Online sharing of physical activity: does it accelerate the impact of a health promotion program?

A. Manzoor, J.S. Mollee, E.F.M. Araújo, A.T. van Halteren, M.C.A. Klein
Behavioural Informatics Group
Department of Computer Science
VU University Amsterdam
Amsterdam, The Netherlands
Email: {a.manzoorrajper, j.s.mollee, e.araujo, a.t.van.halteren, m.c.a.klein}@vu.nl

***Abstract*— Influence on health behavior from peers is well known and it has been shown that participants in an online physical activity promotion program are generally more successful when they share their achievements through an online community. However, more detailed insights are needed into the mechanisms that explain the influence of a community on physical activity level (PAL).**

This paper discusses a detailed analysis of a data set of participants in an online physical activity promotion program. The analysis focuses on a method to compare community members with non-community members that eliminates to a large extent factors that dilute the community effect. We create statistical models that describe the PAL increase at the end of the program. A comparison of these models shows that community members not only have a higher PAL at the start of the program, but also that the PAL increase is significantly greater compared to non-community members.

The results further support the hypothesis that stimulating participants to share their achievements with peers makes physical activity programs more successful to help people achieve a healthy activity level.

***Index Terms*—data analysis; physical activity promotion; social networks; social influence**

I. Introduction

Engaging in sufficient physical activity has many beneficial effects on physical and mental health [1], [2], while low levels of physical activity have been associated with increased risks of cardiovascular diseases, cancer, diabetes, and mental illness [3]. Despite this, a large proportion of the Western population does not meet the guidelines of being moderately to vigorously active for at least 30 minutes on at least five days a week [4]. Therefore, physical activity promotion is a priority in most Western countries and many (online) intervention programs exist. It is important to understand which elements of these physical activity program are effective or could potentially accelerate the impact of these health promotion programs.

Previous research has already revealed that being part of an online social network in a health promotion program is correlated with a higher level of physical activity. In an earlier study based on a data set of 4,333 participants [5], it was shown that the activity level of people that participated in an intervention aiming at stimulating physical activity (for 14 weeks) who became a member of an online community was significantly higher compared to people that chose not to become a member of that community. However, based on this result, it is not possible to conclude that a higher level of physical activity is the *result* of being member of an online social network. It is also possible that there is a selection bias: more active people are more willing to participate in such a community. Besides that, there are other possibly confounding factors. For example, the people participating in a community could be selected in a season that is more suitable for physical activity, the participants could be recruited in companies with higher average activity levels, or the people that opt in for a community are biased with respect to gender or have psychological traits that makes them more intrinsically motivated to share their achievements.

In this paper, we also compare the level of physical activity of people that are member of an online social network within a physical activity promotion program and people that choose not to. We build on the previous work in two ways. First, we perform a very balanced selection in the two groups of people that we compare: we compensate for possible differences in starting season, in country and in gender. This way, we can rule out a number of alternative explanations for the higher lever of physical activity of community members. Second, we compare the *change* in activity level over the time of the intervention, rather than focusing on static differences between the two groups. That is, we see whether the intervention aiming at increasing physical activity has a different effect if people are member of an online community. In other words: does the online community accelerate the impact of the physical activity promotion program? The remainder of this paper is organized as follows. In Section II, we discuss some related work about the influence of online social networks and (mobile) healthy lifestyle interventions on behavior. Section III presents the data, its characteristics and our methods for processing it. In Section IV, we provide statistical analyses to answer our research questions. We conclude with our main findings and directions for future work in Section V and Section VI.

II. Related work

Previous analyses [5] show that there is a positive relation between being part of the online community of a physical activity intervention and the physical activity level of participants. The online community therefore matters. It has also been shown that the number of contacts in the online community does not have a significant effect on the physical activity level, while network density even has a significant, negative effect. On the other hand, adding online community features to an Internet-mediated walking program did not increase average daily step counts, but did reduce participant attrition [6].

Online social interaction plays an important role in forming or adapting some kind of behavior based on the peer's behavior. It has been studied recently that online social networks are equally responsible (as offline networks) in the diffusion of one's emotions to another [7]. It is often difficult to adopt new behavior and adhere to it, but it has been shown that close social circles (such as family, friends, and co-workers) are helpful in sustaining a healthy lifestyle [8], [9]. In [10], the role of online social interactions is discussed in the context of developing and maintaining a healthy lifestyle, e.g. an ambient system can continuously monitor and help people to alter their social ties in order to sustain healthy behavior. Having an infrastructure like a social network already available, social network interventions could be designed to leverage the full potential of a social network [11], for example in case of a health behavior change program.

