An Analysis of Individuals' Behavior Change in Online Groups

David Jurgens[1], James McCorriston[2], and Derek Ruths[2]

[1] Stanford University, Stanford, CA 94095, USA,
jurgens@stanford.edu
[2] McGill University, Montreal, QC, H3A 2A7, Canada,
mccorriston@mail.mcgill.ca, druths@networkdynamics.org

Abstract. Many online platforms support social functions that enable their members to communicate, befriend, and join groups with one another. These social engagements are known to shape individuals' future behavior. However, most work has focused solely on how peers influence behavior and little is known what additional role online groups play in changing behavior. We investigate the capacity for group membership to lead users to change their behavior in three settings: (1) selecting physical activities, (2) responding to help requests, and (3) remaining active on the platform. To do this, we analyze nearly half a million users over five years from a popular fitness-focused social media platform whose unique affordances allow us to precisely control for the effects of social ties, user demographics, and communication. We find that after joining a group, users readily adopt the exercising behavior seen in the group, regardless of whether the group was exercise and non-exercise themed, and this change is not explained by the influence of pre-existing social ties. Further, we find that the group setting equalizes the social status of individuals such that lower status users still receive responses to requests. Finally, we find, surprisingly, that the number of groups one joins is negatively associated with user retention, when controlling for other behavioral and social factors.

1 Introduction

In both online and offline settings, groups and communities provide individuals with opportunities for tie formation and social learning [12] and also expose individuals to new information and peer influence [17]. However, individuals also change their behavior on the basis of their explicit social ties [3, 6], raising the question of what impact group membership specifically has on individuals. Here, we examine how users change their behavior as members of a group in three different settings when controlling for the effects of social ties.

Significant work has shown that individuals change their behavior through explicit peer influence and implicitly observing their peers' actions. For example, the effects of explicit ties within a social network have been shown to manifest in a number of social phenomena such as information diffusion [9, 16], peer influence [74], and exercise frequency [3, 6]. Similarly, other work has examined communities as a whole, showing their potential influence on behavioral aspects such as linguistic norms [24], visual presentation [84], and content moderation [51]. However, little is known about how online groups affect behavior independent of effects from the social network and the whole community, with only a few studies examining their lifecycle [44, 85], their effect on social network formation [8, 54], and how they facilitate political engagement [21, 83] and dietary choice [56].

In this this study, we examine how individuals change their behavior on the basis of group membership, employing a longitudinal dataset from Fitocracy over five years with 477K users and 12K groups. We measure individuals' exercise and communication activities to test for the behavioral effects of group memberships in three contexts: (1) **Physical Activity**: When joining a group, do users change their behavior by becoming more similar to the behaviors seen in the group? (2) **Communication**: How do groups impact a user's requests for help and do they tend to respond to such requests? (3) **User Retention**: Does joining a group make it more likely for a user to stay on the site? Due to the presence of both groups and a social network on Fitocracy, we can disentangle the interaction of these two sources of behavior influence.

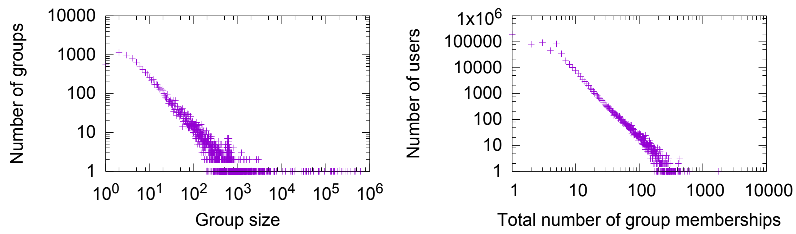
Our work offers three main insights into groups' effects on behavior. First, we demonstrate that joining a group correlates strongly with changes in individuals' behavior, with individuals adopting the exercise behaviors they observe, regardless of whether the group was focused on exercise. Second, individuals modulate their communication strategy for which audience they direct questions to, using groups to answer subjective advice-seeking questions, while asking fact-seeking questions to friends. Further, we find that the penalty for low social status is negated in groups, with individuals of all statuses being equally likely to have their questions answered—unlike when users ask either their social network or others directly in which case high status is critical for a response. Third, surprisingly, increased social and group engagements are not positively associated with new user retention; instead new users are more likely to stay if they engage in the core fitness-related functionality of Fitocracy.

2 Fitocracy Groups

Platform Fitocracy is a social networking platform designed for individuals interested in fitness. Users track activities on a daily basis by selecting from a predefined list of 1,090 exercises, which enables the ability to precisely measure any changes in the activities that were performed. Users, all of whose profiles are public, may optionally self-report their age, gender, and height, with 91% reporting at least two attributes.

Beyond recording workouts, Fitocracy includes common social functionalities. Users may post status updates and comment on others' workouts and statuses. Notably, Fitocracy supports directed communication by enabling users to post comments to another user's wall. All posts are public and may be replied to by any users. No private communication exists on Fitocracy which allows full observability of communication.

Fitocracy includes gamification elements where users receive points for recording workouts. Points determine a user's *level*, which serves as a indicator of their degree of fitness and intensity. Notably, the platform highlights the highest-leveled users and those scoring the most points over different durations. Thus, because of its visibility on the platform, ties to physical prowess, and generalizability across different exercising disciplines, level is a reasonable proxy for *social status*.



(a) Group Sizes (b) Group memberships per user

Fig. 1: Group membership and size rates.

During the lifetime of the platform, Fitocracy added the ability for individuals to find paid coaching and added a knowledge base section with information on diet and exercise. These elements, combined with gamification, can potentially drive an individual to change their behavior beyond what they would do independently.

Dataset User activities, profile, social data, and group memberships were crawled using a web scraper to collect the complete profiles and workout histories of 476,716 users who recorded at least one workout. The total dataset contains 12,522,959 workouts (55.1M activities) over nearly a five year span from February 2011 to December 2015. Users were roughly balanced between genders (48.7% female) and, while skewed slightly younger, contain a sizable older population with a mean age of 29.2 and standard deviation of 19.5 years.

Group Statistics Groups are a widely-used social feature on the platform and are commonly based on sports, diet, lifestyle, or location. Once a user joins a group, the activities recorded by that group's members appear in a separate feed on the user's home page, exposing them to how people in the group exercise. Furthermore, group members may make posts to the group's feed, such as to ask a question.

The membership lists were gathered for 12,669 groups. Group sizes follow a power law curve with $\alpha = 1.96$ (Figure 1a), which was determined to be a better fit than a log-normal distribution at p<0.01 [20, 1]; the largest groups have hundreds of thousands of members. Individuals typically join a handful of groups, with 94% of users belonging to ten or fewer groups (Figure 1b). Groups had low social connectivity with an average clustering coefficient of 0.152, indicating that most people in a group do not follow each other. Further, a group's size and clustering coefficient were not correlated (Pearson's r =-0.01; p=0.22), indicating that smaller groups are not necessarily more connected.

3 Exercise Behavior Change

Individuals exercise for a variety of fitness goals. On Fitocracy, groups can provide social and informational support to individuals for attaining their fitness goals, e.g., observing what types of exercises are done by successful runners and motivation for continued exercise. Here, we ask to what degree do users change their behavior after joining a group? Our study is motivated by recent work [3, 6] showing that the formation of new friendships in a user's social network increases their exercise frequency as measured by daily step count. We examine an analogous condition, asking to what degree do individuals change their exercising behavior when joining a new group and observing the exercises performed by its members.

Experimental Setup

Cohorts Our analysis focuses on individuals who regularly exercise with a minimum level of activity, defined as recording at least one activity twice per week. This selection process controls for the possible effects of changes in exercise frequency (e.g., stopping exercise for several weeks) and ensures accurate estimates of individuals' behavior at any point. To estimate their average behavior, groups are only included if they have at least 100 active members who in total record at least ten workouts a week. Ultimately before-and-after-joining workout histories were gathered for 109,772 individuals across 996 groups. To contrast with the change seen by the study cohort, a *control cohort* was constructed of all users with the same exercise frequency but whom did not join a group during the study period; it consists of 29,520 users.

