

Original Paper

# Factors Associated With Weight Change in Online Weight Management Communities: A Case Study in the LoseIt Reddit Community

Gisele Lobo Pappa<sup>1\*</sup>, PhD; Tiago Oliveira Cunha<sup>1\*</sup>, MSc (Comp Sci); Paulo Viana Bicalho<sup>1\*</sup>, BSc (Comp Sci); Antonio Ribeiro<sup>2,3</sup>, MD, PhD; Ana Paula Couto Silva<sup>1</sup>, PhD; Wagner Meira Jr<sup>1</sup>, PhD; Alline Maria Rezende Beleigoli<sup>2\*</sup>, MD, PhD

<sup>1</sup>Computer Science Department, Universidade Federal de Minas Gerais, Belo Horizonte, Brazil

<sup>2</sup>Internal Medicine Department, Faculdade de Medicina, Universidade Federal de Minas Gerais, Belo Horizonte, Brazil

<sup>3</sup>Centro de Telessaúde do Hospital das Clínicas, Universidade Federal de Minas Gerais, Belo Horizonte, Brazil

\*these authors contributed equally

**Corresponding Author:**

Gisele Lobo Pappa, PhD

Computer Science Department

Universidade Federal de Minas Gerais

Av. Antônio Carlos, 6627

Pampulha

Belo Horizonte, 31270-010

Brazil

Phone: 55 3134097536

Fax: 55 3134095858

Email: [glpappa@dcc.ufmg.br](mailto:glpappa@dcc.ufmg.br)

## Abstract

**Background:** Recent research has shown that of the 72% of American Internet users who have looked for health information online, 22% have searched for help to lose or control weight. This demand for information has given rise to many online weight management communities, where users support one another throughout their weight loss process. Whether and how user engagement in online communities relates to weight change is not totally understood.

**Objective:** We investigated the activity behavior and analyze the semantic content of the messages of active users in LoseIt (r/loseit), a weight management community of the online social network Reddit. We then explored whether these features are associated with weight loss in this online social network.

**Methods:** A data collection tool was used to collect English posts, comments, and other public metadata of active users (ie, users with at least one post or comment) on LoseIt from August 2010 to November 2014. Analyses of frequency and intensity of user interaction in the community were performed together with a semantic analysis of the messages, done by a latent Dirichlet allocation method. The association between weight loss and online user activity patterns, the semantics of the messages, and real-world variables was found by a linear regression model using 30-day weight change as the dependent variable.

**Results:** We collected posts and comments of 107,886 unique users. Among these, 101,003 (93.62%) wrote at least one comment and 38,981 (36.13%) wrote at least one post. Median percentage of days online was 3.81 (IQR 9.51). The 10 most-discussed semantic topics on posts were related to healthy food, clothing, calorie counting, workouts, looks, habits, support, and unhealthy food. In the subset of 754 users who had gender, age, and 30-day weight change data available, women were predominant and 92.9% (701/754) lost weight. Female gender, body mass index (BMI) at baseline, high levels of online activity, the number of upvotes received per post, and topics discussed within the community were independently associated with weight change.

**Conclusions:** Our findings suggest that among active users of a weight management community, self-declaration of higher BMI levels (which may represent greater dissatisfaction with excess weight), high online activity, and engagement in discussions that might provide social support are associated with greater weight loss. These findings have the potential to aid health professionals to assist patients in online interventions by focusing efforts on increasing engagement and/or starting discussions on topics of higher impact on weight change.

**KEYWORDS**

obesity; online social media; weight loss; user behavior; topic modeling

## Introduction

Obesity is a major public health problem that adversely impacts morbidity, mortality, and quality of life. According to the World Health Organization (WHO), the prevalence of obesity has nearly doubled over the last 30 years [1]. According to a survey conducted by the Pew Research Center in September 2013 [2], of the 72% of American Internet users that have looked for health information online, 22% have searched for help to lose or control weight. These numbers show that people are seeking health advice and support online, especially in topic-driven communities and forums, Q&A sites (eg, Quora), and social media, including Twitter, Facebook, and Reddit [3,4].

Along with the growth of online searches for health advice, there has been a shift of social interactions from the real to the virtual world with the popularization of online social networks [5]. Most of these networks consist of communities organized according to user interests, including obesity and weight management groups. In contrast with real-world social groups, online social networks have no time restrictions, may be accessed from anywhere at no cost, and provide a constant source of information, support, and advice [6].

Among the factors that play a major role on the success of weight loss in online social networks are individual features and social embeddedness. Social embeddedness encompasses the structure and intensity of individual connections—reflecting different forms of engagement and participation in the community [3,5]—and it has been investigated across different weight management online communities [6-8]. Although social embeddedness has shown to strongly correlate with weight loss, most studies have mainly looked at the structural connections of the networks (ie, who connects to whom) [9]. Topics discussed by engaged users may also be as important as their engagement in the community. The linguistic style and semantics associated with the topics discussed might also be important features because they provide insights on personality, attitudes, and behavior, which in turn correlate with health outcomes [10]. However, most analyses look at syntactic features of the text, including word counts or the presence of positive or negative constructs [11,12]. Concerning the semantics of the messages, only a few qualitative studies focusing on small groups of users were able to analyze it. This is mainly due to the difficulties of automatically extracting semantics from text.

Analyses focusing on the behavior of users and the syntactic meaning of their texts are becoming common practice, not only in online health communities, but also in other contexts where user behavior in online social media can provide relevant information, including customer behavior [13] and citizens' engagement with politics [14]. For instance, many marketing-related studies have proposed different theoretical frameworks to help understand the behavior of customers in

virtual brand communities [15,16] or to identify the reasons for customers engaging online [17]. These frameworks and associated reasoning, despite being similar to ours, are not directly applicable to the health context. This is because the reasons for patients to engage on health communities are different from those of customers in a brand community and, most importantly, the role of the community toward users in both contexts differs significantly. Health online social network users are motivated mainly by having the opportunity to talk to users going through the same experiences without feeling pressure from society because obesity is still stigmatized. Furthermore, weight-loss communities present a large opportunity for ordinary users and health agents to perform both direct and indirect mental health and social support interventions, improving users' self-motivation to lose weight and increasing their self-esteem [18,19].

In this paper, we describe both the activity behavior and analyze the semantic content of the messages of the LoseIt (r/loseit) online Reddit weight-loss community and investigate whether these factors are associated with weight loss in this online social network using different inputs to a regression model.

