Proceedings of the 56th Hawaii International Conference on System Sciences | 2023

Investigating Churn in Physical Activity Challenges: Evidence from a U.S. Online Social Network

Yi Zhu
University of Minnesota
zhu00331@umn.edu

Ankur Mani
University of Minnesota
amani@umn.edu

Abstract

Physical activities have been found to be positively contagious, as active exercisers tend to motivate their friends to do more exercise. However, it is not clearly understood if inactive exercising behaviors are also socially contagious. As insufficient physical activity is a huge threat to people's health, understanding the potential negative contagion in physical activities is crucial. We approach this problem by studying the effect of individuals' churn of online physical activity challenges relying on the physical activity and a large social network data from a renowned U.S. fitness platform. The underexplored online physical activity challenges provide a natural setup to measure churn and opportunities to study the contagion heterogeneities. Consistent with previous findings, we confirm that physical activity churn is socially contagious. Interestingly, unlike the inside-out positive contagion, our analyses reveal that the contagion of churn happens outside-in on the social network. Implications of such findings are discussed.

Keywords: Social Contagion, Physical Activity, Virtual Challenge, Network Analysis, Churn Analysis

1. Introduction

Physical activities are contagious, as friends' exercising behavior influences individuals' tendency to do exercise (Maturo & Cunningham., 2013; Proestakis et al., 2018). Most existing works conclude that positive contagion usually happens inside-out, as active and influential exercisers in the center of the social network tend to stimulate the physical activities of their friends (or followers) who are located off the network center (Aral & Nicolaides, 2017; Dijk & Treur, 2018). However, negative contagion in the physical activity social network is understudied, and it is not well understood if the findings in the positive contagion hold true for the negative contagion.

Negative contagion in physical activities pertains to the effect of individuals' dormant exercising behavior on their connections, which is the focus of this study. We approach understanding such negative contagion through the evidence of exercise churn, which is important for several reasons. First, churn rate is very high among both new and experienced exercisers (Linke et al., 2011; Lee and Owens, 1986), and insufficient physical activity (partly due to churning midway) is the fourth leading risk factor for mortality around the world (World Health Organization, 2014). Understanding the mechanisms of churn is essential for bringing individuals closer to good health. Second, one may imprudently assume that the effect of churn is similar to the positive contagion and evolves inside-out. However, no work provides explicit evidence to verify such assumption, potentially due to the limitation of an ideal research design, available data, and appropriate measurement of physical activity churn. We employ a special research context, novel measurements, and advanced empirical methodologies to approach the problem.

To achieve our research goal, we utilize the design of virtual physical activity challenges on a renowned fitness platform based in the U.S. The platform introduces apps on smartphones or wearable devices to automatically track and record user data generated during physical activities through the embedded GPS-based sensor. The virtual challenges on the focal platform motivate users to complete short-term exercise goals within a certain period (e.g., walking 10,000 steps a day) to assist adherence to regular exercises. Such designs provide a natural setup for studying users' physical activity churn, as the churn can be easily identified by evaluating users' activity performance during a challenge period against the challenge's goal. We focus on a challenge that motivates users to complete running 100 kilometers every month. Our study is across four months, involving around 20 thousand U.S. users and approximately 800 thousand records of daily running activities. We measure churn by users' dormant days and churn day (more details below) in each month and estimate the contagion based on the characteristics of users' social network connections on the platform. To account for the endogeneity of the estimated contagion effect, we employ an instrumental variable (IV)

URI: https://hdl.handle.net/10125/102906
978-0-9981331-6-4

framework (Angrist & Pischke, 2008), using the exogenous variation of weather across different geographic locations as instruments (more details below).

Consistent with previous findings, we find that physical activity churn is socially contagious. Explicitly, being easy to quit regular physical activity is largely attributed to individuals' friends' high churn rate. Interestingly, unlike the inside-out positive contagion, our analyses reveal that the contagion of churn happens outside-in. I.e., the negative contagion in physical activities spreads from the sub-central or peripheral individuals to central individuals on the social network.

The key contribution of our work is to shed light on the practical and theoretical unknowns of negative contagion in physical activities. We disentangle this contagion using the natural setup of online physical activity challenges that have received little attention in previous work and empirical methodologies supported by fine-grained physical activity and social network data. Our work contributes to several streams of literature. First, we add to the study of physical activity contagion, investigating the phenomenon from the perspective of churn. We confirm that the positive contagion found in existing literature also exists in the churn (negative) contagion but diffuses in a different fashion. Second, we contribute to the understanding of physical activity adherence by presenting empirical evidence that physical activity churn is strongly infectious, inferring that adherence is mediated by social influence.