Groups Groups may form for different purposes so all groups in this study were categorized as either exercise-focused (27.7%), such as Running or Powerlifting, or those focused on a non-exercise theme (72.3%) such as Vegans or Video Games. This distinction allows comparing groups where the expected impact of the group's knowledge and conversation is on an individual's exercising behavior versus another aspect of their life. We further divide these groups into one of eight themes: social (40.8%), shared interests (18.3%), general fitness (10.6%), challenge (8.8%), sport-specific (6.3%), city or regional affiliation (6.3%), dietary (4.6%), and weight-loss (2.3%). A small number of groups (1.2%) were unable to be categorized due to no clear description or title and are excluded from this study. Challenge groups ask members to perform specific exercises and show members' rankings according to the challenge goal.

Exercise Measurement A natural method of comparing individuals' exercising behavior is to simply compare the frequencies with which they perform each exercise. However, directly comparing exercises can fail to recognize thematic similarities in behavior, e.g., that "trail running" and "jogging" are highly related, and that individuals performing such exercises have the same behavior in practice. Therefore, we adopt the approach of Jurgens *et al.* [43] for capturing high-level fitness behaviors. Here, behaviors are probability distributions over exercises and are learned by training a Latent Dirichlet Allocation (LDA) model [13] on individuals' workout histories. Much like how an LDA model identifies topics of related words in text documents, our model identifies behaviors of thematically-related exercises. The authors show that the LDA model is robust to number of topics and captures salient behaviors for a wide-range of choices. Here, we opt for 100 behaviors to capture fine-grained changes.

The LDA model is trained from documents where each represents all the exercises performed by a user; to provide more behavioral consistency when learning the model, we construct the training data from documents consisting of the exercises recorded by a user within a single month, where that user has recorded at least two workouts a week. The final training contains a total of 656,802 documents.

Measuring Behavioral Similarity The LDA model infers a probability distribution over exercise-based behaviors (topics) from a user's activities. To compare behaviors,

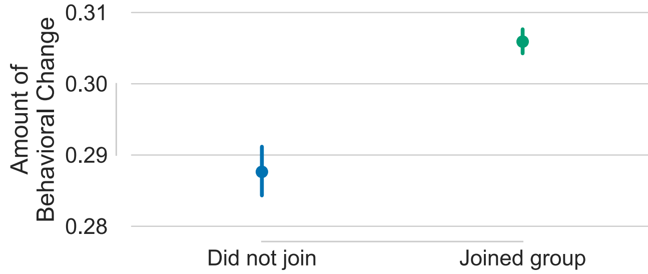


Fig. 2: The cohort of users who joined a group (shown left in green) had a statistically significant increase in the amount of change in the following month, compared with the control cohort of users who did not join a group but had equivalent levels of activity (shown right in blue). Bars show 95% confidence intervals computed through bootstrap resampling and points denote the mean change.

we calculate the distance between two distributions by computing Jensen-Shannon Divergence (JSD). Behavior similarity is then measured as $1-$ JSD, where 1 indicates maximal similarity. We note that when comparing the behaviors of a group and one of its members, the distribution of the group is inferred from the activities of all other members, *excluding* those of the individual under analysis.

Did users change behavior after joining a group?

To assess whether joining a group was associated with behavior change, we compared the behaviors for individuals in the study cohort for their activities in the month prior to joining and the month after. By contrasting these differences with the changes seen in an analogous period for the control cohort, shown in Figure 2, we observe that users who join a group change their behavior significantly more than those in the control cohort. This initial result should not be directly interpreted as the group causing the change itself. An alternative interpretation is that joining a group signals an intent to change behavior; therefore any observed difference in the groups could be due to selection bias, rather than the effects of group memberships. Nonetheless, there is a clear signal that when a user does join a group, their behavior is likely to change more.

Do users adopt the group's behavior?

When an individual joins the group, do they change by adopting the behaviors of the group? We test for group-specific change by measuring behavioral similarity before and after joining with respect to the group's behavior during the relevant time periods. Here, we consider two cohorts: (1) users who join an *Exercise-focused group*, e.g., Running, (2) users who join a *Non-exercise focused group*, e.g., Vegetarians. Behavior change is compared against a *Null model* of the same individuals but where change is measured as if they had joined a *random group* at that same point in time. This null model captures the scenario where the user has the same motivation for change (as signaled by joining a group) but is not exposed to the exercises of that group. For the null model, we sample 30 random groups per user and require that a random group have the same activity level.

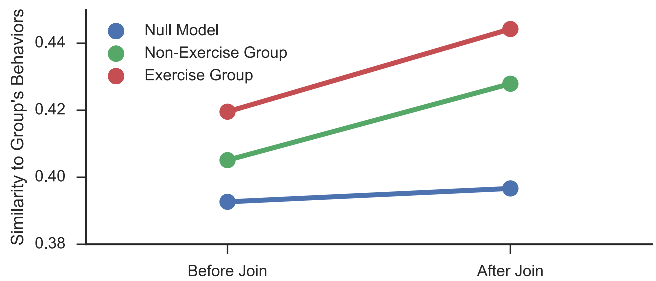


Fig. 3: Users had a substantially larger change towards the behavior of the group they joined than would be expected than by chance, as shown with the null model. Bars show bootstrapped 95% confidence intervals but are too small to be seen and points denote mean similarity.

The exercise, non-exercise, and null conditions allow testing for different hypotheses of exercise behavior change:

H1: The platform as a whole is converging towards a common behavior so all users become more similar to each other every month. If true, we should expect an increase in user similarity with the random group's behavior in the null model.
H2: An individual joins a group due to interest in the group's theme and therefore are likely to adopt behaviors with respect to that theme. If true, we should expect only users who joined exercise-related groups to become more similar to the group's exercise behavior.
H3: Individuals adopt the behavior they are exposured to within a group. If true, we should expect to see increased similarity for both exercise and non-exercise groups.

Results At the outset, it is important to note that most studies on community-driven behavior change (including the present study) are observational, which limits the extent to which causality can be inferred unless experimental conditions allow for a natural experiment. That said, the findings of prior work and the findings here are strongly suggestive of some kind of causal mechanism. Establishing such causality is an important direction for future work. Here, we consider the alternative ways in which our findings could inform such investigations.

Individuals adopted the exercise behavior they observed in groups, *independent of the exercise-focus of the group*, as seen in Figure 3. Individuals in both exercise-related and non-exercise groups became substantially more similar to the group's behavior, far beyond what is expected by chance, as reflected in the null model. The increase in behavioral similarity was consistent across group types, shown Figure 4 in Appendix A, However, the degree of adoption varied by type of group, with users joining Challenge groups changing the most and those joining location, sports, or shared-interests groups changing the least. Individuals are known to adopt the norms and behaviors of communities upon joining [47] and this tendency has been seen in topical online communities with changes to linguistic behaviors [76, 24, 59, 55, 14]. Hence, we interpret the consistent shift in user exercise behavior towards the group's behavior as evidence that individuals adopt the offline norms of online groups. Our result also concurs with the observations of prior work's surveys of Fitocracy that showed perceived social norms and feedback from "likes" provided strong incentives for sustained behavior [35]. Examining the null model, we found weak support for H1, where there was a tiny but statistically-significant increase in user similarity month-over-month for the platform as a whole ($p<0.01$ using the two-sample Kolmogorov-Smirnov test); given the exercise focus the platform, we speculate that this effect is due individuals' continued lifestyle improvements. The consistent trend of users in all group types to adopt their group's exercises supports H3, with a substantially larger effect size than H1 ($p<0.01$ using the two-sample Kolmogorov-Smirnov test), and did not support evidence of change only for exercise-focused groups only (H2). Thus, our results suggest that an individual observing a group could drive behavior change, though our experimental conditions cannot fully establish causality.