## Methods

### Data Collection

Reddit is an online forum organized in subcommunities by areas of interest called subreddits. We used Python Reddit Application Programming Interface (API) Wrapper (PRAW) [20], a Python package that eases access to Reddit's official API, to collect the dataset from a Reddit subreddit called "LoseIt" [21], which is a community in which people interact about weight-loss issues [22]. Reddit users may submit content, such as texts or direct links to other sites, both collectively referred to as "posts." The community can then vote posted submissions up (upvotes) or down (downvotes) as a sign of "liking" or "disliking" it. Users can also reply to posts with comments. One interesting aspect of Reddit is the anonymity of user accounts. In this study, users were identified by their system log-in. Data on real-world characteristics (age, gender, anthropometric measures), online activity behavior, and linguistic style were collected in English-written posts, comments, and tags in LoseIt from August 2010 to November 2014 [23].

### Study Measures

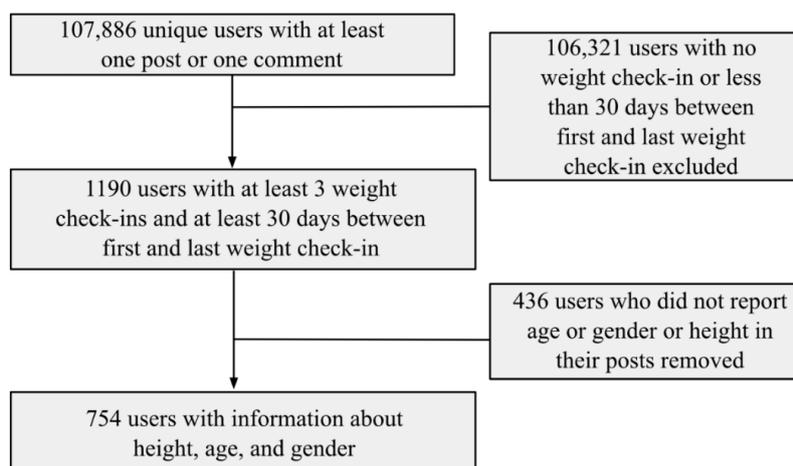
Self-reports of weight, known as weight check-ins, included start weight, current weight, and goal weight. To deal with multiple reports of start weight (65/1190, 5.46%) per user, we selected the ones with the longer interval between start weight and current weight. To be considered for analysis, this interval had to be at least 30 days in an attempt to minimize the effect of short-term weight variations not related to fat loss on our outcome [24]. For a subset of users with this data available (n=1190), we collected data on age, gender, and height, and

calculated body mass index (BMI) as  $BMI = \text{weight (kg)} / \text{height (m}^2\text{)}$ , obtaining a final sample of 754 users, as depicted in Figure 1. We categorized BMI according to the WHO criterion [25] of normal ( $BMI < 25 \text{ kg/m}^2$ ), overweight ( $25 < BMI < 30 \text{ kg/m}^2$ ), class I obesity ( $30 < BMI < 35 \text{ kg/m}^2$ ), class II obesity ( $35 < BMI < 40 \text{ kg/m}^2$ ), and class III obesity ( $BMI \geq 40 \text{ kg/m}^2$ ).

Measures of online activity were undertaken in relation to individual use (per user) and temporal activity. Data on the number of posts and comments (per time and per user), number

of upvotes received per post and comment, number of comments received per post, and the number of weight check-ins were collected. We also computed the lifespan of a user, defined as the period of time between his/her first and last activity, and assessed the proportion of days online (days when a user posted or commented) and the number of active weeks. For assessing temporal patterns of activities, we derived a weekly temporal series for each user life span in LoseIt that estimated user participation over time (ie, we computed his/her number of activities in that week for each week in the lifespan of the user).

**Figure 1.** Flowchart of the user selection process.



## Semantic Analysis of Text

After removal of special characters and stop words (the most common words in a language), we created a matrix of posts/comments per terms (words) using the term frequency [26]. This matrix was used as input to a semantic analysis of the messages through topics using latent Dirichlet allocation (LDA) with Gibbs sampling [27]. LDA assumes that there are hidden variables (topics) that explain the similarities between observable variables (posts or comments). The parameter for the number of topics was empirically defined as 50 and the method run for 2000 iterations. For each post/comment, LDA returns the probability of it being associated with a topic. Each topic, in turn, is represented as a set of words, in which each word is also associated with a probability of describing a topic. For each topic, we manually associated a label with the message given by its descriptive words and associated posts/comments.

## Statistical Analyses

Dataset characteristics were described by mean and standard deviation (SD), median and interquartile range (IQR), or frequencies.

We performed three linear regression models to investigate the association between weight change over at least 30 days and users' real-world characteristics (model 1: age, gender, BMI), activity behaviors (model 2: number of activities, comments received, upvotes received, number of weight check-ins, lifespan, percentage of days active, number of days active, number of days offline, number of weeks active, and if the user has a verified email), and semantics extracted from text (model

3: the 26 most-discussed topics of posts and comments selected from a set of 100 topics using the Akaike information criterion) on weight change (over at least 30-day interval). For model 3, summed and normalized probabilities returned by the LDA for each user's post or comment per topic were given as input to the regression model. Regression models were compared using the coefficient of determination ( $R^2$ ) and the adjusted  $R^2$ , which also accounts for the number of variables included in the model.

To evaluate the generalization power of the models, we executed a 10-fold cross-validation [28]. This method divides the data into 10 nonoverlapping folds and uses nine parts to fit the model and one (also called test set) to measure its generalization performance. This procedure is repeated 10 times, each time with a different fold as the test. At the end of this process, the mean and standard deviation over the 10 folds are reported, together with the correlation between the real and predicted values in the user test set. All data were analyzed using R (MASS and caret packages).

## Results

### Data Sample Analysis

Over the 4-year follow-up, 107,886 of 252,279 (42.76%) Reddit users were active (ie, users with at least one post or comment). Among these, 101,003 (93.62%) wrote at least one comment and 38,981 (36.13%) wrote at least one post. For posts, 1621 of 107,886 users (1.50%) contributed more than five times, suggesting that this group starts several discussions around weight-loss problems. These users were the ones who remained

active in the network for longer, with a median of 72 (IQR 78) active weeks versus a median of 1 (IQR 18) active weeks for the remaining users. We observed that more than 95.39% (64,281/67,387) of posts had at least one upvote. Considering

the top 5.00% (3370/67,387) most upvoted posts, they all received more than 100 upvotes, suggesting that some posts do call the attention of the community and receive its support. Characteristics of the dataset are shown in [Table 1](#).