With the rise of mobile technology, there has also been a steep increase in the number of healthy lifestyle interventions that are available through a smartphone. As of May 2016, the number of apps in the Health & Fitness category has grown to 67,552 for the Google Play Store [12] and 68,248 for the iTunes App Store [13]. A systematic review of apps that promote physical activity has shown that even though most apps apply only a few behavior change techniques [14] [15], a majority of these apps (approximately 58%) provided a form of social support or social change [14]. This was done, for example, through providing chat possibilities among users or through enabling a link to an external virtual social network, where users could share their goals or achievements [14].

It is widely believed that mobile technology can be a useful tool to promote physical activity among a large part of the population. First, average smart phone ownership numbers are high: 68% in the United States, with higher numbers among young adults (86% in ages 18-29 years and 83% in ages 30-49 years) [16], and 80% in The Netherlands [17] in the third quarter of 2015. This means that interventions designed for smartphones can theoretically reach a large number of people. Second, mobile interventions are always accessible to the user, and also allow for continuous monitoring and (if applicable) feedback. Also, similar to interventions delivered over the Internet, mobile interventions can reduce stigma and lower the barrier for people to address their (health) issues [18]. In combination with the relatively high number of apps that enable social support or social change, these advantages of mobile interventions imply that smartphone apps are a very suitable means to guide social influence for behavior change.

III. Data set

This section describes the data set that is used for the analysis. In Section III-A, we describe the process of data collection and the resulting data. Section III-B describes the way in which we processed the data to select suitable subsets, and some of the structural characteristics of the selected social network components are presented in Section III-C.

A. Data collection

The analysis presented in this paper uses a data set of people (n=50,000) that participated in an online physical activity promotion program. The promotion program has three different phases. The first phase is a one-week assessment period, that is used to evaluate the user's activity level during his/her daily routine. The assessment is followed by the second phase: a 12-week plan that aims to gradually increase the user's activity level towards a specified end goal. The goal is determined based on the physical activity reported during the assessment week. After the plan, the members of the program can opt to start a new 12-week plan to further increase their activity level or simply continue with the activity goal set during the last week of their program. This constitutes the third phase.

The activity promotion program provides an online community and joining this social network is optional for the users. Each member of the community can connect to other users (i.e., become online friends), exchange messages and see the relative achievements of themselves and their connections (which are only visible after the participants confirm their connection). Around 5,000 people in the data set opted to join the online community.

The participants in the program wear an activity monitor device that measures their physical activity level (PAL). When they register to the program via the website, the participants fill in their gender, age, and nationality. In addition, the data set contains information about the date that people start the program, the company they work in (if the program is offered via a company), and their friendship connections with other participants. In order to ensure anonymity of the participants, their age was omitted from the data set before the analysis.

B. Data selection

As our aim is to compare the change in the physical activity level of people that are part of a community with people that are not member of a community, we select two subsets of the data. The first subset is the intervention group: participants in the physical activity promotion program that opted in for the community. The second subset is the control group: participants in the physical activity promotion program that did not opt in for the community. In this section, we describe how we selected those two subsets.

The data is represented in two files. One file is a GEXF (Graph Exchange XML Format) file and represents the network structure of the community. The other file consists of the PAL values of all participants, and their characteristics, such as gender, BMI (Body Mass Index), corporation and country.

As we want to be able to consider the mutual effect of friendship relations on the activity level of participants, we select a number of connected components from the community. A connected component (or just component) of an undirected graph is a subgraph in which any two nodes are connected to each other by edges, and which is connected to no other nodes in the supergraph. In order to extract the components, we use Python's NetworkX library [19], which is based on the community detection algorithm Tarjan's algorithm with Nuutila's modifications [20], [21]. The algorithm is based on the principle of strongly connected components, where each node in a graph has a bidirectional connection. The total number of communities that are found by the algorithm is 395. One of them is a large community with 3,926 participants; the second largest community consists of 42 participants. Figure 1 shows an overview of the number of participants in each of the components. The components are ordered by size, and the largest component (of 3,962 participants) is left out.

For all connected components, we extract the PAL (physical activity level) values for the individuals in each of the components. Since there are multiple consecutive plans (i.e., periods of twelve weeks in which people are stimulated to increase their activity level), the PAL values used in the analysis represent the first 12-week plan, in order to ensure fair comparisons.