Is change driven by explicit social ties?

Given that prior studies have shown that an individual's behavior is influenced by their friends [5, 82], one possible explanation for the behavior change is the presence of friends within the group. That is, when a user joins a group that contains their friends, they change to match the behavior of their friends within the group, rather than being influenced by the group itself. However, an individual's demographics and prior behavior are also known to affect the degree to which they change their exercising behavior [22, 78, 58, 69, 77]. Thus, we test two hypotheses for what factors explain the degree of change seen by an individual: (1) **H4**: behavior change is driven by the influence of existing relationships in the group and thus the percentage of the group that is the user's friends predicts the magnitude of the behavior change and (2) **H5**: behavior change is driven by demographic factors of the individual.

To test the impact of friendship, we construct a mixed-effect linear regression model to predict the degree of change towards the group's mean behavior. We include random effects for both the individual and group to control for within-subject and within-group variation. As fixed effects, we include an individual's demographic features, their frequency of posting to the group, group size, and group type. Following common practice when regressing on variables from different scales, all numeric variables are centered and standardized to enable comparison in effect sizes per standard deviation of change.

Results The regression results in Table 1 show that an individual's demographics and prior behavior have a much larger impact on the degree of their change than prior social ties. An individual's prior similarity to the group's behavior is the largest predictor of increased conformity, indicating that people change the most when joining groups whose activities are more familiar. Further, users who have exercised more frequently or more intensely (i.e., have a high level) change more. The coefficients for age and gender concur with prior work that had shown that (1) older individuals perceive more limitations in their exercise abilities [22, 78] and thus, may be less likely to attempt new behaviors and (2) inconclusive results of gender on exercising behavior when controlling for age [10, 68], suggesting its effect should not be significant. These results confirm H5 that demographics are in part responsible for the degree of change.

Examining group variables, the percentage of friends in the group has a statistically-significant positive association with behavior change towards the groups, which confirms H6. However, its effect size is the second smallest of all significant variables;

Variable	Coefficient	Variable	Coefficient
User's initial similarity to the group	0.062^{***}	Type: dietary	-0.012^{***}
User's level	0.015^{***}	Type: fitness	0.003
User's age	-0.007^{***}	Type: interests	-0.016^{***}
Gender: male	-0.002	Type: location	-0.027^{***}
Gender: unspecified	0.065^{***}	Type: social	-0.010^{***}
% of group friended	0.002^{***}	Type: sport	-0.001
# of posts to group wall	0.001^{***}	Type: weight-loss	-0.003
Group's mean level	0.004^{***}	Group size	-0.002^{***}
Intercept	0.029^{***}		
N 109,772	Marginal R^2 0.404	Conditional R^2 0.676	

Table 1: Regression coefficients for predicting the degree of behavior change towards the group's behavior show demographic and behavior factors were stronger predictors of change than the percentage of friends in the group. This suggests that peer influence from friends is not the driving force behind behavior change seen when a user joins a group. *** denotes $p<0.01$. Due to normalization, coefficients should be interpreted as change per standard deviation of difference in the variable. For categorical coding, the reference gender is female and group type is challenge.

i.e., having a standard deviation more friends in a group will have a smaller behavior change than having a similar magnitude of change in another variable. As such, *we do not interpret the change seen when joining a group to be primarily explained by explicit friendships within the group*.

More generally, our findings offer potential insight for how to introduce behavioral interventions where individuals are placed into online groups to support behavior change [11, 66, 57], e.g., dietary or exercise. Given the choice between placing the individual in a group with more friends versus a group with a more familiar behavior, our results suggest that the individual would change more if placed in the latter group.

4 Requesting Help from Peers

Individuals frequently ask questions to their peers in online platforms [52, 73]. This behavior, known as social search, leverages the social capital of an individual for mobilizing their peers to respond [62, 67, 80]. On Fitocracy, individuals have the three communication forms (posting to your own wall, another's wall, or to a group), with each reaching distinct audiences and incurring different social capital costs. We examine how individuals modulate their communication behavior on the basis of being members of groups in two settings: (1) how audience affects the types of questions asked, (2) what social factors predict whether a question will be answered.

Who is asked what?

Groups provide an avenue for individuals to pose questions to a large audience and, due to their membership, potentially increase the diversity in response beyond that available

	Own wall		*Other's Wall*		*Group*	
Type	Asked	w. reply	Asked	w. reply	Asked	w. reply
Advice	35.0	25.7	6.5	76.9	51.5	90.2
Fact Seeking	29.5	35.6	7.0	71.4	20.5	92.6
Invitation	1.5	33.3	2.5	100.0	3.5	57.1
Participation	1.5	33.3	0.0	-	6.0	75.0
Favor	2.5	20.0	0.0	-	0.0	-
Rhetorical	18.5	37.8	5.5	63.6	4.5	88.8
Personal	5.5	36.3	53.5	76.6	11.0	95.4
Social	6.0	25.0	25.0	70.0	3.0	66.7

Table 2: Individuals modulate the questions they ask by audience as seen in the stark differences in question distributions. No Offer type questions were seen.

through existing ties. We hypothesize that the low social capital cost associated with asking the social ties implicit in a group should cause users to modulate their behavior in terms of which questions are asked to groups versus friends (directly). To underscore, our question here is not about the *topic* a user might ask about. Rather, the hypothesis concerns what kind of questions a user might pose on a given topic and how those questions would be posed: the null hypothesis being that the user does not differentiate between the user-to-user and group contexts.

Methodology A question-response dataset was created by automatically extracting posts with at least one sentence ending with a '?' and then using a series of heuristics to remove noise from platform-generated messages and messages not expressing a question, e.g., "huh?" In total, users posted 44,543 questions on their own wall, 87,205 questions to the wall of another users, and 61,969 questions to groups.

Questions were classified using a scheme adapted from Ellison *et al.* [27] and Morris *et al.* [62] to match the question types seen in Fitocracy (described in full in Appendix B, Table 5). Question types capture the broad intentions of users during social search, e.g., asking for advice. Two hundred questions from each setting were then randomly sampled and annotated by two annotators, who had an agreement of 0.657, measured using Krippendorff's α, which indicates substantial agreement [7]. After annotation was complete, both annotators adjudicated all items for the final labels.

Results Given the three potential audiences to which a question could be posed, individuals displayed clear preferences for the types of questions asked to each. Table 2 displays the breakdown of question types seen for each audience and the percentage of the questions that received a response. Two main trends are seen in these results.

First, individuals used groups and their social network primarily for seeking information rather than for social purposes. However, the type of information asked for differed: subjective questions for advice were posed more to groups, while factual questions were asked more frequently to followers; both differences are significant at $p<0.01$. This result contrasts with the analysis of Morris *et al.* [62], who found no statistically-significant difference between factual and advice question frequencies on Facebook and Twitter, which are analogous to own-wall and group posting conditions, respectively. We interpret this difference as suggesting that, given equal opportunity between asking a peer or group audience, individuals may choose not to mobilize their social capital when asking a question without an objective answer—thereby incurring the social costs associated with asking peers [62]—and instead seek out answers to subjective questions from individuals with whom they have no social relationship.

Despite the features on the Fitocracy site designed to highlight and promote highly proficient individuals, the questions directly posed to other individuals on their wall were primarily social and personal in nature, rather than requests for advice from experts. This result is made more striking in comparison to the clear information gathering behavior seen for questions posted to a user's wall and groups. Surprisingly, despite the social nature of the questions, the majority of questions were asked to those without social relationships. Indeed, 85.7% questions were replied to by strangers while only 14.3% were replied to by friends, suggesting a general openness beyond that predicted by social ties.