**Table 1.** Overall characteristics of the LoseIt dataset.

Statistics	Median (IQR)	Mean (SD)
Posts per day	45.0 (30.0)	45.5 (22.7)
Posts per user	0 (1.0)	0.7 (1.8)
Comments per day	599.0 (333.0)	586.7 (264.3)
Comments per user	2 (4.0)	7.9 (34.4)
Upvotes per post	6.0 (16.0)	35.7 (126.7)
Upvotes per comments	2.0 (2.0)	3.1 (11.4)
Weight check-ins per user	2 (2.0)	2.7 (3.6)
Activities per user	2 (5.0)	8.5 (35.3)

For weight check-ins, 8834 of 107,886 (8.18%) users had at least one weight check-in. Among them, 6622 users declared a start weight, 6303 a current weight, and 3074 a goal weight. Among all users who declared a start weight, 65 (0.1%) declared it more than once, with 34 of 65 (52%) decreasing its value. The same happened with goal weights, which might indicate that 1.01% (31/3074) of the users reassessed their goal weight. In total, 19 of 31 users (62%) with more than one report reduced their initial goal weight, whereas 12 of 31 (38%) increased it.

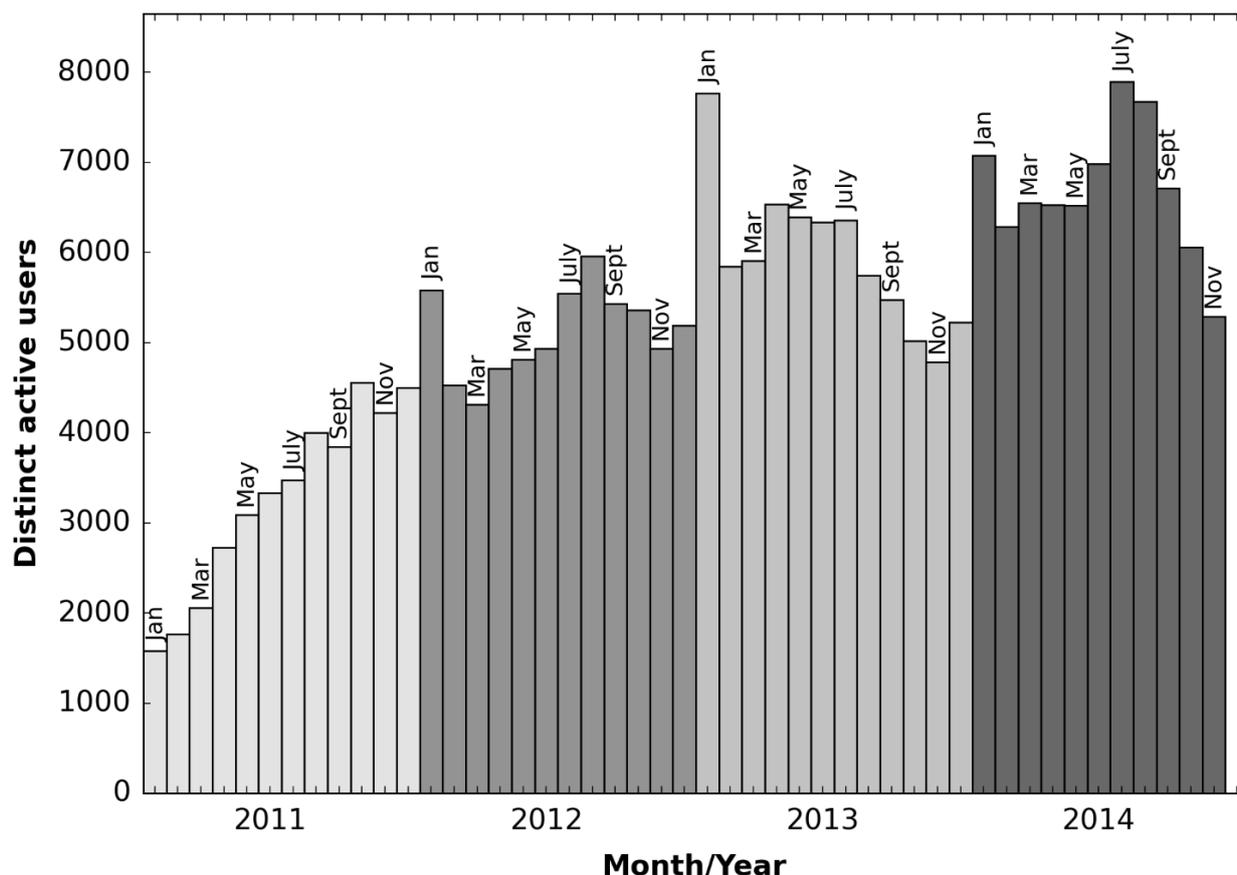
### User Online Activity

[Figure 2](#) shows the distribution of the number of distinct active users in the community over the follow-up period. Overall, the number of unique active users increased over time, with high values in January (motivated by New Year's resolutions) and July to August (northern hemisphere summer).

[Table 2](#) shows information about the user lifecycle. Considering the dates of the first and last activity, 50,705 of 107,886 (46.99%) users had a lifespan of one day, meaning they had a single activity in the network. Nevertheless, such short lifespans do not necessarily mean they left the community because they may have been simply acting as lurkers (ie, readers only). In contrast, 10,874 of 107,886 (10.08%) users had a lifespan longer than 1 year (449 days), and 543 of 107,886 (0.50%) reached 2 years (682 days). The longest life span was 1566 days (223 weeks). Disregarding the users who were active only once, for the 57,181 remaining users, the activity peak of 37,802 (66.10%) users was in week 1. Further, 45,403 of 57,181 (79.40%) users in the dataset reached their peak before week 10. In contrast, the peak for 3.54% (2024/57,181) of users was after week 40.