The rest of the paper is organized as follows. In the next section, we review relevant literature, after which we describe the data we use in the analyses and identify the measurement of variables. Next, we present our estimating methodologies and main results. We conclude with a discussion of the work's implications and avenues for future research.

2. Literature Review

2.1. Physical Activity Contagion

Social contagion, also known as social influence, is "the change in an individual's thoughts, feelings, attitudes, or behaviors that results from interaction with another individual or a group." (Rashotte, 2007) Physical activity is typical human behavior, and friends' physical activities may influence individuals' physical activities. In the current study context, if an individual participates in an online challenge and sees her friend's fast progress in completing the challenge goal, she may be motivated to do more exercise or pick up what she has dropped for days. In contrast, when the individual's friends make slow progress, she may subconsciously find it okay to remain sedentary. The topic of social influence in physical activity is not new, but related studies under the online context and based on large social networks only emerged in recent years, thanks to the advancement of social media platforms (Aral & Nicolaides, 2017; Dijk & Treur, 2018).

Numerous works have demonstrated strong social contagion in physical activities. In a review of 35 research papers related to social influence in physical activities, Maturo and Cunningham (2013) find substantial evidence in 30 papers (85%) that children's physical activity is associated with the physical activity of their friends, suggesting that children may be guided by their friends' behavior. In middle-aged adults, social influence enhances the opportunity to bond, feel more confident about participating in exercise and facilitate adherence (White et al., 2005; Dijk & Treur, 2018). However, social support may not facilitate physical activity in older adults (Brassington et al., 2002; De La Cámara et al., 2020).

Most of these studies either employ field experiments that rely on limited collected samples or only focus on local social effects without considering the effects of large networks and settings with online-offline interactions. Additionally, to the best of our knowledge, no research has touched on the social influence in completing the online physical activity challenges, from which we can infer the contagion effect in physical activity churn. Disentangling such effects from individuals' completion of virtual challenges is the focus of this paper.

2.2. Physical Activity Adherence

Adherence to regular physical activities has been difficult for many individuals. Many studies have tried to understand the reasons for quitting physical activities. Glaros and Janelle (2001) report increased adherence in those who varied their activity type compared to those who did not vary. The role of activity type in exercise adherence is further mediated by enjoyment, as when individuals enjoy what they are doing, they tend to continue doing it (Dacey et al., 2003). Other studies point the factors affecting physical activity adherence to activity settings, i.e., the location where exercise takes place. Morey et al. (2003) suggest that home-based exercise programs or exercising in the near-home environment enhance adherence. Hallam and Petosa (2004) find that the workplace may facilitate adherence due to the convenience and social support available.

Moreover, previous works have also documented the enhanced adherence to physical activity due to increased self-efficacy. Self-efficacy refers to an individual having confidence to participate in physical activity (Mcauley & Blissmer, 2000). Studies have demonstrated that building self-efficacy using social modeling and persuasion can enhance adherence to long-term physical activity (Mcauley et al., 1994). Another study suggests that when participating in an activity without a "leader," individuals may have less self-efficacy because they are expected to do the activity on their own (Eys & Brawley, 2018). In contrast, in situations that a "leader" presents, self-efficacy was higher because individuals knew they had someone to guide them through the physical activity (Rhodes et al., 2001). These findings further open the study of the impact of social support on physical activity adherence.

Literature related to self-efficacy and social support in affecting physical activity adherence has investigated the role friends play in individuals' physical activity. However, the social influence under focus pertained to small-scale individual connections, and the study settings were all offline. In the current study context, we employ a large online social network containing over 20,000 nodes to study social contagion in physical activity churn. We approach the investigation from the perspective of network science, extending the insights from local network structures (consisting of only a few nodes) to global network structures (consisting of abundant nodes).

3. Background and Data

To analyze the contagion in physical activity churn, we rely on publicly available physical activity and social network data from a renowned U.S. online exercise platform. To obtain the physical activity data, the platform allows its users to track their physical activities by using its app on smartphones or wearable devices. The embedded GPS-based sensor automatically records data generated during physical activities, including activity measures (e.g., distance, duration, heart rate), date, and time of the day. To obtain the social network data, the platform allows users to connect with one another by following and being followed by other users. Thus, constructing a social network. As long as the user keeps her activity profile public, these social features keep the user posted on her friends' activity updates immediately after each activity is completed.