Not all detected network components are used in the analysis. We select only components with (1) a limited number of participants for whom other data is missing and (2) a minimal difference between the plan start dates of the members of the component. For some participants in the online community, no other personal data or PAL data was available. Only components with at most one such participant were eligible for inclusion. For the (earliest and latest) start dates of the plans in the component, the maximum difference is four months. This is done to ensure that the participants in the community were using the program around the same time, so the community was 'active'. As a result, we discard the largest component, because of the fact that the earliest and latest start dates are three years apart. The second component with 42 nodes is not included in the analysis because a lot of data is missing for that component.

This selection process yielded ten of such connected components, consisting of 109 individuals in total. We left out 25 individuals for whom PAL values were missing for one or more weeks, for instance because they dropped out of the program. Eventually, this resulted in 84 individuals in the intervention group.

For the control group, we select a set of individuals who did not opt for the community, but who are otherwise similar to the participants in each of the components in the intervention group. We balance the data with respect to the following characteristics: the participants work in the same companies, their plan earliest and latest start and date are similar to the corresponding component, and their gender ratio is also similar to the matching component. As the number of non-community individuals is much larger than the number of individuals within a community, we randomly select a set of around five times the size of the number of people in the community component with corresponding characteristics, resulting in a set of 501 people. For this data set, we also avoid including individuals with missing data, i.e. individuals who dropped out of the program or who had missing PAL data for one or more weeks. In total, this resulted in a set of 498 participants. Based on the selection of the individuals in both groups, the PAL values are extracted for the two subsets.

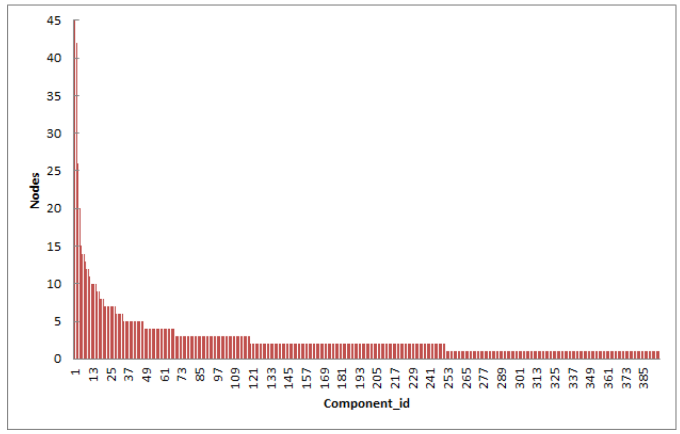

Fig. 1. Number of nodes in each of the components

A summary of the selected data is given in Table I and Table II. In Table I, each row shows several meta-data characteristics of the selected components of the network. 'Component' is the id of the network component, and 'Number of Participants' represents the total number of participants in the component. It can be seen that the size of the selected components varies between 7 and 26 participants. 'Dropouts' shows the number of people that were omitted from the component, because at least one week of PAL data was missing. The 'Start (earliest)' and 'Start (latest)' columns show the earliest or the latest date on which people in a component started their first plan.

The last column shows the number of individuals with similar characteristics identified in the non-community data set and the number of them that were randomly selected for the control group. As mentioned earlier, since the number of people that did not opt in for the community is much larger than the number of people who did, the size of the control group is about five times the intervention group size. For example, for component A (consisting of 26 participants), 130 participants were randomly selected from a set of 2,735 individuals with similar characteristics. However, for some components, we could not find enough individuals with similar characteristics for the non-community data subset. For example, only six individuals were found for the non-community data subset corresponding to component G.

Table II illustrates different characteristics of people in the components, such as their nationality. The ‘Country’ column shows that the participants in each of the communities are from the same country, namely Germany, the Netherlands or the USA. ‘Number of Corporations’ shows whether all people in a certain component work in the same or different organizations. It is possible that people in a community work in different organizations, like in components C, H and J. In rest of the components, the participants all work in the same company. The column ‘Gender Ratio’ provides information about the ratio of male and female participants in each of the communities. ‘Average BMI’ represents the average BMI for each of the components.

C. Structural analysis of the components

As described in the previous section, we selected 10 components from the community for the analysis, ranging from 7 to 26 participants each and with different configurations. The difference in the structural characteristics between the components can be seen in Figure 2. Social network analyses were run on the components in order to understand the structure of the connections.