Second, groups provide a significantly higher response rate for questions that a user posted to their own wall and are in line with those seen for community question-answering sites [36, 38]. In contrast, the response rate when an individual asks their social network via a wall post is similar to that seen on Twitter when using general-purpose hashtags [40] rather than those when asking on social networking sites [61]. One explanation could be that friends tend to answer questions posted in groups, which was seen for Twitter Q&A hashtags [73], where most responses were from individuals with established social connections. However, we find that few questions posted to groups had responses by friends (13.9%). We examine the social factors in these responses next.

Who receives a response?

As potential members of each audience, individuals have the option of responding to questions. Social theory suggests that individuals with high social status are more likely to have their questions answered, as lower-status individual aim to acquire social capital through fulfilling these requests [70, 33, 42]. Alternatively, group membership may elicit altruistic behavior due to the perception of a shared social affiliation [12, 30]. As a result, group members may respond without incentivization regardless of an individual's status. Finally, linguistic signals of politeness convey deference and respect and can incentivize individuals to respond [75, 18, 4, 72, 37]. These behavioral theories can be operationalized as three hypotheses for whether a question receives a response.

H6: High social status individuals are more likely to receive a response.
H7: Group members are more likely to receive a response due to affiliation benefits.
H8: Asking politely increases the likelihood of a response.

These hypotheses offer alternate explanations of how individuals behave when choosing to answer questions, raising the question of which hypotheses are valid for each Fitocracy audience, e.g., do group members still require high social status to receive response or is the group affiliation alone sufficient? Following, we test each hypothesis for the three audiences to understand the impact of group membership.

Methodology We construct separate mixed-effect logistic regression models for each audience on the binary variable of whether a question receives a response. The models

	Own Wall	Other's Wall	Group
Gender: Male	-0.510^{***}	-0.110^{***}	-0.196^{***}
Age	0.045^{*}	-0.124^{***}	-0.129^{***}
Level	0.256^{***}	0.261^{***}	-0.003
# of Followers	0.647^{***}	-0.008	-0.020
# of Friends	0.027	0.005	-0.026
# of Previously asked questions to this audience	0.020	-0.100^{***}	-0.022
Message length	0.099^{***}	0.204^{***}	0.106^{***}
Message politeness	0.013	0.046^{***}	0.060^{***}
Intercept	-1.656^{***}	-1.049^{***}	-0.508^{***}
N	28,712	55,987	31,908
Marginal R^2	0.120	0.077	0.078
Conditional R^2	0.263	0.252	0.280
Note:			*p<0.1; **p<0.05; ***p<0.01

Table 3: Regressions on whether a question will receive an answer. Results for additional question content variables are reported in Appendix D in Tables 6, 7, and 8. All numeric variables are z-scored. For categorical coding, the reference gender is Female.

use fixed effects for sociological and textual variables to predict whether the question has a response, described in full in Appendix C. We use random effects for the location where the question is asked and the individual asking the question. As in previous models, all numeric variables are normalized such that coefficients reflect impact per standard deviation of change. To control for collinearity between variables, models were constructed using step-wise variable deletion, also described in Appendix C. We fit the model to using 116,607 questions for which we have users' demographic information.

Results Audiences differed widely according to which social factors were predictive of them replying to a question (Table 3), with our three hypotheses varying by audience. Individuals posting to groups did not require high social status (i.e., level) to receive a response, confirming H7 that group membership affords social benefits and provides no evidence for the role of status (H6); in contrast, high status plays a substantial role when a user asks on their own wall or the wall of another, suggesting H6 holds in these settings. Surprisingly, politeness (H8) was only significantly associated with increased response rates for questions posted to groups or to other's walls, though its effect size was relatively small. Demographic factors played a substantial role in response rates, with men and older users being much less likely to receive a response. However, social connectedness did not play a significant role except for the number of followers when an individual posts to their own wall, which is expected since it reflects how may people might see the question. More broadly, these findings suggest that groups can a play critical role for new users. Because a new user is unlikely to have high social status or a large number of followers, they are much less likely to have their questions answered *unless* they ask in a group setting where these factors do not matter.

5 User Retention

Online communities often change over time, with new users joining a site and older users departing at certain intervals. The departure of users can come from a variety of reasons, such as a lack of active connections to other users [26, 29] or changes in culture [24]. While relationships may fill the social needs of individuals, groups can potentially provide another form of social support, creating a sense of virtual community that keeps users engaged on the platform [65].

In examining group's impact on user retention, prior work has largely examined groups on social platforms where social engagement and communication are the primary uses of the platform [81, 50, 45, 46]. In contrast, as an exercise-oriented platform, Fitocracy provides workout tracking functionality as its core service, with social and group functionalities as additions. This distinction enables us to study the impact of groups on retention independent of the purpose of the platform. Therefore, we test how the likelihood of retention of new users is affected by (a) joining a group and communicating within it and (b) engaging in social networking and directed communication.

Methodology The Fitocracy platform publicly displays the dates when users record each workout and when users join groups. However, no information is provided about when users begin following another user. Therefore, to precisely track the impact of social network, we created a complete longitudinal dataset for 10,000 new users, by crawling these users profiles every day over a four month period to capture how their social network evolved, in addition to their workout and group activities. We construct our set of users from the "WTF – Welcome to Fitocracy" group, which all new users automatically join upon signing up for the site. Once crawled, we restrict our analyses to those that report their age, leaving a total of 7781 users.

To analyze user retention, we construct logistic regression models with the dependent variable of whether a user will record any kind of activity (post, comment, or workout) in the following month. Two models were designed with different blocks of variables to highlight the separate contributions of (a) user demographics and (b) social and group interactions. Due to access limitations, we were unable to continuously track changes to who follows each user. As in previous models, all numeric variables were standardized so that the variable coefficients can be compared. To reduce collinearity, we used step-wise variable deletion on the fixed effects to remove the variable with the largest variance inflation factor (VIF) until all variables had VIF ≤ 5; ultimately eight variables were removed. Each model is provided with the first month of the user's activity on the site.

Results Most of the variance in user retention is explained by individuals' demographic information and whether they actively engage in the site's core functionality of recording activities. Table 4 shows the results of the two regression models, with only small improvements to the model when adding social and group variables. The positive relations from the combination of the level and number of recorded workouts variables suggests that fitness enthusiasts are more likely to stay on the site, as these are the cohort of users who workout frequently or who record intensive exercises.

Surprisingly, for the most part, increases in social connectivity, interactions, and group memberships were not associated with increased retention. Indeed, of the social variables, only users posting to their own wall was positively associated and significant

	Null Model	Social Model
Intercept	−4.02***	−3.93***
# of workouts	0.39***	0.33***
Age	−0.33***	−0.23**
Gender: male	0.28*	0.11
Gender: unstated	1.99***	1.31***
Level	1.45***	1.55***
# of friends		−0.02
# of groups joined		−0.30***
# of posts received		−0.44
# of group comments		0.26
# of posts to others		0.65*
# of wall posts		0.16**
# of questions replied to		−0.12
Pseudo R^2	0.47	0.48
N	7773	7773

Table 4: Regression coefficients for user retention show that workout-associated variables were positive predictive of user retention, rather and social variables. *, **, and *** denote p<, 0.10, 0.05 and 0.01, respectively. All numeric variables are z-scored. For categorical coding, the reference gender is Female.

at $p<0.05$; however, these posts often describe fitness-related content, rather than being social. The number of group memberships was strongly negatively associated with user retention. We speculate that this behavior is analogous to that seen in social platforms such as Facebook when individuals are overexposed to other individuals and become dissatisfied [63, 32, 19, 28]. Here, when new users rapidly join a large number of groups immediately upon joining, they increase their exposure to individuals with much higher fitness status; this exposure in turn leads to lower satisfaction with their own activities and eventual dropout. Further, we can hardly expect community membership and user retention to be a simple function. For these reasons, future work is needed to test this hypothesis.