**Table 2.** Statistics about users' lifecycles on LoseIt.

Lifecycle	Median (IQR)	Mean (SD)
Life span (days)	2 (132.0)	124.60 (238.64)
Percentage of days online	3.81 (9.54)	11.11 (17.97)
Weeks with activities	1 (2.0)	3.05 (5.09)
Peak of activity (week)	1 (0.0)	5.35 (15.88)

**Figure 2.** Number of distinct active users on LoseIt per month. Different colors correspond to different years.

### Semantics Extracted From Text

Table 3 shows the 10 topics most frequently found in posts and messages, which are representations of the semantics of the posts. Note that by simply looking at the words describing topic 1, for instance, one might say they do not necessarily refer to healthy food, but the qualitative analysis together with the messages used allow us to state that. Also, notice that we may have more than one topic with the same label. This happens because the method used to extract the topics (LDA) does not guarantee they are unique. Examples of topics include information about workouts (topics 4 and 8) and the best ways

to control calorie intake (topics 3 and 5). The topics in the comments are not that different from those in posts, although a strong presence of user support is frequently found in comments. The following comment illustrates the type of message we were dealing with:

*Good work dude, I am in the same boat you are...I recently started strong lifts as well as the low-carb high-protein diet it's awesome and I rarely feel hungry. Keep it up dude and when you feel comfortable post some pics. When I hit my first milestone of 50 lbs I will be posting some...hopefully soon down to go.*

**Table 3.** Most frequently discussed topics in the LoseIt community for users' (N=38,981) posts (n=67,387) and users' (N=101,003) comments (n=771,146).

Given label	Top descriptive words	Posts, n (%)
<b>Posts</b>		
1. Healthy food	Chicken, lunch, diner, breakfast, salad	3489 (5.18)
2. Clothing	Size, fit, clothes, shirt, pant	2873 (4.26)
3. Calories count	Calorie, day, counting, per, intake	2550 (3.78)
4. Workout	Run, running, mile, minute, walk	2305 (3.42)
5. Calories count	Using, use, mfp, track, myfitnesspal	1976 (2.93)
6. Looks	Fat, body, muscle, stomach, skin	1952 (2.89)
7. Habits	Water, drink, soda, drinking, cut	1933 (2.87)
8. Workout	Cardio, workout, minute, training, lifting	1875 (2.78)
9. Support	Loseit, thank, everyone, post, thanks	1831 (2.72)
10. Unhealthy foods	Food, eat, pizza, one, ate	1803 (2.67)
<b>Comments</b>		
1. Support	Great, look, awesome, job, amazing	44,829 (5.81)
2. Food	Chicken, cheese, salad, vegies, cup	38,960 (5.05)
3. Clothing	Size, fit, skin, clothes, loose	34,155 (4.43)
4. Calories count	Calorie, day, counting, eat, deficit	21,458 (2.78)
5. Calories count	Use, scale, track, using, mfp	21,081 (2.73)
6. Weight control	Week, pound, month, lost, two	20,610 (2.67)
7. Self-esteem	See, look, picture, progress, difference	20,588 (2.67)
8. Workout	Run, running, minute, mile, walking	20,450 (2.65)
9. Support	Feel, like, better, feeling, much	19,249 (2.49)
10. Testimony	Self, not, problem, control, issue	19,232 (2.49)

### Understanding Weight Change

As showed in [Figure 1](#), data on real-world characteristics were only available for 754 users; hence, the investigation of the factors associated with weight change is limited to this subsample. In this sample, 56.5% (426/754) of users were female and the mean age was 26 (SD 6) years. The mean

difference between the start and goal weights was 35.6 (SD 24.8) kg; 3.7% (28/754) of users gained weight (mean 3.88%, SD 4.04), 3.5% (25/754) maintained weight, and 92.9% (701/754) lost weight. Overall, 514 of 754 users (68.2%) moved to a healthier category while participating in the LoseIt community as depicted in [Table 4](#).

**Table 4.** User distribution according to initial and final obesity categories.

Initial category	Final category, n					Total
	Normal	Overweight	Obese I	Obese II	Obese III	
<b>Total</b>						
Normal	25	2				27
Overweight	70	58	1			129
Obese I	37	105	46	1		189
Obese II	10	52	76	32	1	171
Obese III	4	36	61	63	74	238
<b>Female</b>						
Normal	23	1				24
Overweight	60	39	1			100
Obese I	21	65	23	1		110
Obese II	8	23	35	11	1	78
Obese III	2	14	30	27	41	114
<b>Male</b>						
Normal	2	1				3
Overweight	10	19				29
Obese I	16	40	23			79
Obese II	2	29	41	21		93
Obese III	2	22	31	36	33	124

Among weight losers, 350 lost up to 14.2% of their start weight and the top 10.0% (70/701) of weight losers lost between 29.4% and 57.7% of their start weight. Median weight loss was 41.6 (IQR 41.4) kg. In general, weight losers were active users; on average, they contributed mean 3.55 (SD 6.11) posts, made mean 59.21 (SD 164.01) comments, and received mean 50.26 (SD 106.05) comments.

Table 5 shows the regression coefficients found in the models for different variables along with their standard errors (SE) and

*P* values. A positive coefficient indicates the variable is correlated to an increase in weight loss, whereas a negative coefficient indicates the variable is related to a decrease in weight loss. The *P* value for each term tests the null hypothesis that the coefficient is equal to zero (no effect). A low *P* value (<.05) indicates that you can reject the null hypothesis (ie, the predictor is likely to be a meaningful addition to the model because changes in its value are related to changes in the response variable).

**Table 5.** Variable coefficients (and their standard errors) found by three linear models accounting for different factors associated with percentage weight loss.