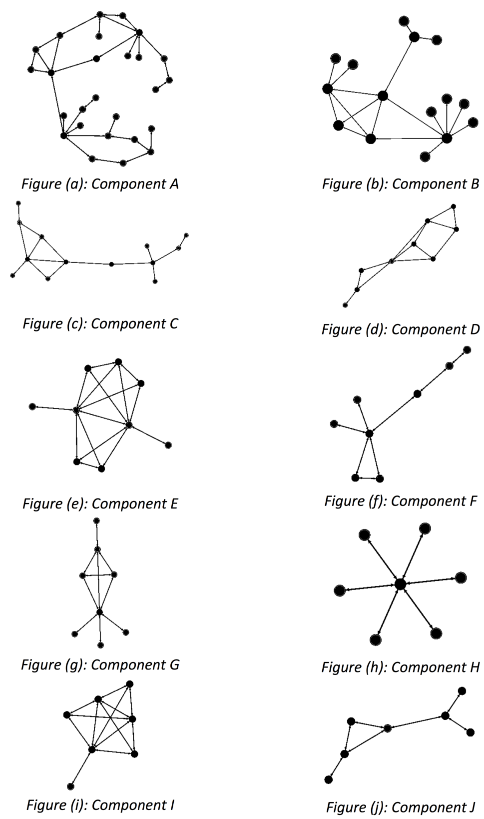

Fig. 2. The network components used for the community analysis.

The components are mostly sparse networks with a low average degree and low clustering coefficient, meaning that the neighbors of each node are not well connected among themselves. All nodes within each component belong to the same country, the countries being the Netherlands, the USA and Germany. Because of the nature of the online friendship connections, all connections in the network are bidirectional.

Details of two components will be given to further illustrate the data. Component I has the highest average density and clustering coefficient, both more than 60%. It also presents a small diameter, which means that the nodes are very well connected, and are very close to each other. In this network, the degree of the nodes ranges from 2 to 12, with every connection being bidirectional. One of the nodes with the highest degree is connected to all the other nodes in the network, having an important role for the social influences in this component.

Component E is also a well-connected component with a small network diameter, as in most social networks in real life. This component has an average density and clustering coefficient of around 50%, which makes the network well connected, but not very dense. Two nodes have only one connection, and the rest of the network presents a very good clustering coefficient. Table III shows the detailed characteristics of this component.

IV. Results

We perform several steps to answer our main question: is the change in the physical activity level of people that are part of an (online) community different from people that are not member of such a community?

A. Visual comparison

Our first analysis is based on a visual comparison of the differences between the two groups. The average PAL values for both groups during twelve weeks (84 days) are shown in Figure 3. The figure illustrates that more active people are part of the community, since their average PAL is higher than the average PAL of the non-community participants. It also shows that the linear trendline of both groups has a different slope.

B. Multiple linear regression model

For a more thorough analysis, we use statistical methods. In the second step of the analysis, a multiple linear regression model is fitted to predict the average physical activity level at the end of the program (i.e., the last three weeks) based on whether a person is member of the community and the average PAL at the start of the program as predictors. For the average PAL at the start of the program, we consider only the second and third week. The first week is left out, because this week is usually a bit atypical, presumably due to novelty effects of starting the program.

To measure the average difference between groups, a dummy variable (*Community*) is coded with the value ’1’ if a person is in the community and ’0’ if the person did not opt in for the community. The results are illustrated in Table IV. A significant regression model was found ($F(2,579) = 227$, $p < .001$). The model accounts for 44% of the variance in the PAL values of the participants at the end of the program, $R^2 = .4395$. Both predictor variables, *Start-PAL* and

TABLE I
META-PROPERTIES OF SELECTED COMPONENTS

Component	Number of Participants	Dropouts	Start (earliest)	Start (latest)	Number of Participants Non-Community
A	26	0	25/01/2010	22/03/2010	130 / 2,735
B	15	4	15/02/2010	26/04/2010	70 / 838
C	13	4	18/05/2009	17/05/2010	65 / 178
D	9	1	16/05/2009	20/07/2009	45 / 74
E	9	3	25/01/2010	12/04/2010	45 / 2,839
F	8	0	25/05/2009	19/04/2010	40 / 608
G	8	6	19/04/2010	21/06/2010	6 / 6
H	7	6	15/02/2010	07/06/2010	30 / 35
I	7	1	22/02/2010	22/03/2010	35 / 358
J	7	0	02/03/2009	27/07/2009	35 / 335