We interpret the negative association of social and group variables with increased user retention as reflecting the purpose of the Fitocracy website itself. While the site provides many social features, its core functionality is for tracking workouts. Users who actively seek such functionality are more likely to stay, whereas making friends and joining groups are not central to the focus of the platform and therefore do not keep them active. More broadly, this insight suggests that the success of an online platform with social features is dependent on the site's ability to engage users with its core functionality and is less dependent on the social networking functionality available.

Our finding that social and group features do not strongly predict user retention stands in contrast to the Fitocracy user surveys [34, 35] that found social factors were strongly associated with the user's perceived enjoyment of the platform and planned continued use. However, they surveyed individuals who had been actively using the site, rather than new users, which we examine. This difference suggests that social connections enrich the experience of established users, keeping them engaged longer, but that when initially joining, in the absence of social capital and strong ties, the core functionality drives longevity rather than the formation of social ties.

More broadly, our finding also relates to studies showing that social connections lead to increased user engagement [8, 53, 15] and retention [41, 71, 25]. While our results initially seem conflicting, these studies were done on social platforms where interpersonal communication is central to the platform's service, in contrast to the functionality of Fitocracy. Thus, our findings actually confirm those of earlier studies: because the sites are designed for social engagement, increased use of its core functionality should be strongly associated with user retention.

Furthermore, we observed Simpson's paradox in our regression: when using only the number of groups joined as the predictive variable, its coefficient was positive (0.151) and significant at $p<0.01$; however, when the full list of variables are included, its coefficient flips its sign, indicating a confounder variable was missing from the simplified regression and the overall impact of group joins is detrimental to retention.

In light of other work, our results point to a deeper mechanism behind retention: the ability of a site to attract and engage individuals in its core functionality predicts the size of its user base. More broadly, our results suggest that for online platforms, the addition of social and group functionality is not immediately necessary nor beneficial; such features only benefit the retention of established users, rather than new users.

6 Conclusion

Groups in online social networks can have a profound impact on the behaviors of individuals. Through a large-scale analysis of hundreds of thousands of users on Fitocracy, a social media platform dedicated to fitness and workout tracking, we demonstrate the impact of groups in three core contributions. First, the act of joining a group is strongly associated with an individual changing their behavior to be more similar to the group's; further, this behavior change is not explained through social influence and is substantially larger than expected by chance. Second, we demonstrate that individuals modulate their communication strategy by preferring to seek subjective answers and advice from group members, while asking factual questions more to friends. This difference suggests a strategic choice in the willingness to mobilize social capital when a objective answer may not be available. Additionally, we find that when individuals had the option of answering a question, they are more likely to respond to those asked in a group, independent of the relative social status of the asker, which suggests online groups promote in-group altruistic behavior. Third, we find that neither group nor social activities strongly contribute to the retention of new users on the platform, but rather retention is explained most by the individual's engagement with the platform's core functionality, i.e., workout tracking. Viewed with prior studies on user churn in social networking sites which found retention was increased by social relationships, our work points to the more fundamental mechanism being user engagement in the platform itself (e.g., tracking for Fitocracy, being social for social platform) rather than participation in social functionalities. Beyond Fitocracy, our results demonstrate that social groups can serve as a primary form of information for individuals aiming to change their behavior, which has broader implications when designing policies and campaigns to raise awareness about health topics [64], which to-date have seen only modest benefits from incorporating social media [49, 57].

Acknowledgments

We thank the reviewers for their insightful and detailed comments, with special thanks to Reviewer 3 for their nuanced analysis. The first author would also like to thank Tim Althoff, Will Hamilton, and Jure Leskovec for their helpful discussion and feedback. Finally, we thank the Fitocracy developers for creating their platform and making it accessible to all.

A Appendix: Additional Behavior Change Results

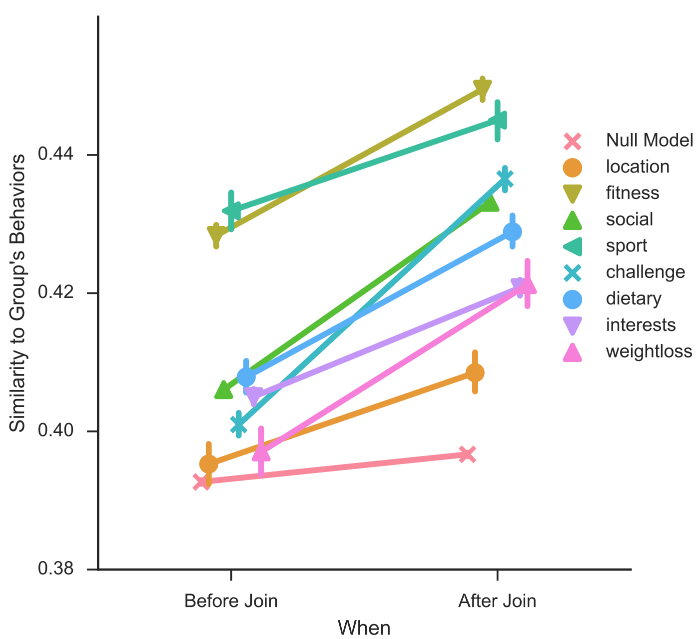


Fig. 4: Users change their behavior to be more similar to the group's behavior in the month after joining, regardless of group type. Symbols have added x-axis jitter for improved visual clarity. Bars show bootstrapped 95% confidence intervals. The large effect seen in the challenge groups suggests that public competition in online groups can facilitate individuals changing their behavior to meet their goals. Indeed, surveys, Fitocracy users rated the gamification aspects of the site highly for their enjoyment and motivation [31, 34].

B Appendix: Question Classification

Class	Description
Advice	Asking for a recommendation or opinion ("What is the best leg exercise?")
Fact Seeking	Asking a concrete question ("What's the world record marathon time?")
Invitation	Inviting others to a social event ("Anyone want to train together in Boston?")
Participation	Asking about attending or participating in an event or social activity ("Is anyone doing this road race this weekend?")
Offer	Offering services or help ("Does anyone want a used treadmill?")
Favor	Requesting a favor from another user ("Mind checking my form in this video?")
Rhetorical	Asking with no intention for a response ("am I right or what?")
Personal	A factual question posed to a user or group about their habits, lifestyle, or some other aspect of their person ("what's your diet like?").
Social	An open-ended, generic question to group or person, often with a subjective nature ("hows everyone doing today?")

Table 5: The classification of user question on social search and applied in Section 4, adapted from Ellison et al. [27] and Morris et al. [62]

C Appendix: Details on Question Response Regression

Social variables include the user's gender, age, level, number of followers, number of friends, and the number of times the user has asked this audience a question previously. Text features are partially drawn from the setup for Althoff *et al.* [2], which examined requests for favors on Reddit. We measure the overall politeness of a request with the model of Danescu-Niculescu-Mizil *et al.* [23]. To capture broad content variations, we (1) train a 20-topic LDA model [13] to generate a distribution of topics for each question, described in Appendix B, Table 9 and (2) count the relative frequency of word categories from the Linguistic Inquiry and Word Count (LIWC) [79] lexicon. Question sentiment is measured using word frequencies from the NRC sentiment and emotion lexicons [48, 60]. Finally, we include the relative frequency of hedges and modals in the question [39].

The mixed-effect logistic regression was constructed using step-wise variable deletion, which removed the variable with the highest variance inflation factor (VIF) until all variables had a VIF < 5. This process removed 9 variables all of which were LIWC lexical categories. The final model had 94 variables.