The focal platform holds many physical activity challenges, aiming to motivate regular exercise by encouraging users to achieve a specific goal within a certain period (e.g., week, month). Such designs provide a natural setup for studying users' churning behavior, as churn can be easily identified by evaluating users' activity performance during a challenge period against the challenge's goal. This work focuses on users' activities in the challenge that motivates participants to complete a 100-kilometer running distance every month. Our study sample consists of 21,354 US users who participated in the challenge from February 2021 to May 2021 and ran at least one day in each month. We download approximately 800 thousand publicly available records of their daily running distances. Because the challenge period is a month, and we are interested in studying users' churning behavior in completing the monthly goals, we aggregate the daily distance data by calculating the monthly mean to create a monthly panel.

Based on the 21,354 US users in our study sample, we collect the social network data. In particular, we download these users' publicly available information of the people they follow on the platform to construct a social network within these users. The data download was done at the beginning of February 2021 to reflect the network at the starting point of the study period. On the network, each node represents a user, and each edge represents a following/being followed relation. We remove users who are not included in our study sample or isolated in the network. Our eventual network consists 19,159 nodes and 97,660 edges. The network's largest component contains 18,115 nodes, and the second largest component contains five nodes. With the largest component close to the size of the whole network and the second largest component much smaller than the largest component, our network is well-connected (Newman et al., 2001). Our final panel data consists of 76,636 user-month observations (19,159 ×4), which is our unit of analysis.

3.1. Dependent Variables

We consider two outcomes to measure a user's churning behavior. First, to study the contagion effect, we measure the tendency to churn by a user's number of dormant days in each month, which we define as the

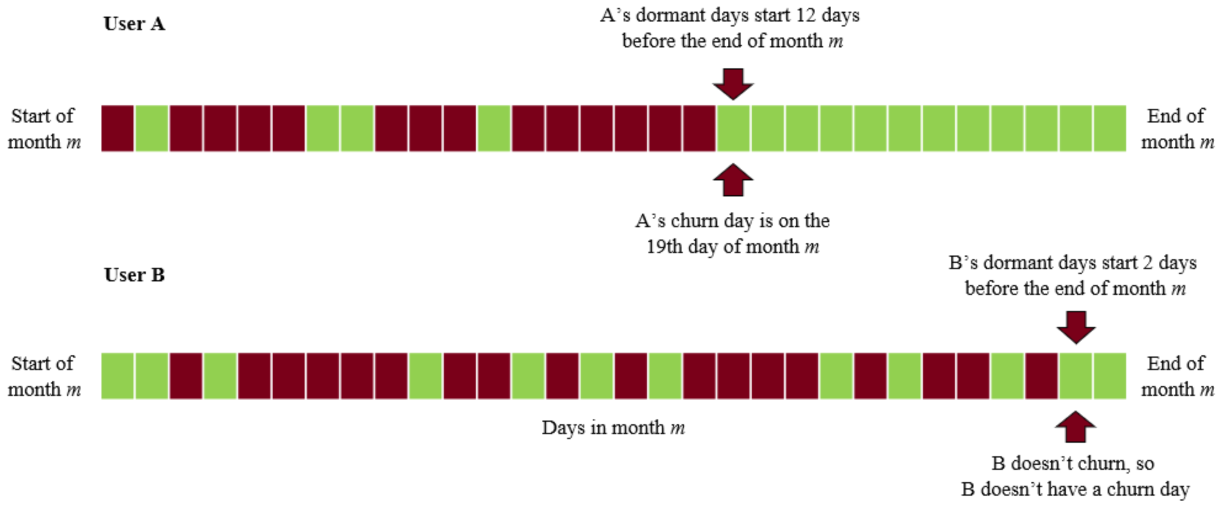

Figure 1. Dormant days and churn day

number of consecutive inactive days (recorded zero running distance) until the end of the month. Figure 1 presents an example of physical activities for user A and user B in month m. User A and user B start to have their consecutive inactive days 12 and 2 days before the end of month m. Thus, their number of dormant days is 12 and 2, respectively. Second, to study the spreading fashion of the contagion, we first defined churn as having at least five dormant days in a month. The cutoff is borrowed from previous literature that typically defines a lapse in a new behavior as having a week or less of days returning to previous unhealthy behavior (CDC, 2012; Mohseni et al., 2022). We also used seven days as the cutoff for robustness check. We then define a user's churn day as the ordinal value of her first dormant day relative to the beginning of the month. If a user doesn't churn (i.e., number of dormant days less than five), then we don't consider this user in the analysis of churn day. In Figure 1, user A's churn day is on the 19th day of month m. Because B's number of dormant days is two, B doesn't churn, and we don't consider B in the churn day analysis.

