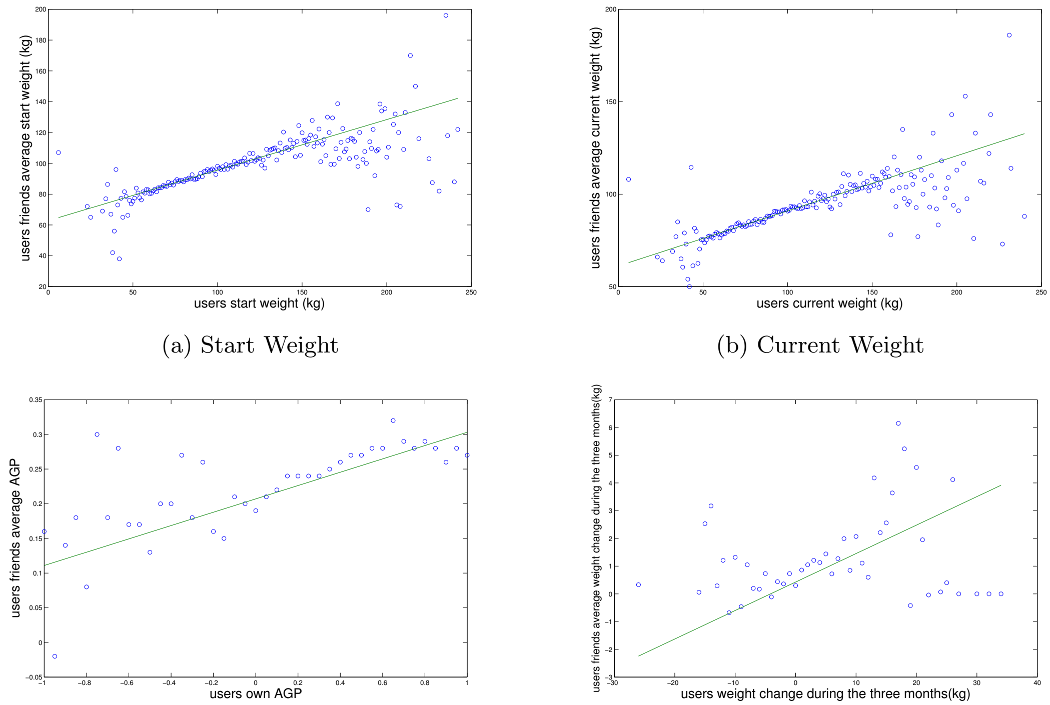
6.2 Correlation with friends' attributes

Intuitively we know that friends are alike: they likely grew up together, share similar education background, like same food, and do some activities together. In this section, we study the correlation between the users and their friends' weight-change behaviors. We chose all the public users who have at least one non-private user as their friend to be the subjects of this study, which gave us 16,883 users in total.

We retrieved all the 16,883 users' start weight, and rounded their weight into integers (e.g. 68.6 kg is rounded to 69 kg). Then these users' friends' average start weight is computed. We used the users' own start weight as the x-axis, and their friends' average start weight as the y-axis. This, however, may lead to the situation that the same x value has several different corresponding y values, which means that some users may have the same start weight, but their friends' average start weight are not the same; thus there will be multiple y values corresponding to a single x value. Although the correlation can be shown in such a graph, to make it more clear, we took a further step to compute the average on each x value, and the results are shown in Figure 13(a).

It can be easily seen that there is a strong correlation between the users' start weight and their friends' start weight. We fit these points with linear regression and got the green line $y = 0.3280x + 62.7966$. The correlation coefficient is 0.9469. The correlation does not look strong when the x value is beyond the range of [50:150], but in fact there are less than 200 users outside of that range, causing larger variations. The goal weight correlation chart is similar to the start weight chart, with the x shift a bit to the left (not shown). Based on these results we can see that the users with larger start weight are more likely to be friends with those whose start weight is also large. Similarly, the users whose goal weights are close tend to become friends on FatSecret.

Figure 13(b) shows the correlation between the users and their friends' average current weight. This graph looks similar to the one for start weight, but the implications are different. There are two possible reasons for the correlation of current weight. First, the users who have similar current weight are more likely to be friends because they may join the same groups, such as "Get under 200" group for members wanting to reduce their weight to be less than 200 pounds. Second, the users with similar start weight have already become friends and they keep more or less similar pace of losing or gaining weight. Thus their current weight will also be correlated. While it is not clear whether these



(a) Start Weight

(b) Current Weight

(c) Achieved Goal Percentage

(d) Three-Month Performance

Figure 13: Correlation with the users' friends.

two are the only reasons or which one is the main reason for friends to share similar current weight, our other analysis shows that friends do tend to maintain similar weight change pace.