D Appendix: Additional Question Response Regression Results

	Own Wall	Other's Wall	Group
hedges	0.001	-0.028^{**}	0.016
modals	-0.019	-0.078^{***}	-0.103^{***}
NRC_trust	-0.007	0.013	0.086^{***}
NRC_fear	0.008	0.037^{**}	-0.060^{**}
NRC_negative	0.020	-0.020	0.039
NRC_sadness	-0.024	-0.032^{*}	0.045^{*}
NRC_anger	0.034	0.022	-0.020
NRC_surprise	-0.004	-0.043^{***}	-0.060^{**}
NRC_positive	0.053^{*}	-0.019	0.036
NRC_disgust	-0.029	-0.021	-0.006
NRC_joy	-0.025	0.008	-0.088^{***}
NRC_anticipation	-0.019	0.012	0.019
LIWC_Inhib	0.024	-0.004	0.018
LIWC_Space	0.008	-0.042^{***}	-0.048^{**}
LIWC_Filler	0.040^{*}	0.009	0.003
LIWC_Ipron	0.041	0.031^{***}	0.073^{***}
LIWC_Time	-0.051^{**}	-0.007	0.001
LIWC_Quant	0.071^{***}	-0.001	0.094^{***}
LIWC_Discrep	-0.004	-0.018	-0.013
LIWC_You	0.081	-0.023	0.029
LIWC_Cause	-0.035	0.056^{***}	0.026
LIWC_Prep	0.109^{***}	0.044^{***}	0.132^{***}
LIWC_Relig	-0.039	-0.002	-0.001
LIWC_Body	0.001	-0.002	-0.001
LIWC_We	-0.061^{*}	-0.019^{*}	-0.040^{**}
LIWC_Assent	0.019	-0.026^{**}	-0.057^{**}
LIWC_Incl	0.017	-0.025^{**}	0.026
LIWC_Leisure	-0.039	-0.026^{**}	0.055^{***}
LIWC_AuxVb	0.102^{***}	0.030^{**}	0.158^{***}
LIWC_Hear	0.048^{**}	-0.006	0.013
LIWC_They	0.011	0.006	0.015
LIWC_Posemo	-0.029	-0.049^{***}	0.002
LIWC_Article	0.045^{**}	0.018	0.062^{***}
LIWC_Excl	0.074^{***}	0.002	0.099^{***}
LIWC_Home	-0.018	-0.006	-0.021
LIWC_Friends	0.012	0.017^{*}	-0.007
LIWC_Present	-0.014	-0.021	0.035
LIWC_Numbers	0.028	0.001	-0.011
Note:			*p<0.1; **p<0.05; ***p<0.01

Table 6: Additional regression coefficients for textual features in the mixed-effect logistic regression model predicting whether a question will receive a response, shown in 3. All numeric variables are z-scored.

	Own Wall	Other's Wall	Group
LIWC_I	0.160^{***}	0.022	0.113^{***}
LIWC_Work	-0.034	0.007	-0.066^{***}
LIWC_Tentat	0.023	0.037^{**}	0.044^{**}
LIWC_Ingest	0.045^{**}	0.017	0.046^{***}
LIWC_Motion	-0.001	-0.033^{***}	-0.006
LIWC_Anger	-0.021	0.004	-0.038
LIWC_Achiev	0.005	-0.011	0.050^{**}
LIWC_Swear	-0.004	0.00002	-0.028
LIWC_Death	-0.019	-0.0004	0.005
LIWC_Social	0.036	0.051^{***}	0.125^{***}
LIWC_Nonflu	0.020	0.016^{*}	-0.023
LIWC_Family	0.022	-0.023^{**}	0.005
LIWC_Feel	-0.0004	0.008	0.030^{*}
LIWC_Certain	0.024	-0.004	0.022
LIWC_Insight	0.069^{***}	0.052^{***}	0.027^{*}
LIWC_Humans	0.099^{***}	-0.005	0.037^{**}
LIWC_Sad	0.013	0.032^{**}	-0.005
LIWC_Past	-0.054^{*}	-0.045^{***}	-0.094^{***}
LIWC_See	-0.011	-0.0002	-0.008
LIWC_Future	-0.008	-0.028^{**}	-0.009
LIWC_Adverbs	-0.031	-0.035^{***}	0.039^{*}
LIWC_SheHe	0.006	0.001	-0.009
LIWC_Money	-0.004	-0.014	0.011
LIWC_Negate	-0.002	-0.065^{***}	0.006
LIWC_Health	0.020	-0.006	0.027
LIWC_Conj	0.026	0.037^{***}	0.057^{**}
LIWC_Anx	0.009	0.003	0.022
LIWC_Negemo	0.054^{*}	-0.012	0.102^{***}
LIWC_Sexual	0.025	-0.015	0.028
Note:			*p<0.1; **p<0.05; ***p<0.01

Table 7: Additional regression coefficients for textual features in the mixed-effect logistic regression model predicting whether a question will receive a response, shown in 3. All numeric variables are z-scored.

	Own Wall	Other's Wall	Group
topic_1	0.066	-0.139^{***}	-0.129^{***}
topic_2	0.034	-0.053^{***}	-0.008
topic_3	0.076^{**}	-0.048^{**}	0.046^{**}
topic_4	0.073^{**}	-0.021	0.057^{***}
topic_5	0.110^{*}	-0.095^{***}	-0.044
topic_6	0.034	-0.144^{***}	-0.090^{***}
topic_7	0.099	-0.305^{***}	-0.106^{*}
topic_8	0.023	-0.001	-0.115^{***}
topic_9	0.083^{**}	-0.039^{**}	0.060^{***}
topic_10	0.057^{*}	-0.024^{*}	-0.051^{**}
topic_11	0.069^{**}	-0.087^{***}	0.011
topic_12	0.056^{*}	-0.063^{***}	-0.006
topic_13	0.034	-0.119^{***}	-0.053^{**}
topic_14	0.022	-0.128^{***}	-0.076^{***}
topic_15	-0.008	-0.148^{***}	-0.141^{***}
topic_16	-0.005	-0.087^{***}	-0.045^{**}
topic_17	0.059^{**}	-0.051^{***}	0.027^{*}
topic_18	0.028	-0.072^{***}	-0.113^{***}
topic_19	0.068^{*}	-0.095^{***}	-0.021
Note:	*p<0.1; **p<0.05; ***p<0.01		

Table 8: Additional regression coefficients for textual features in the mixed-effect logistic regression model predicting whether a question will receive a response, shown in 3. All numeric variables are z-scored.

Topic #	Most Probable 20 Words
0	follow play i'm back game hey playing love games read martial favorite kind practice good you're arts book i've video
1	hey how's hope haven't workouts what's you're man back props things time prop good long bro miss dude coming workout
2	run running i'm week time training miles marathon half mile i've minutes today distance runs day long ran start tips
3	protein eat i'm good eating food healthy day diet drink water don't meal make breakfast i've ideas suggestions sugar it's
4	weight i'm fat i've lbs lose body calories day muscle week loss eating diet days eat lost pounds gain good
5	follow back goals hey you're fitness fito how's good awesome great luck happy fitocracy nice workouts what's training i'm coming
6	app fitocracy i'm group post workouts people find can't runkeeper don't it's site workout track iphone friends fito hey feed
7	starting check strength log running routine don't workouts forget show started amazing wanted bodyweight exercise view case guide guides trainers
8	i'm run hey area race year weekend tough running mudder live good marathon training group meet gym half follow spartan
9	i'm gym good weight strength lifting training i've program week start weights starting workout work don't routine exercises body workouts
10	people don't gym i'm it's guy you're make friends feel love didn't today that's girl time awesome told friend life
11	i'm i've time back it's feel work working week don't months weeks gym year tips started hard good ago exercise
12	back pain i'm i've knee it's left hurt good squats running sore shoulder lower today advice suggestions leg don't exercises
13	day today workout i'm work gym week days time morning tomorrow feel back rest good night it's feeling working run
14	workout log points workouts track fitocracy exercises i'm add exercise can't quests don't find time logging quest today day class
15	challenge level points group i'm join start challenges fitocracy day today hey duel time you're gonna month make i'll quests
16	bike points log count walking work hours today ride stairs exercise track walk workout miles day time cycling i'm running
17	shoes running i'm good wear pair suggestions i've buy music don't recommendations it's run fit wearing thinking love size nike
18	follow pic profile back props thx picture love prop nice based workout hey awesome you're fast what's lol where's great
19	weight press squats bar bench ups squat i'm exercises pull log barbell reps exercise leg set sets dumbbell machine back

Table 9: The most probable words for the topics used in predicting whether a question receives a response using the mixed-effect regression analysis in Table 2. Topics regression coefficients are reported in Table 8.