Variable	Model 1		Model 2		Model 3	
	Coefficient (SE)	<i>P</i> value	Coefficient (SE)	<i>P</i> value	Coefficient (SE)	<i>P</i> value
<b>Real world</b>						
BMI	0.517 (0.090)	<.001	0.415 (0.090)	<.001	0.371 (0.044)	<.001
Gender	-1.357 (0.140)	<.001	-1.979 (1.570)	<.001	-1.471 (0.781)	<.001
Age	0.235 (1.630)	.88	0.126 (0.140)	.36	0.118 (0.069)	.09
<b>Online behavior</b>						
Activity			-0.008 (0.005)	.11	-0.006 (0.001)	.01
Comments received			0.120 (0.007)	<.001	0.005 (0.003)	.12
Upvotes received			0.002 (0.001)	.04	0.002 (0.000)	<.001
Number of weights			0.270 (0.120)	.02	0.288 (0.058)	<.001
Weeks active			0.008 (0.070)	.90	0.089 (0.033)	<.001
Off days			-0.004 (0.002)	.04	-0.004 (0.001)	<.001
<b>Post topics</b>						
Weight control apps					-57.26 (26.68)	.03
Self-esteem (1) <sup>a</sup>					118.50 (24.49)	<.001
Self-esteem (2) <sup>a</sup>					48.03 (27.47)	.08
Workout					37.22 (20.05)	.06
Friendship					83.72 (28.31)	.003
Health information					44.84 (31.49)	.15
Weight goals					77.90 (26.99)	.004
Feelings					-125.00 (37.61)	<.001
Asking for advice					68.97 (44.70)	.12
Motivation					68.51 (52.50)	.009
Weight check-ins					42.02 (26.25)	.10
Family					-47.16 (30.94)	.12
<b>Comment topics</b>						
Body transformations					112.60 (59.71)	.06
Asking for help					235.70 (96.420)	.01
Body measures					-290.90 (108.20)	.007
Changing lifestyle					-219.20 (141.40)	.12
Counting calories					187.40 (71.55)	<.001
Asking for support					-134.20 (63.51)	.03
Friendship					223.40 (105.20)	.03
Workout					295.30 (72.43)	<.001
Motivation (1) <sup>a</sup>					304.80 (149.60)	.04
Weight goals					243.00 (66.50)	<.001
Motivation (2) <sup>a</sup>					251.80 (100.50)	.01
Healthy foods					173.70 (91.41)	.05
Gratitude					160.20 (72.06)	.02
Starting again					-231.40 (106.10)	.02

Variable	Model 1		Model 2		Model 3	
	Coefficient (SE)	<i>P</i> value	Coefficient (SE)	<i>P</i> value	Coefficient (SE)	<i>P</i> value
Lifestyle change					221.00 (99.84)	.03

<sup>a</sup> Topics generated by latent Dirichlet allocation may overlap. The posts and comments for both self-esteem and motivation appear in two different topics, here indicated by (1) and (2) after their names.

Regarding the association between weight change and real-world characteristics, online activity behavior, and the semantics from text, we found that a higher initial BMI, higher number of weight check-ins and upvotes, the amount of time spent in the community, and the topics of comments and posts (self-esteem, friendship, motivation, asking for help, counting calories, workout, gratitude) were directly and independently associated with weight loss. Conversely, some topics (weight control apps, feelings, body measures, asking for support) were inversely associated with weight loss.

The addition to variables related to online activity behavior and topics to the model led to an improvement on its ability to explain weight loss (from  $R^2=.18$  in model 1 to  $R^2=.26$  in model 2 and from  $R^2=.26$  in model 2 to  $R^2=.39$  in model 3, respectively). The adjusted  $R^2$  did not change for models 1 and 2, and decreased from .39 to .36 in model 3. As assessed by the cross-validation procedure, our final model showed generalization power and robustness with an  $R^2$  mean of 0.39 (SD 0.01) and a weight loss correlation of 0.83 (SD 0.01).

## Discussion

Knowing about characteristics, experiences, and interests of users might be of great value for improving weight-loss strategies in online communities. Our analyses have shown that the LoseIt users' activities are usually concentrated in their first 10 weeks in the community, when they are probably most motivated with the weight-loss process. Most of the active users reported that they lost weight, with approximately 70% moving to a lower obesity category. In addition, we found that users who reported weight change within a 30-day minimum interval were predominantly women, who declared themselves as overweight or obese. Female gender, BMI at baseline, high levels of online activity and of support (as measured by upvotes), and specific topics discussed within the community independently determined weight loss.

The predominance of women in comparison to men in our study may reflect both the higher prevalence of overweight and obesity among women in the general population [29] and a greater awareness of excessive weight among women in comparison to men [30]. Moreover, it may reflect the fact that women are more likely to pursue weight control [30,31] and to use the Internet to search for health information than men [32]. However, male sex was associated with greater weight loss. In real-world interventions, gender differences in success rates following short-term weight-loss programs are controversial [33], despite males' higher metabolic and lipolytic rates [34]. In practice, these differences in gender composition of online communities might suggest that several dimensions of weight history that are known to be different between men and women

should be considered for customizing apps for weight loss. For instance, the online community can be programmed to approach emotional eating (in response to mood) and social eating (eating in social situations) differently according to the gender of the user because emotional eating has been reported to be more common in women and social eating more common in men [33].

In this study, BMI was positively associated with weight loss. Because weight and BMI were self-reported, we might imply that the higher the baseline BMI, the greater the perception of excessive weight and/or dissatisfaction with weight status. Weight perception accuracy and dissatisfaction with weight status have been consistently associated with trying to lose weight and better weight control among obese and overweight adults in population-based studies [35]. However, due to the cross-sectional nature of our study, it is not possible to establish causality between self-declaration of high BMI and successful weight loss.

The number of weight reports, which may be a proxy of regular weight automonitoring, was directly associated with weight loss. Regular self-weighing has been reported as a useful tool for weight loss because it might provide feedback on energy balance status and consequent improvement in self-regulation [36]. It is not possible to establish a causal link between more frequent reporting and losing weight, but because individuals with greater weight loss might be more motivated to report their weight, it has been reported that self-reports of weight are more accurate among those who lose weight than among those who do not [37].

Various measures of online activity were independent predictors of weight loss. Higher participation levels in the online social network might unveil higher levels of self-motivation, which has been associated with better weight-loss outcomes following real-world interventions as well [38]. Therefore, developing strategies that improve self-motivation and maintain users active in the online community should be a cornerstone of Web-based weight-loss apps. These strategies might involve a more personalized approach with a deeper understanding of the uniqueness of the situation and needs of each user. In contrast with real-world situations, it has been reported that, among patients who seek health care for weight loss, the perception of the individual's needs by health professionals was associated with higher rates of weight loss and adoption of healthy lifestyle habits [39]. Understanding that users have different demands in regard to the various types of social support [40], and developing computational algorithms to recognize these differences, might be key points to offer personalized assistance and enhance users' activity in online social network.

Patients report that nonjudgmental and empathic interactions with health professionals are key points to achieve success in

weight-loss programs in the real world [41]. Upvotes in online communities might lead to a sense of social belonging and improve perceived empathy of the user within the community [40]. This probably explains the positive association we found between the number of upvotes and weight loss. A satisfactory interaction between users can be accomplished in an online social network by moderation of posts and comments that might be stigmatizing or demotivating and implementation of rating systems that allow users to rate the usefulness of posts and comments.