As previous results show that the users and their friends have strong correlations on their start weight, goal weight, and current weight, it is expected that their WC and GCP, calculated from these three values, will also likely have strong correlations. Our results confirmed these expectations (Figures not shown here). When a user's WC is in the range from -10 to 40 kg, she and her friends have strongest weight change correlation (there are 99.5% of the users in this correlation range).

Achieved Goal Percentage (AGP) is a better metric than WC to gauge the users' weight change progress, as it may directly reflect the users' satisfaction on their progress towards their goals. Figure 13(c) shows the AGP correlation of the users and their friends. The correlation is strongest when x ranges from -20% to 50%. Despite the variations of the values outside of that range, the correlation coefficient is still higher than 0.8.

Since the users have different registration time, simply comparing their WC or AGP may not seem to be fair. So we selected all the 11,301 users who have at least one public friend on February 20 and remained active till May 20, and we tracked their weight changes during these three months. The performance of the friends' weight changes over the same period still have strong correlations as shown in Figure 13(d). We can ignore the users whose x value is less than -10 or greater than 20, as there are just 29 and 12 respective users. The plot shows that there is a linear increase for x from -10 to 20, and a higher rise for x greater than 15. This result further shows the strong weight-change influence among friends.

6.3 Weight change propagations

Previous results show that the users and their friends' weight changes are strongly correlated. In this section we use statistical methods to quantify the probability of a user's weight-change level given her friends' performance. Here we focus on the AGP attribute, and we define a threshold T as the boundary condition to consider whether a user has made significant weight-change progress. If a user's AGP is greater than T, we consider that she is a "success." We can of course vary the values of the T to see its impact on the results. For example, take 30% as the threshold. The users whose AGP greater than 30% are considered to have successfully changed significant weight towards their goals.

What is the probability for a user to be successful if we know one of her friends have succeeded in weight change, according to T? For comparison purposes we can construct a synthetically generated network as the baseline, and the network consists of the same number of nodes and edges as the actual FatSecret social network. Also we randomly assign the same number of success and non-success nodes in the synthetic network. We then calculated the empirical probabilities of a node to be a success if one of its neighbors is a success in both actual and synthetic networks. We then computed the ratio of these two probabilities. We varied the threshold T from 10% to 100% with a 10% interval to determine "success." We generated 1000 synthetic networks for each threshold value, and computed the mean value of the probabilities of these 1000 networks. The 1000 values are also sorted from low to high and we took the 26th and 975th

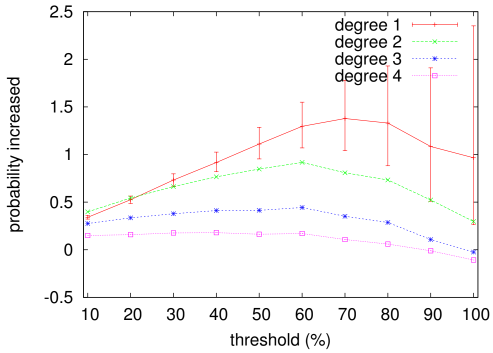


Figure 14: Weight-change influence propagation with different thresholds.

values as the 95% confidence interval. A similar technique is used for studying obesity spread in a real-world social network [3].

Figure 14 shows the probability ratio results (the curve marked with degree 1, comparing the users and their direct friends). When the threshold is set to 60%, the probability of a user is a success, when one of his friend is a success, is 106.9% higher than in a synthetic graph. As all friendships are mutual, we can also say the probability of a user's friend is a success, when this user is a success, is 106.9% higher than a synthetic network. We found that with the threshold increases, the confidence interval becomes wider as there are less and less success nodes that lead to the randomness of generating an edge between a success node and another success node. Also we can observe an ascending and then a descending trend of the curve when the threshold increases. The reason is that when the threshold is small, there are a lot of success users so that it is easy to find a success user in the synthetic graph who connects with another success user. With the increase of the threshold, the number of success users becomes smaller, leading to the chance of finding a success-to-success edge becoming smaller and smaller. On the other hand, due to the clustering effect (friends have correlated AGP) of the observed actual social network, it is still relatively easier to find such edges than in a synthetic network. So the difference of the probabilities between the two networks increased. When the threshold is large enough, however, the number of success users in the observed actual network also dropped quickly, thus causing a declining tail.