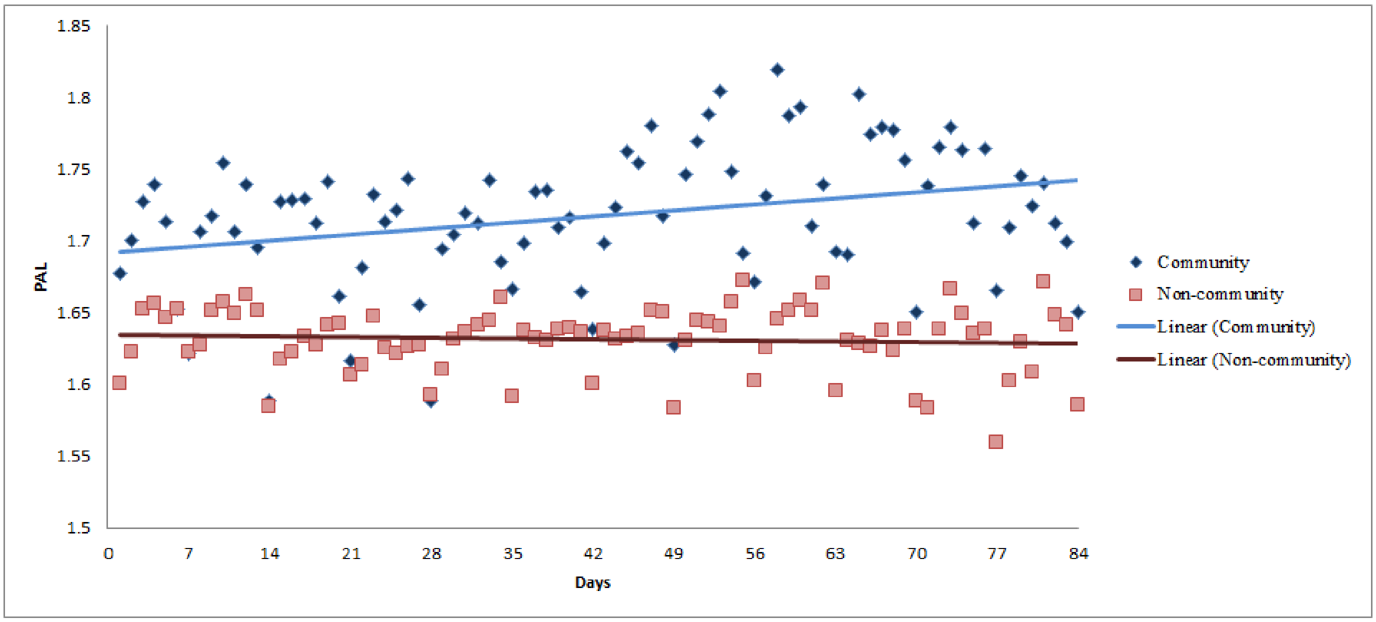

Fig. 3. Physical activity levels (PALs) of community vs. non-community participants during 84 days of the intervention, including linear trendlines.

TABLE II
CHARACTERISTICS OF PARTICIPANTS IN SELECTED COMPONENTS

Component	Country	Number of Companies	Gender Ratio (M/F:%)	Average BMI
A	DE	1	M:88.5, F:11.5	25.85
B	NL	1	M:100.0, F:0.0	25.74
C	US	3	M:47, F:53	24.37
D	US	1	M:89, F:11	31.76
E	DE	1	M:34, F:66	23.97
F	US	1	M:87, F:13	30.78
G	NL	1	M:58, F:42	25.60
H	NL	4	M:100.0, F:0.0	32.10
I	NL	1	M:86, F:14	28.05
J	DE	3	M:86, F:14	25.65

TABLE III
DETAILED CHARACTERISTICS OF COMPONENT E

Number of Nodes	9
Edges	32
Average Degree	3.556
Average Path Length	1.583
Network Diameter	3
Density	0.444
Average Clustering Coefficient	0.622
Country	Germany

Community, are statistically significant, $p < .05$. The model shows that the predicted PAL for the last three weeks is equal to 0.23562 + 0.05061 * *Community* + 0.85041 * *Start-PAL*, where *Community* is 0 or 1. The model signifies that being member of a community is associated with an increase of approximately 0.05 in physical activity level.