The finding that adding topics discussed in posts and comments to the model that explains weight change led to a great improvement in the coefficient of determination of the model suggests that an online community should provide different types of social support to users to be effective because most topics reflect some kind of support. Social support in health communities has been shown to be associated with maintenance of health behavior change [42] and better weight-loss outcomes in real-world interventions [42,43]. Most of the topics that were significant determinants of weight loss can be associated with three of the four types of social support. Emotional (ie, provision of empathy) and/or appraisal (ie, provision of information that is useful for self-evaluation purposes) support might have been provided to users in posts and comments that discussed self-esteem, motivation, and friendship, for example. Informational (ie, provision of advice, suggestions, and information that may be used to solve problems) and appraisal support might have been provided by discussions regarding counting calories, body transformations, workout, healthy food, lifestyle changes, asking for help, and gratitude. Implementing computational strategies to recognize the topics discussed by users and making active efforts to maximize the engagement of users in forums that address these topics might be effective tools in online weight-loss programs.

A few topics we expected to show positive association with weight loss, such as weight control apps, body measures, feelings, and asking for support, showed a negative association to it. Regarding weight control apps, we can infer that users who talked about them in LoseIt might not have followed the interventions of the app appropriately because they may not have decided yet which weight control app to use. Concerning body measures, users talking about it might have been frustrated by their results or might have set unrealistic weight goals. Both conditions might be associated with less successful weight loss. Users who talked about feelings or asked for support probably

did not feel appropriately supported, which might be related to unsuccessful weight loss.

We acknowledge some limitations of our study. The exclusion of less-engaged LoseIt users might have introduced a selection bias, which limits the generalization of the findings to active users of a closed online social network. Although it is not possible to exclude the possibility that unique users have multiple system log-ins, this issue probably did not influence our weight loss analyses due to the large sample size. Additionally, as in other studies with online communities, we relied on self-reports of weight, which may have reduced accuracy. The use of regular expressions for obtaining weight check-ins and real-world variables limited our data extraction and precluded us from having these data from the complete dataset. Furthermore, it was not possible to evaluate the influence of other factors that might impact LoseIt users' weight change, such as dieting and exercise planning. Due to the cross-sectional nature of the study, it is not possible to infer causality between the investigated factors and weight loss.

As major strengths, we must highlight the use of a quantitative methodology, which automatically looks at our measures of interest and allows for an analysis of a big sample, the long-term follow-up of the social network, and the investigation of the influence of real-world, online behaviors and semantics of online discussions on weight change reported by users of the online social network. We focus on a content-centered community instead of a user-centered one (ie, users in Reddit do not have connections to people but to posts), where the role of semantics may be even more accentuated. The semantics, analyzed with LDA, is a good way to summarize the topics being discussed as well as represent the interests of individual users.

In conclusion, our findings show that users who are more active and engage in discussions that provide emotional, informational, and appraisal social support are the ones with more successful weight loss in the online social network LoseIt. With increasing access worldwide to computers, apps, and the Internet, as well as the substantial amount of health care resources demanded by the high prevalence of obesity—particularly at young ages—online communities might be important public health strategies for obesity treatment and prevention. To potentialize the benefit of these communities, specific features that increase online activity should be investigated and incorporated to the online social network. Furthermore, activities that stimulate social support among users might be key points in the planning of online weight-loss interventions.

---

## Acknowledgments

This work was partially supported by LG Electronics, CNPq, Fapemig, and the Programa de Pós Graduação em Ciências da Saúde: Infectologia e Medicina Tropical, Faculdade de Medicina da Universidade Federal de Minas Gerais. The authors also would like to thank InWeb-Brazilian National Institute for the Web, and the Centro de Telessaúde do Hospital das Clínicas da Universidade Federal de Minas Gerais. Alline M Beleigoli was funded by Programa de Pós Doutorado from CAPES.

---

## Conflicts of Interest

None declared.