References

1. Alstott, J., Bullmore, E., Plenz, D.: powerlaw: a python package for analysis of heavy-tailed distributions. PloS one 9(1), e85777 (2014)
2. Althoff, T., Danescu-Niculescu-Mizil, C., Jurafsky, D.: How to ask for a favor: A case study on the success of altruistic requests. In: ICWSM (2014)
3. Althoff, T., Jindal, P., Leskovec, J.: Online actions with offline impact: How online social networks influence online and offline user behavior. In: WSDM (2017)
4. Andersson, L.M., Pearson, C.M.: Tit for tat? the spiraling effect of incivility in the workplace. Academy of management review 24(3), 452–471 (1999)
5. Aral, S., Muchnik, L., Sundararajan, A.: Distinguishing influence-based contagion from homophily-driven diffusion in dynamic networks. Proceedings of the National Academy of Sciences 106(51), 21544–21549 (2009)
6. Aral, S., Nicolaides, C.: Exercise contagion in a global social network. Nature Communications 8 (2017)
7. Artstein, R., Poesio, M.: Inter-coder agreement for computational linguistics. Computational Linguistics 34(4), 555–596 (2008)
8. Backstrom, L., Huttenlocher, D., Kleinberg, J., Lan, X.: Group formation in large social networks: membership, growth, and evolution. In: KDD (2006)
9. Bakshy, E., Rosenn, I., Marlow, C., Adamic, L.: The role of social networks in information diffusion. In: WWW (2012)
10. Bauman, A.E., Sallis, J.F., Dzewaltowski, D.A., Owen, N.: Toward a better understanding of the influences on physical activity: the role of determinants, correlates, causal variables, mediators, moderators, and confounders. American Journal of Preventive Medicine 23(2), 5–14 (2002)
11. Van den Berg, M.H., Schoones, J.W., Vlieland, T.P.V.: Internet-based physical activity interventions: a systematic review of the literature. Journal of Medica Internet Research 9(3), e26 (2007)
12. Billig, M., Tajfel, H.: Social categorization and similarity in intergroup behaviour. European Journal of Social Psychology 3(1), 27–52 (1973)
13. Blei, D.M., Ng, A.Y., Jordan, M.I.: Latent dirichlet allocation. The Journal of Machine Learning Research 3, 993–1022 (2003)
14. Budak, C., Agrawal, R.: On participation in group chats on twitter. In: WWW (2013)
15. Burke, M., Marlow, C., Lento, T.: Feed me: motivating newcomer contribution in social network sites. In: CHI (2009)
16. Cheng, J., Adamic, L., Dow, P.A., Kleinberg, J.M., Leskovec, J.: Can cascades be predicted? In: WWW. pp. 925–936. ACM (2014)
17. Cheung, C.M., Chiu, P.Y., Lee, M.K.: Online social networks: why do students use facebook? Computers in Human Behavior 27(4), 1337–1343 (2011)
18. Chilton, P.: Politeness, politics and diplomacy. Discourse & Society 1(2), 201–224 (1990)
19. Chou, H.T.G., Edge, N.: they are happier and having better lives than i am: the impact of using facebook on perceptions of others' lives. Cyberpsychology, Behavior, and Social Networking 15(2), 117–121 (2012)
20. Clauset, A., Shalizi, C.R., Newman, M.E.: Power-law distributions in empirical data. SIAM review 51(4), 661–703 (2009)
21. Conroy, M., Feezell, J.T., Guerrero, M.: Facebook and political engagement: A study of online political group membership and offline political engagement. Computers in Human Behavior 28(5), 1535–1546 (2012)
22. Cross, S., Markus, H.: Possible selves across the life span. Human development 34(4), 230–255 (1991)

23. Danescu-Niculescu-Mizil, C., Sudhof, M., Jurafsky, D., Leskovec, J., Potts, C.: A computational approach to politeness with application to social factors. In: ACL (2013)
24. Danescu-Niculescu-Mizil, C., West, R., Jurafsky, D., Leskovec, J., Potts, C.: No country for old members: User lifecycle and linguistic change in online communities. In: WWW (2013)
25. Dror, G., Pelleg, D., Rokhlenko, O., Szpektor, I.: Churn prediction in new users of yahoo! answers. In: WWW. pp. 829–834. ACM (2012)
26. Ducheneaut, N., Yee, N., Nickell, E., Moore, R.J.: Alone together?: exploring the social dynamics of massively multiplayer online games. In: CHI. pp. 407–416. ACM (2006)
27. Ellison, N.B., Gray, R., Vitak, J., Lampe, C., Fiore, A.T.: Calling all facebook friends: Exploring requests for help on facebook. In: ICWSM (2013)
28. Garcia, D., Mavrodiev, P., Casati, D., Schweitzer, F.: Understanding popularity, reputation, and social influence in the twitter society. Policy & Internet (2017)
29. Garcia, D., Mavrodiev, P., Schweitzer, F.: Social resilience in online communities: The autopsy of friendster. In: COSN (2013)
30. Goette, L., Huffman, D., Meier, S.: The impact of group membership on cooperation and norm enforcement: Evidence using random assignment to real social groups. American Economic Review 96(2), 212–216 (2006)
31. Goh, D.L., Razikin, K.: Is gamification effective in motivating exercise? In: Kurosu, M. (ed.) Human-Computer Interaction: Interaction Technologies, Lecture Notes in Computer Science, vol. 9170, pp. 608–617. Springer International Publishing (2015), `http://dx.doi.org/10.1007/978-3-319-20916-6_56`
32. Gonzales, A.L., Hancock, J.T.: Mirror, mirror on my facebook wall: Effects of exposure to facebook on self-esteem. Cyberpsychology, Behavior, and Social Networking 14(1-2), 79–83 (2011)
33. Gray, R., Ellison, N.B., Vitak, J., Lampe, C.: Who wants to know?: question-asking and answering practices among facebook users. In: CSCW (2013)
34. Hamari, J., Koivisto, J.: Social motivations to use gamification: An empirical study of gamifying exercise. In: ECIS. p. 105 (2013)
35. Hamari, J., Koivisto, J.: "working out for likes": An empirical study on social influence in exercise gamification. Computers in Human Behavior 50, 333–347 (2015)
36. Harper, F.M., Raban, D., Rafaeli, S., Konstan, J.A.: Predictors of answer quality in online q&a sites. In: CHI (2008)
37. Holmes, J., Stubbe, M., et al.: Power and politeness in the workplace: A sociolinguistic analysis of talk at work. Routledge (2015)
38. Horowitz, D., Kamvar, S.D.: The anatomy of a large-scale social search engine. In: WWW (2010)
39. Hyland, K.: Metadiscourse: Exploring Interaction in Writing. London: Continuum (2005)
40. Jeong, J.W., Morris, M.R., Teevan, J., Liebling, D.J.: A crowd-powered socially embedded search engine. In: ICWSM (2013)
41. Joyce, E., Kraut, R.E.: Predicting continued participation in newsgroups. Journal of Computer-Mediated Communication 11(3), 723–747 (2006)
42. Jung, Y., Gray, R., Lampe, C., Ellison, N.: Favors from facebook friends: unpacking dimensions of social capital. In: CHI. pp. 11–20. ACM (2013)
43. Jurgens, D., McCorriston, J., Ruths, D.: An analysis of exercising behavior in online populations. In: ICWSM (2015)
44. Kairam, S.R., Wang, D.J., Leskovec, J.: The life and death of online groups: Predicting group growth and longevity. In: WSDM (2012)
45. Karnstedt, M., Rowe, M., Chan, J., Alani, H., Hayes, C.: The effect of user features on churn in social networks. In: WebScience (2011)
46. Kayes, I., Chakareski, J.: Retention in online blogging: A case study of the blogster community. IEEE Transactions on Computational Social Systems 2(1), 1–14 (2015)