C. Linear mixed model

The regression model described in the previous section only compares the PAL at the start with the PAL at the end. A linear mixed model can be used to take into account all days of data (except for the first week, as mentioned above). Since the data are longitudinal by nature, we follow the approach as outlined in [22]. A sample of the data is shown in Table V. Each row represents one day's PAL for an individual, and there are 77 rows (eleven weeks) for each individual. As discussed in [22], we first conduct the test using a simple model based

TABLE IV
ANALYSIS USING MULTIPLE LINEAR REGRESSION

	Estimate	Std. Error	*p*-value	95% C.I.
(Intercept)	0.23562	0.06743	<.001	[0.103183, 0.36806]
Start-PAL	0.85041	0.04091	<.001	[0.77005, 0.93076]
Community=1	0.05061	0.02327	.0300	[0.00490, 0.09631]

on the generalized least square method and later add random effects to the intercepts in the simple model to see if the two models differ significantly. For this purpose, R's NLME library is used [23]. Since we are primarily interested to see whether being in a community makes a difference over time, the model includes an interaction term, i.e. a product of *Community* and *Time*. The results of the simple model (without random effects) are shown in Table VI. Here, being part of the community is taken as the reference group (*Community*=1), in contrast to the model presented in Table IV. The estimates associated with the predictor variables indicate the effect of the program on the PAL. So, the interaction term tests whether the effect of the program on the PAL of the participants is different for people inside or outside the community. The results show that this is indeed the case: people perform differently in the two groups. In this analysis, not being part of the community is again associated with a lower PAL value (with a difference of approximately 0.06).

TABLE V
PHYSICAL ACTIVITY LEVEL DATA IN LONG FORMAT

Id	Time	PAL	Community
1	8	1.57800	1
1	9	1.85780	1
1	10	1.78080	1
.	.	.	.
.	.	.	.
582	82	1.5803	0
582	83	1.7658	0
582	84	1.4576	0

TABLE VI
ANALYSIS USING GENERALIZED LINEAR REGRESSION

	Estimate	Std. Error	*t*-value	*p*-value
(Intercept)	1.6933389	0.009814656	172.53167	<.001
Community=0	−0.0586920	0.010610159	−5.53168	<.001
Time	0.0005821	0.000192112	3.02993	.0024
Community=0 : Time	−0.0006525	0.000207683	−3.14177	.0017

The results of the more advanced random intercept model are shown in Table VII. In this model, we account for the fact that the start PAL (i.e., the intercept) of each of the individuals is different by adding a mixed effect for this value. The results show that there are only some small differences in the standard error compared to the simple model. Similar to the model in Table VI, being part of the community is taken as the reference group (*Community*=1). The intercept therefore represents the predicted PAL scores for the people in the community, and the estimated coefficient for *Community*=0 indicates the difference between the predicted PAL for the people in the community and the people who are not part of the community. The coefficient of *Time* indicates that for every unit of time, there is an increase of 0.0005821 in the PAL for people in the community. The estimated coefficient for the interaction term represents the difference in the slope for the two groups. In other words, the interaction term tells us that the two groups (community vs. non-community) show a significantly different change in PAL over a period of twelve weeks.

The likelihood ratio test is often conducted to test the significance of predictor variables, i.e. to compare the fit of one model (with a reduced set of predictors variables) to the fit of another model (with a complete set of predictor variables). Here, we also use this test to see which model provides a better fit for data. Model 1 is based on a generalized linear regression (Table VI) and model 2 is based on a linear mixed effects model (Table VII). The latter includes all the variables of model 1, plus an additional mixed effect for the individuals' intercepts. The results are shown in Table VIII. The null hypothesis (stating that the between-subject variation in the intercept is equal to zero) is rejected, $\chi^2(1)$ = 17882.63, $p < .001$. This tells us that adding a random effect for the individuals to the model is a significant improvement, therefore the mixed effect model provides better fit for the data.

TABLE VII
ANALYSIS USING LINEAR MIXED EFFECTS MODELING

	Estimate	Std. Error	df	*t*-value	*p*-value
(Intercept)	1.6933389	0.02388	44230	70.901	<.001
Community=0	−0.0586920	0.02581	580	−2.273	.0234
Time	0.0005821	0.00015	44230	3.791	<.001
Community=0 : Time	−0.0006525	0.00016	44230	−3.931	<.001

V. DISCUSSION

The main research question that is investigated in this paper is whether the intervention aiming to increase physical activity has a different effect if people are member of an online community. Two statistical analyses were performed. In the first analysis, a significant linear regression model was found. Based on the adjusted R^2, we conclude that 44% of the variance in the PAL values is explained by this model. In the second analysis, a linear mixed model was fitted on the whole data set (eleven weeks), which shows that there is a significant difference between the increase in PAL of the two groups, even when a random factor for the start PAL is taken into account. It can thus be inferred from the results that on average people have a larger difference in activity level between the start and end of the program when they are member of an online community, compared to people that are not part of this community. This answers our question positively.