Another interesting question is how far the weight-change influence between friends can propagate in the social network. For example, if a user loses weight very fast, his friends may also likely lose weight relatively fast as they are influenced by that user. Then that user's friends' friends (2 degrees away from the original user) may also be influenced by that user's friends, so that they may also lose weight faster than others. How many degrees can weight-change influence ripple through the social network? Using the same technique described above we can also calculate the probability of a user is a success, when one of her 2-degree friend is a success, in both networks and compute the ratio. Similarly we can do this for degree 3, 4, and so on. If the probably ratio between the actual and synthetic networks for degree d becomes too small, we consider the influence has exhausted at this degree.

Figure 14 shows that a user's two-degree friends will have at most 91% higher possibility, when the threshold is set to 60%, of being a success than in a synthetic network, when the user himself is a success. A user's three-degree friends will have at most 44% higher possibility, when the threshold is set to 60%, of being a success than in a synthetic network when the user himself is a success. The probability of friends being success in fourth degree of separation of a user is around 15% higher than in a synthetic network, and becomes only 5% for the fifth-degree friends. Thus we say the weight-change influence's propagation distance is at most four degrees. This is different than the results obtained by Christakis and Fowler [3], in which they show the weight-change influence in a real social network reaches the 3rd degree. It is interesting to see that the influence goes further in an online social network, at least for FatSecret, than a real-world one, probably because the FatSecret network consists of like-minded users and their weight changes are highly correlated as previous results show.

We also evaluated the dynamic extent of the social influence using AGP. Unlike the static association we studied above, here we want to find out if one of a user's friends *becomes* successful from non-successful, what is the probability increase for the user herself also *change* from non-success to success. To calculate this probability increase, a logistic regression model with time-lagged measurement is used.

Similar to the Christakis and Fowler study [3], we also assume a user's current success is related with her previous success, her friend's current success and her friend's previous success. For each friend the user has, we let

$$self_c = \beta_0 + \beta_1 * friend_p + \beta_2 * self_p + \beta_3 * friend_c$$

where $self_c$ and $self_p$ are the current and previous success of the user herself (values to be 0 or 1). $friend_c$ and $friend_p$ are the current and previous success of the user's friend. β_0–β_3 are effect parameters [7]. Based on the dataset stated above, we estimated the coefficient covariance matrix $[\beta_0, \beta_1, \beta_2, \beta_3]$. By drawing 1000 random sets of estimates from this matrix, we calculated 95% confidence intervals by simulating the first difference in the $friend_c$'s contemporaneous success (changing from non-success to success) and assuming mean values for all other variables [7]. The result is shown in the Figure 15. When the threshold is set to 50% or 60%, the chance of a user to become successful appeared to increase by over 60% (95% confidence interval is from 34 to 90) if one of her friends became successful. When the threshold is set to 30%, 40%, 70% or 90%, the chance of a user to become successful increased over 25% if one of his friends became successful. Also, the logistic model showed a user's current success is proportional to his friends' current success while inversely proportional to their previous success.

7. DISCUSSIONS

Our analysis results show that the more weigh-ins recorded by the users, the more likely they will succeed in weight changes. FatSecret currently uses periodic emails to remind its users to weigh-in if they have not done so for some time. While we do see some users respond to such reminders, it could be more effective to deliver such reminders to the users' mobile phones. It will also make it much easier for the users to weigh-in if the website can connect with Internet-based body scales, such as the Withings devices, which can automatically send the users' weight to a backend server

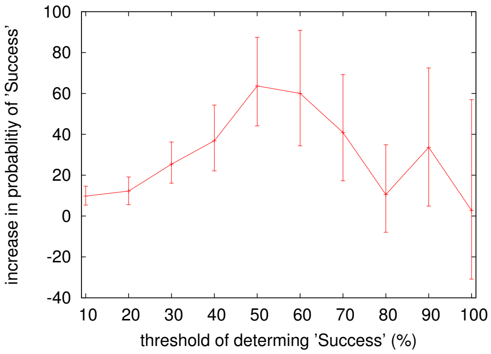


Figure 15: Success probability increases with friends becoming successful.

without any manual effort. Currently FatSecret has a default setting that shares the users' weight with everyone, which may make some weight-conscious users feel uncomfortable. Changing the default weight sharing as private or friends only may be another way to reduce obstacles for the users to weigh-in.