47. Kelman, H.: Social influence and linkages between the individual and the social system: Further thoughts on the processes of compliance, identification, and internalization. In J. Tedeschi (Ed.), Perspectives on social power (1974)
48. Kiritchenko, S., Zhu, X., Mohammad, S.M.: Sentiment analysis of short informal texts. Journal of Artificial Intelligence Research (JAIR) 50, 723–762 (2014)
49. Korda, H., Itani, Z.: Harnessing social media for health promotion and behavior change. Health promotion practice 14(1), 15–23 (2013)
50. Kumar, S., Zafarani, R., Liu, H.: Understanding user migration patterns in social media. In: AAAI (2011)
51. Lampe, C., Resnick, P.: Slash (dot) and burn: distributed moderation in a large online conversation space. In: CHI. pp. 543–550. ACM (2004)
52. Lampe, C., Vitak, J., Gray, R., Ellison, N.: Perceptions of facebook's value as an information source. In: CHI (2012)
53. Lento, T., Welser, H.T., Gu, L., Smith, M.: The ties that blog: Examining the relationship between social ties and continued participation in the wallop weblogging system. In: WWW (2006)
54. Li, H., Bhowmick, S.S., Sun, A.: Casino: towards conformity-aware social influence analysis in online social networks. In: CIKM (2011)
55. Luczak-Roesch, M., Tinati, R., Simperl, E., Van Kleek, M., Shadbolt, N., Simpson, R.J.: Why won't aliens talk to us? content and community dynamics in online citizen science. In: ICWSM (2014)
56. Ma, X., Chen, G., Xiao, J.: Analysis of an online health social network. In: Proceedings of the 1st ACM international health informatics symposium. pp. 297–306. ACM (2010)
57. Maher, C.A., Lewis, L.K., Ferrar, K., Marshall, S., De Bourdeaudhuij, I., Vandelanotte, C.: Are health behavior change interventions that use online social networks effective? a systematic review. Journal of Medical Internet Research 16(2) (2014)
58. Marcus, B.H., Selby, V.C., Niaura, R.S., Rossi, J.S.: Self-efficacy and the stages of exercise behavior change. Res. Quarterly for Exercise and Sport 63(1), 60–66 (1992)
59. Michael, L., Otterbacher, J.: Write like i write: Herding in the language of online reviews. In: ICWSM (2014)
60. Mohammad, S.M., Kiritchenko, S.: Using hashtags to capture fine emotion categories from tweets. Computational Intelligence 31(2), 301–326 (2015)
61. Morris, M.R., Teevan, J., Panovich, K.: A comparison of information seeking using search engines and social networks. In: ICWSM (2010)
62. Morris, M.R., Teevan, J., Panovich, K.: What do people ask their social networks, and why?: a survey study of status message q&a behavior. In: CHI (2010)
63. Muise, A., Christofides, E., Desmarais, S.: More information than you ever wanted: Does facebook bring out the green-eyed monster of jealousy? CyberPsychology & behavior 12(4), 441–444 (2009)
64. Nakhasi, A., Shen, A.X., Passarella, R.J., Appel, L.J., Anderson, C.A.: Online social networks that connect users to physical activity partners: a review and descriptive analysis. Journal of medical Internet research 16(6) (2014)
65. Newell, E., Jurgens, D., Saleem, H.M., Vala, H., Sassine, J., Armstrong, C., Ruths, D.: User migration in online social networks: A case study on reddit during a period of community unrest. In: ICWSM (2016)
66. Norman, G.J., Zabinski, M.F., Adams, M.A., Rosenberg, D.E., Yaroch, A.L., Atienza, A.A.: A review of ehealth interventions for physical activity and dietary behavior change. American journal of preventive medicine 33(4), 336–345 (2007)
67. Paul, S.A., Hong, L., Chi, E.H.: Is twitter a good place for asking questions? a characterization study. In: ICWSM (2011)

68. Plotnikoff, R.C., Mayhew, A., Birkett, N., Loucaides, C.A., Fodor, G.: Age, gender, and urban–rural differences in the correlates of physical activity. Preventive medicine 39(6), 1115–1125 (2004)
69. Prochaska, J.O., Velicer, W.F.: The transtheoretical model of health behavior change. American journal of health promotion 12(1), 38–48 (1997)
70. Raban, D., Harper, F.: Motivations for answering questions online. New media and innovative technologies (2008)
71. Ren, Y., Harper, F.M., Drenner, S., Terveen, L.G., Kiesler, S.B., Riedl, J., Kraut, R.E.: Building member attachment in online communities: Applying theories of group identity and interpersonal bonds. Mis Quar. 36(3), 841–864 (2012)
72. Rogers, P.S., Lee-Wong, S.M.: Reconceptualizing politeness to accommodate dynamic tensions in subordinate-to-superior reporting. Journal of Business and Technical Communication 17(4), 379–412 (2003)
73. Rzeszotarski, J.M., Spiro, E.S., Matias, J.N., Monroy-Hernández, A., Morris, M.R.: Is anyone out there?: unpacking q&a hashtags on twitter. In: CHI (2014)
74. Salganik, M.J., Dodds, P.S., Watts, D.J.: Experimental study of inequality and unpredictability in an artificial cultural market. Science 311(5762), 854–856 (2006)
75. Samuel, G.O.: Language and politics: Indirectness in political discourse. Discourse & Society 8(1), 49–83 (1997)
76. Schoenebeck, S.Y.: The secret life of online moms: Anonymity and disinhibition on youbemom. com. In: ICWSM (2013)
77. Sherwood, N.E., Jeffery, R.W.: The behavioral determinants of exercise: implications for physical activity interventions. Annual review of nutrition 20(1), 21–44 (2000)
78. Stults-Kolehmainen, M.A., Ciccolo, J.T., Bartholomew, J.B., Seifert, J., Portman, R.S.: Age and gender-related changes in exercise motivation among highly active individuals. Athletic Insight 5(1), 45 (2013)
79. Tausczik, Y.R., Pennebaker, J.W.: The psychological meaning of words: Liwc and computerized text analysis methods. Journal of language and social psychology 29(1), 24–54 (2010)
80. Teevan, J., Morris, M.R., Panovich, K.: Factors affecting response quantity, quality, and speed for questions asked via social network status messages. In: ICWSM (2011)
81. Teng, C.Y., Adamic, L.A.: Longevity in second life. In: ICWSM (2010)
82. Umberson, D., Crosnoe, R., Reczek, C.: Social relationships and health behavior across life course. Annual review of sociology 36, 139 (2010)
83. Vissers, S., Stolle, D.: Spill-over effects between facebook and on/offline political participation? evidence from a two-wave panel study. Journal of Information Technology & Politics 11(3), 259–275 (2014)
84. Yee, N., Bailenson, J.N., Urbanek, M., Chang, F., Merget, D.: The unbearable likeness of being digital: The persistence of nonverbal social norms in online virtual environments. CyberPsychology & Behavior 10(1), 115–121 (2007)
85. Zhu, H., Kraut, R.E., Kittur, A.: The impact of membership overlap on the survival of online communities. In: CHI (2014)