---

## References

1. World Health Organization. Global Health Observatory data: obesity. 2009. URL:[http://www.who.int/gho/ncd/risk\\_factors/obesity\\_text/en/Archived](http://www.who.int/gho/ncd/risk_factors/obesity_text/en/Archived) [accessed 2015-09-15] [WebCite Cache ID 6gFkthqAL]
2. Fox S, Duggan M. Health online 2013. Pew Research Center. 2013 Jan 15. URL:<http://www.pewinternet.org/2013/01/15/health-online-2013/> [accessed 2015-09-01] [WebCite Cache ID 6gFIZZiOD]
3. Eysenbach G, Powell J, Englesakis M, Rizo C, Stern A. Health related virtual communities and electronic support groups: systematic review of the effects of online peer to peer interactions. *BMJ* 2004 May 15;328(7449):1166 [FREE Full text] [doi: [10.1136/bmj.328.7449.1166](https://doi.org/10.1136/bmj.328.7449.1166)] [Medline: [15142921](https://pubmed.ncbi.nlm.nih.gov/15142921/)]
4. Laranjo L, Arguel A, Neves AL, Gallagher AM, Kaplan R, Mortimer N, et al. The influence of social networking sites on health behavior change: a systematic review and meta-analysis. *J Am Med Inform Assoc* 2015 Jan;22(1):243-256. [doi: [10.1136/amiajnl-2014-002841](https://doi.org/10.1136/amiajnl-2014-002841)] [Medline: [25005606](https://pubmed.ncbi.nlm.nih.gov/25005606/)]
5. Rad AA, Rahmandad H. Reconstructing online behaviors by effort minimization. In: Greenberg AM, Kennedy WG, Bos ND, editors. *Social Computing. Behavioral-Cultural Modeling and Prediction*. Berlin: Springer; 2013:75-82.
6. Turner-McGrievy GM, Tate DF. Weight loss social support in 140 characters or less: Use of an online social network in a remotely delivered weight loss intervention. *Transl Behav Med* 2013;3(3):287-294. [doi: [10.1007/s13142-012-0183-y](https://doi.org/10.1007/s13142-012-0183-y)]
7. Black LM, Bute JJ, Russell LD. "The secret is out!": Supporting weight loss through online interaction. In: Shedletsky L, Aitken JE, editors. *Cases on Online Discussion and Interaction: Experiences and Outcomes*. Hershey, PA: IGI Global; 2010:351-368.
8. Hwang KO, Ottenbacher AJ, Green AP, Cannon-Diehl MR, Richardson O, Bernstam EV, et al. Social support in an Internet weight loss community. *Int J Med Inform* 2010 Jan;79(1):5-13 [FREE Full text] [doi: [10.1016/j.ijmedinf.2009.10.003](https://doi.org/10.1016/j.ijmedinf.2009.10.003)] [Medline: [19945338](https://pubmed.ncbi.nlm.nih.gov/19945338/)]
9. Ma X, Chen G, Xiao J. Analysis of an online health social network. In: *Proceedings of the 1st ACM International Health Informatics Symposium*. 2010 Presented at: ACM International Health Informatics Symposium; Nov 11-12, 2010; Arlington, VA p. 297-306. [doi: [10.1145/1882992.1883035](https://doi.org/10.1145/1882992.1883035)]
10. Culotta A. Estimating county health statistics with Twitter. In: *Proceedings of the SIGCHI Conference on Human Factors in Computing Systems*. 2014 Presented at: 32nd Annual ACM Conference on Human factors in Computing Systems; Apr 26-May 1, 2014; Toronto, ON p. 1335-1344. [doi: [10.1145/2556288.2557139](https://doi.org/10.1145/2556288.2557139)]
11. De Choudhury M, Sushovan DE. Mental health discourse on reddit: self disclosure, social support, anonymity. In: *Proceedings of the International AAAI Conference On Weblogs And Social Media*. 2014 Presented at: International AAAI Conference On Weblogs And Social Media; Jun 1-4, 2014; Ann Arbor, MI p. 71-80.
12. Eschler J, Dehlawi Z, Pratt W. Self-characterized illness phase information needs of participants in an online cancer forum. In: *Proceedings of the Ninth International AAAI Conference On Weblogs And Social Media*. 2015 Presented at: Ninth International AAAI Conference On Weblogs And Social Media; May 26-29, 2015; Oxford, UK p. 101-110.
13. Brodie RJ, Ilic A, Juric B, Hollebeek L. Consumer engagement in a virtual brand community: an exploratory analysis. *J Bus Res* 2013;66(1):105-114. [doi: [10.1016/j.jbusres.2011.07.029](https://doi.org/10.1016/j.jbusres.2011.07.029)]
14. Zúñiga G, Jung H, Valenzuela S. Social media use for news and individuals' social capital, civic engagement and political participation. *J Comput-Mediat Comm* 2012;17(3):319-336. [doi: [10.1111/j.1083-6101.2012.01574.x](https://doi.org/10.1111/j.1083-6101.2012.01574.x)]
15. Hollebeek LD, Glynn MS, Brodie RJ. Consumer brand engagement in social media: conceptualization, scale development and validation. *J Interact Mark* 2014;28(2):1094-1065. [doi: [10.1016/j.intmar.2013.12.002](https://doi.org/10.1016/j.intmar.2013.12.002)]
16. Vivek SD, Beatty SE, Morgan RM. Customer engagement: exploring customer relationships beyond purchase. *J Marketing Theory Practice* 2012 Jul 2;20(2):122-146. [doi: [10.2753/MTP1069-6679200201](https://doi.org/10.2753/MTP1069-6679200201)]
17. Hennig-Thurau T, Gwinner KP, Walsh G, Gremler DD. Electronic word-of-mouth via consumer-opinion platforms: what motivates consumers to articulate themselves on the Internet? *J Interact Marketing* 2004;18(1):38-52. [doi: [10.1002/dir.10073](https://doi.org/10.1002/dir.10073)]
18. Biyani P, Caragea C, Mitra P, Yen J. Identifying emotional and informational support in online health communities. In: *Proceedings of the 25th International Conference on Computational Linguistics*. 2014 Presented at: 25th International Conference on Computational Linguistics; Aug 23-29, 2014; Dublin, Ireland p. 827-836.
19. Puhl R, Brownell K. Bias, discrimination, and obesity. *Obes Res* 2001;9(12):788-805. [Medline: [11743063](https://pubmed.ncbi.nlm.nih.gov/11743063/)]
20. PRAW. PRAW: the Python Reddit Api wrapper. URL:<https://praw.readthedocs.org/en/v3.0.0/> [accessed 2016-03-24] [WebCite Cache ID 6gFtNMQeM]
21. reddit. LoseIt. URL:<https://www.reddit.com/r/loseit> [accessed 2016-03-24] [WebCite Cache ID 6gFtn4xcr]
22. reddit. 22 guidelines for posting in the Loseit reddit community. 2015. URL:<https://www.reddit.com/r/loseit/wiki/faq> [accessed 2016-03-24] [WebCite Cache ID 6gFtvRlwX]
23. Python. langdetect. 2015. URL:<https://pypi.python.org/pypi/langdetect/1.0.1> [accessed 2016-03-24] [WebCite Cache ID 6gFuExB9P]
24. Maclean PS, Bergouignan A, Cornier M, Jackman MR. Biology's response to dieting: the impetus for weight regain. *Am J Physiol Regul Integr Comp Physiol* 2011 Sep;301(3):R581-R600 [FREE Full text] [doi: [10.1152/ajpregu.00755.2010](https://doi.org/10.1152/ajpregu.00755.2010)] [Medline: [21677272](https://pubmed.ncbi.nlm.nih.gov/21677272/)]
25. [No authors listed]. Obesity: preventing and managing the global epidemic. Report of a WHO consultation. *World Health Organ Tech Rep Ser* 2000;894:i-xii, 1. [Medline: [11234459](https://pubmed.ncbi.nlm.nih.gov/11234459/)]