However, these findings do not yet answer the question why this is happening. It is still unclear what kind of social phenomenon causes this effect of the community on the

TABLE VIII
COMPARISON OF STANDARD LINEAR REGRESSION MODEL WITH RANDOM INTERCEPT MODEL

Model	df	logLik	Chi Sq.	p-value
standard (1)	5	−15712.890		
advanced (2)	6	−6771.574	17882.63	<.001

participants' physical activity levels. One hypothesis is that it is caused by social contagion, i.e. the process of influencing others (sometimes unconsciously) via a network of social relations [24]. Another possible explanation is social support, in the sense that community members help each other in performing physical activities (e.g., doing sports together) [25]. Yet another hypothesis is that social comparison is the driving factor, i.e. that people that chose to share their physical activity level online are stimulated by the achievements of others [26]. These questions require further investigation and provide directions for further research.

The visual representation in Figure 3 of the PAL during the period of the intervention shows – apart from a different slope for the two groups – also two other interesting aspects. First, a regular pattern of peaks and dips in both groups can be seen. Since each participant always starts his/her plan on a Monday, the data is aligned per weekday. Our explanation is that the dips correspond to weekends, in which people are on average less active. Second, we see that the PAL of the people that are not part of a community does not increase at all, even though they participate in a physical activity promotion program. There is no obvious explanation for this fact, it seems that the intervention without the community aspect is only effective for the first two weeks.

VI. CONCLUSION

Sharing achievements online in a physical activity promotion program does make a difference for the users of the program. A data set of approximately 50,000 individuals was used to extract data that ensured a fair comparison between participants that are part of the online community and participants that are not. From the set of approximately 5,000 individuals that opted in for the community (consisting of a collection of several smaller connected components), a number of components was selected based on specific inclusion criteria. Based on the characteristics of those sub-communities, similar individuals were found from the set of individuals who were not tied to any community.

The two data sets were analyzed and compared with each other. We were able to conclude that there is a difference in PAL, as the users in the community are already more active at the start. This confirms findings from earlier work [5]. Also, we were able to conclude that the PAL of people in a community shows an increase that is significantly greater compared to non-community users. Since we balanced the data sets for possibly confounding factors like gender, time of the year and corporation, it is very likely that the fact that people are member of the community is the dominant factor that makes a difference for their increase in physical activity level. Although we do not know the direction of the causal relation, our findings are a valuable step towards answering the question "does online sharing of physical activity accelerate the impact of a health promotion program". We can conclude that participating in an online social network for sharing activity data is associated with an increase in physical activity. Similarly, not being part of a community is correlated with no increase in physical activity. These results further support the hypothesis that enabling participants to share their achievements with peers makes physical activity programs more successful to help people achieve a healthy activity level.

In future work, we plan to use an existing computational model of social contagion [27] to see whether this model can explain and predict the change. Also, it would be interesting to consider the effect of other factors on the physical activity level, such as the community size and structure. That way, research can further uncover phenomena that are at the basis of the beneficial effects of online social networks in health promotion programs.

Acknowledgments: This research is supported by Philips and Technology Foundation STW, Nationaal Initiatief Hersenen en Cognitie NIHC under the partnership program Healthy Lifestyle Solutions.

E.F.M. Araújo's stay at the VU University Amsterdam is funded by the Brazilian Science without Borders Program, through a fellowship given by the Coordination for the Improvement of Higher Education Personnel CAPES (reference 13538-13-6).

REFERENCES

[1] V. S. Conn, A. R. Hafdahl, and D. R. Mehr, "Interventions to increase physical activity among healthy adults: meta-analysis of outcomes," *American journal of public health*, vol. 101, no. 4, pp. 751–758, 2011.

[2] R. M. Eime, J. A. Young, J. T. Harvey, M. J. Charity, W. R. Payne *et al.*, "A systematic review of the psychological and social benefits of participation in sport for children and adolescents: informing development of a conceptual model of health through sport," *Int J Behav Nutr Phys Act*, vol. 10, no. 98, p. 1, 2013.

[3] I.-M. Lee, E. J. Shiroma, F. Lobelo, P. Puska, S. N. Blair, P. T. Katzmarzyk, L. P. A. S. W. Group *et al.*, "Effect of physical inactivity on major non-communicable diseases worldwide: an analysis of burden of disease and life expectancy," *The lancet*, vol. 380, no. 9838, pp. 219–229, 2012.

[4] W. L. Haskell, I.-M. Lee, R. R. Pate, K. E. Powell, S. N. Blair, B. A. Franklin, C. A. Macera, G. W. Heath, P. D. Thompson, and A. Bauman, "Physical activity and public health: updated recommendation for adults from the american college of sports medicine and the american heart association," *Circulation*, vol. 116, no. 9, p. 1081, 2007.