Like real-world friendships, online social connections also appear to have great influence on the users' health behaviors. Our results show that with support from similar weight users who want to achieve similar weight-change goals, a user is much more likely to make positive progress. If a user's friend makes significant progress on weight loss, the probability of that user's own weight-loss success is much greater than a user whose friends are not making progress. Thus a health OSN provider can segment its users into groups based on their weight and goal levels, identify those who are making good progress in these groups, and encourage them with incentives to be more active and to reach out befriending with other users. These "influencers" will be visible in the whole network and their weight loss progress will have rippling effects and change behaviors for a large portion of the network. For new users, the health OSN provider could be more proactive encouraging them making friends on the website, by suggesting them to other friends who may like similar foods and exercise routines.

The effectiveness of using an OSN for health management can be viewed that the OSN users accumulate the *social capital* over time [4], including knowledge sources and supporting relationships, which helps keeping the users motivated and engaged. On the other hand, sharing health information through OSNs raises privacy issues [1, 13]. The OSN operators must take careful steps to ensure the participants' privacy. In addition, we do not claim that social capital is the only factor influencing weight changes. For example, it is well known that smoking behavior has strong correlation with weight changes [3]. The data we collected simply does not have such information so that we can not perform such analysis.

8. CONCLUSION

In this paper we present an empirical analysis of a modern health-based online social network (OSN). To the best of our knowledge, our five-month study with data collected from more than 107,000 users is the first large-scale analysis of a modern health OSN. We found that a whole range of users whose current weight from 50 to 200 kg were interested in losing weight, with several popular goal weights. The number of weigh-ins and the number of friends correlate positively with the users' weight-change progress. If a user's friend makes significant progress on weight loss, the probability of that user's own weight-loss success is much greater than a user whose friends are not making progress. A user's weight loss may positively influence those separated as far as four-degree away in the social network. These results suggest that the OSN provider should use a variety of mechanisms to encourage users to regularly weigh-in and strengthen the social connections, which will help its users manage their weight more effectively.

9. REFERENCES

[1] A. Acquisti and R. Gross. Imagined communities: Awareness, information sharing, and privacy on the facebook. In *Privacy Enhancing Technologies (PET'06)*, 2006.

[2] A.-L. BarabÃąsi. The origin of bursts and heavy tails in humans dynamics. *Nature*, 435:207–211, 2005.

[3] N. A. Christakis and J. H. Fowler. The spread of obesity in a large social network over 32 years. *The New England Journal of Medicine*, 357:4, July 2007.

[4] N. B. Ellison, C. Steinfiels, and C. Lampe. The benefits of facebook "friends": Social capital and college students' use of online social network sites. *Computer-Mediated Communcation*, 12(4), 2007.

[5] N. Hellmich. Rising obesity will cost u.s. health care $344 billion a year. *USA Today*, November 2009.

[6] M. Hitti. Keeping food diary helps lose weight. *WebMD Health News*, July 2008.

[7] G. King, M. Tomz, and J. Wittenberg. Making the most of statistical analyses: Improving interpretation and presentation. *American Journal of Political Science*, 44(2):341–355, April 2000.

[8] N. Li and G. Chen. Analysis of a location-based social network. In *International Symposium on Social Intelligence and Networking (SIN)*, Vancouver, Canada, August 2009.

[9] D. Maloney-Krichmar and J. Preece. The meaning of an online health community in the lives of its members: Roles, relationships and group dynamics. In *Proceedings of the International Symposium on Technology and Society*, 2002.

[10] D. Maloney-Krichmar and J. Preece. A multilevel analysis of sociability, usability, and community dynamics in an online health community. *ACM Transactions on Computer-Human Interaction*, 12(2), June 2005.

[11] M. McClusky. The nike experiment: How the shoe giant unleashed the power of personal metrics. *Wired*, June 2009.

[12] A. Mislove, M. Marcon, K. P. Gummadi, P. Druschel, and B. Bhattacharjee. Measurement and analysis of online social networks. In *Proceedings of the 5th ACM/USENIX Internet Measurement Conference (IMC'07)*, San Diego, CA, October 2007.

[13] J. S. Olson, J. Grudin, E. Horvitz. A study of preferences for sharing and privacy. In *Conference on Human Factors in Computing Systems (CHI'05)*, Portland, OR, 2005.