26. Manning CD, Raghavan P, Schütze H. Introduction to Information Retrieval. New York: Cambridge University Press; 2008.
27. Blei D, Ng A, Jordan M. Latent Dirichlet allocation. *J Mach Learn Res* 2003 Jan;3:993-1022 [[FREE Full text](#)]
28. James G, Witten D, Hastie T, Tibshirani R. An Introduction to Statistical Learning With Applications in R. London, UK: Springer; 2013.
29. Garawi F, Devries K, Thorogood N, Uauy R. Global differences between women and men in the prevalence of obesity: is there an association with gender inequality? *Eur J Clin Nutr* 2014 Oct;68(10):1101-1106. [doi: [10.1038/ejcn.2014.86](#)] [Medline: [24918120](#)]
30. Yaemsiri S, Slining MM, Agarwal SK. Perceived weight status, overweight diagnosis, and weight control among US adults: the NHANES 2003-2008 Study. *Int J Obes (Lond)* 2011 Aug;35(8):1063-1070. [doi: [10.1038/ijo.2010.229](#)] [Medline: [21042327](#)]
31. Tsai SA, Lv N, Xiao L, Ma J. Gender differences in weight-related attitudes and behaviors among overweight and obese adults in the United States. *Am J Mens Health* 2015 Jan 15:pil. [doi: [10.1177/1557988314567223](#)] [Medline: [25595019](#)]
32. Bidmon S, Terlutter R. Gender differences in searching for health information on the internet and the virtual patient-physician relationship in Germany: exploratory results on how men and women differ and why. *J Med Internet Res* 2015;17(6):e156 [[FREE Full text](#)] [doi: [10.2196/jmir.4127](#)] [Medline: [26099325](#)]
33. Forster JL, Jeffery RW. Gender differences related to weight history, eating patterns, efficacy expectations, self-esteem, and weight loss among participants in a weight reduction program. *Addict Behav* 1986;11(2):141-147. [Medline: [3739799](#)]
34. Wyatt HR, Grunwald GK, Seagle HM, Klem ML, McGuire MT, Wing RR, et al. Resting energy expenditure in reduced-obese subjects in the National Weight Control Registry. *Am J Clin Nutr* 1999 Jun;69(6):1189-1193 [[FREE Full text](#)] [Medline: [10357738](#)]
35. Blomain ES, Dirhan DA, Valentino MA, Kim GW, Waldman SA. Mechanisms of weight regain following weight loss. *ISRN Obesity* 2013;2013:1-7 [[FREE Full text](#)]
36. Raynor HA, Jeffery RW, Ruggiero AM, Clark JM, Delahanty LM, Look AHEAD (Action for Health in Diabetes) Research Group. Weight loss strategies associated with BMI in overweight adults with type 2 diabetes at entry into the Look AHEAD (Action for Health in Diabetes) trial. *Diabetes Care* 2008 Jul;31(7):1299-1304 [[FREE Full text](#)] [doi: [10.2337/dc07-2295](#)] [Medline: [18375417](#)]
37. Jerome GJ, Dalcin A, Coughlin JW, Fitzpatrick S, Wang N, Durkin N, et al. Longitudinal accuracy of web-based self-reported weights: results from the Hopkins POWER Trial. *J Med Internet Res* 2014;16(7):e173 [[FREE Full text](#)] [doi: [10.2196/jmir.3332](#)] [Medline: [25042773](#)]
38. Elfhag K, Rössner S. Who succeeds in maintaining weight loss? A conceptual review of factors associated with weight loss maintenance and weight regain. *Obes Rev* 2005 Feb;6(1):67-85. [doi: [10.1111/j.1467-789X.2005.00170.x](#)] [Medline: [15655039](#)]
39. Pollak KI, Østbye T, Alexander SC, Gradison M, Bastian LA, Brouwer RJ, et al. Empathy goes a long way in weight loss discussions. *J Fam Pract* 2007 Dec;56(12):1031-1036. [Medline: [18053443](#)]
40. Nambisan P. Information seeking and social support in online health communities: impact on patients' perceived empathy. *J Am Med Inform Assoc* 2011 May 1;18(3):298-304 [[FREE Full text](#)] [doi: [10.1136/amiajnl-2010-000058](#)] [Medline: [21486888](#)]
41. Rao G, Burke LE, Spring BJ, Ewing LJ, Turk M, Lichtenstein AH, American Heart Association Obesity Committee of the Council on Nutrition, Physical Activity/Metabolism, Council on Clinical Cardiology, Council on Cardiovascular Nursing, Council on the Kidney in Cardiovascular Disease, Stroke Council. New and emerging weight management strategies for busy ambulatory settings: a scientific statement from the American Heart Association endorsed by the Society of Behavioral Medicine. *Circulation* 2011 Sep 6;124(10):1182-1203 [[FREE Full text](#)] [doi: [10.1161/CIR.0b013e31822b9543](#)] [Medline: [21824925](#)]
42. Verheijden MW, Bakx JC, van Weel C, Koelen MA, van Staveren WA. Role of social support in lifestyle-focused weight management interventions. *Eur J Clin Nutr* 2005 Aug;59 Suppl 1:S179-S186. [doi: [10.1038/sj.ejcn.1602194](#)] [Medline: [16052189](#)]
43. Flores-Gómez I, Bacardí-Gascón M, Armendáriz-Anguiano AL, Pérez-Morales ME, Jiménez-Cruz A. [Evidence of social support as therapy for weight loss: a systematic review]. *Nutr Hosp* 2012;27(5):1422-1428 [[FREE Full text](#)] [doi: [10.3305/nh.2012.27.5.5877](#)] [Medline: [23478687](#)]

## Abbreviations

- API:** application programming interface
- BMI:** body mass index
- LDA:** latent Dirichlet allocation
- WHO:** World Health Organization

*Edited by G Eysenbach; submitted 29.03.16; peer-reviewed by S Bidmon, S Tsai; comments to author 25.07.16; revised version received 05.09.16; accepted 22.10.16; published 16.01.17*

*Please cite as:*

*Pappa GL, Cunha TO, Bicalho PV, Ribeiro A, Couto Silva AP, Meira Jr W, Beleigoli AMR*

*Factors Associated With Weight Change in Online Weight Management Communities: A Case Study in the LoseIt Reddit Community*  
*J Med Internet Res 2017;19(1):e17*

*URL: <http://www.jmir.org/2017/1/e17/>*

*doi: [10.2196/jmir.5816](https://doi.org/10.2196/jmir.5816)*

*PMID: [28093378](https://pubmed.ncbi.nlm.nih.gov/28093378/)*

©Gisele Lobo Pappa, Tiago Oliveira Cunha, Paulo Viana Bicalho, Antonio Ribeiro, Ana Paula Couto Silva, Wagner Meira Jr, Alline Maria Rezende Beleigoli. Originally published in the Journal of Medical Internet Research (<http://www.jmir.org>), 16.01.2017. This is an open-access article distributed under the terms of the Creative Commons Attribution License (<http://creativecommons.org/licenses/by/2.0/>), which permits unrestricted use, distribution, and reproduction in any medium, provided the original work, first published in the Journal of Medical Internet Research, is properly cited. The complete bibliographic information, a link to the original publication on <http://www.jmir.org/>, as well as this copyright and license information must be included.