[5] M. Groenewegen, D. Stoyanov, D. Deichmann, and A. van Halteren, "Connecting with active people matters: the influence of an online community on physical activity behavior," in *Social Informatics*. Springer, 2012, pp. 96–109.

[6] C. Richardson, L. Buis, A. Janney, D. Goodrich, A. Sen, M. Hess, K. Mehari, L. Fortlage, P. Resnick, B. Zikmund-Fisher, V. Strecher, and J. Piette, "An online community improves adherence in an internet-mediated walking program. part 1: Results of a randomized controlled trial," *J Med Internet Res*, vol. 12, 2010.

[7] L. Coviello, Y. Sohn, A. D. Kramer, C. Marlow, M. Franceschetti, N. A. Christakis, and J. H. Fowler, "Detecting emotional contagion in massive social networks," *PloS one*, vol. 9, no. 3, p. e90315, 2014.

[8] R. S. Zimmerman and C. Connor, "Health promotion in context: the effects of significant others on health behavior change," *Health Education & Behavior*, vol. 16, no. 1, pp. 57–75, 1989.

[9] L. H. McNeill, M. W. Kreuter, and S. Subramanian, “Social environment and physical activity: a review of concepts and evidence,” *Social science & medicine*, vol. 63, no. 4, pp. 1011–1022, 2006.
[10] W. Van Breda, J. Treur, and A. Van Wissen, “Analysis and support of lifestyle via emotions using social media,” in *Social Informatics*. Springer, 2012, pp. 275–291.
[11] M. C. A. Klein, A. Manzoor, J. S. Mollee, and J. Treur, “Effect of changes in the structure of a social network on emotion contagion,” in *Web Intelligence (WI) and Intelligent Agent Technologies (IAT), 2014 IEEE/WIC/ACM International Joint Conferences on*, vol. 3. IEEE, 2014, pp. 270–277.
[12] AppBrain, May 2016.
[13] PocketGamer.biz, May 2016.
[14] A. Middelweerd, J. S. Mollee, C. N. van der Wal, J. Brug, and S. J. te Velde, “Apps to promote physical activity among adults: a review and content analysis,” *Int J Behav Nutr Phys Act*, vol. 11, no. 1, p. 97, 2014.
[15] C. Abraham and S. Michie, “A taxonomy of behavior change techniques used in interventions.” *Health psychology*, vol. 27, no. 3, p. 379, 2008.
[16] PewResearchCenter, October 2015.
[17] GfK, December 2015.
[18] F. Griffiths, A. Lindenmeyer, J. Powell, P. Lowe, and M. Thorogood, “Why are health care interventions delivered over the internet? a systematic review of the published literature,” *Journal of medical Internet research*, vol. 8, no. 2, 2006.
[19] A. A. Hagberg, D. A. Schult, and P. J. Swart, “Exploring network structure, dynamics, and function using NetworkX,” in *Proceedings of the 7th Python in Science Conference (SciPy2008)*, Pasadena, CA USA, Aug. 2008, pp. 11–15.
[20] R. Tarjan, “Depth-first search and linear graph algorithms,” *SIAM journal on computing*, vol. 1, no. 2, pp. 146–160, 1972.
[21] E. Nuutila and E. Soisalon-Soininen, “On finding the strongly connected components in a directed graph,” *Information Processing Letters*, vol. 49, no. 1, pp. 9–14, 1994.
[22] P. D. Bliese and R. E. Ployhart, “Growth modeling using random coefficient models: Model building, testing, and illustrations,” *Organizational Research Methods*, vol. 5, no. 4, pp. 362–387, 2002.
[23] J. Pinheiro, D. Bates, S. DebRoy, D. Sarkar, and R Core Team, *nlme: Linear and Nonlinear Mixed Effects Models*, 2016, r package version 3.1-128. [Online]. Available: http://CRAN.R-project.org/package=nlme
[24] G. Schoenewolf, “Emotional contagion: Behavioral induction in individuals and groups.” *Modern Psychoanalysis*, 1990.
[25] S. E. Cohen and S. Syme, *Social support and health.* Academic Press, 1985.
[26] L. Festinger, “A theory of social comparison processes,” *Human relations*, vol. 7, no. 2, pp. 117–140, 1954.
[27] T. Bosse, R. Duell, Z. A. Memon, J. Treur, and C. N. Van Der Wal, “A multi-agent model for emotion contagion spirals integrated within a supporting ambient agent model,” in *International Conference on Principles and Practice of Multi-Agent Systems*. Springer, 2009, pp. 48–67.