3.2. Independent Variables

Social contagion comes from users' friends (or peers). We define a focal user's friends as the users who the focal user follows. We then investigate the social contagion on users' number of dormant days in each month based on three variables of interest. First, we compute the average monthly number of dormant days among the focal user's friends. Second, we compute the average monthly distance gap among the focal user's friends. The distance gap is defined as the eventual running distance shortage compared with the 100-kilometer goal in each month. Third, we compute the monthly fraction of the focal user's friends who churn. All variables are based on users' friends' churning behavior.

3.3. Control Variables

To control for potential confounding factors that influence the estimation of the contagion effect, we collect each user's weather and demographic information. Our weather data includes daily temperature, precipitation, and wind speed, which we aggregate at the user-month level. The demographic data contains age, weight, and gender. Additionally, the contagion effect estimation may also deviate from the truth if we don't consider the position of each user in the social network (Tang et al., 2009; Liao et al., 2018). We use three network embeddedness to represent different measurements of a user's position in the network: the degree centrality, the eigenvector centrality, and the betweenness centrality. To rule out potential endogeneity, we compute the average value of these control variables across all friends of each user to be included in our instrumental variable framework (more details later). We present the summary statistics of our panel data in Table 1.

4. Econometric Methodologies

4.1. Naive Correlation

To estimate the contagion, we first employ an ordinary least squares model and fit the equation:

$$D_{it} = \delta \cdot D_{it}^{p} + W_{it} + Z_i + N_i + \eta_i + v_t + \varepsilon_{it} \quad (1)$$

Table 1. Summary statistics of panel data

Variable	Obs	Mean	SD.	Min	Max
Panel A: Dependent and independent variables					
Dormant days (dependent variable)	76,636	1.877	3.883	0	30
Mean friends dormant days	76,636	1.610	2.452	0	30
Mean friends distance gap (km)	76,636	11.752	17.075	0	99.664
Fraction of friends who churn (%)	76,636	6.796	16.335	0	100
Panel B: Users' control variables					
Mean monthly temperature (°F)	76,636	51.768	13.331	-9.600	83.719
Mean monthly precipitation (inch)	76,636	0.076	0.065	0	1.589
Mean monthly wind speed (mph)	76,636	14.460	2.666	1.555	28.243
Age (years)	76,636	37.001	12.043	20	70
Weight (kg)	76,636	69.136	15.669	50	100
Gender (male)	76,636	0.725	0.446	0	1
Degree centrality	76,636	0.303	0.486	0.047	18.771
Eigenvector centrality	76,636	1.887	6.663	8.14E-22	190.210
Betweenness centrality	76,636	0.182	1.041	0	76.593
Panel C: Users' friends' control variables					
Mean friends monthly temperature (°F)	76,636	51.697	12.496	-9.600	83.719
Mean friends monthly precipitation (inch)	76,636	0.075	0.056	0	0.811
Mean friends monthly wind speed (mph)	76,636	14.463	2.229	1.555	28.239
Mean friends age (years)	76,636	37.054	9.692	20	70
Mean friends weight (kg)	76,636	69.049	10.719	50	100
Mean friends proportion of male	76,636	0.742	0.308	0	1
Mean friends degree centrality	76,636	0.001	0.001	4.70E-05	0.019
Mean friends eigenvector centrality	76,636	0.006	0.013	8.14E-25	0.185
Mean friends betweenness centrality	76,636	0.001	0.004	0	0.077

where focal user i's number of dormant days in month t (D_{it}) is regressed on i's peers' mean number of dormant days in month t (D_{it}^{p}), mean weather experienced by i in month t (W_{it}), i's demographic characteristics (Z_i), i's network centralities (N_i), state fixed effect (η_i), time fixed effect (v_t), and an error term ε_{it}. The coefficient δ measures the association between the focal user's tendency to churn and her friends' tendency to churn. This estimating equation suffers from many well-known empirical challenges in identifying causal peer effects, including i) homophily, i.e., the tendency for users to choose similar friends; ii) omitted factors, i.e., the tendency for connected users to be exposed to the same external stimuli; iii) simultaneity, i.e., the tendency for users to co-affect each other; and iv) measurement errors, i.e., the tendency for the platform to inaccurately record users' self-report data or problematic data due to app glitch.

4.2. IV Framework

To avoid such empirical challenges, we employ the IV framework developed by former econometricians, which untangles endogeneity using variations created by instrumental variables as a shock to the endogenous explanatory variable to estimate its causal effect on the outcome (Angrist & Krueger, 2001). It solves the endogeneity problems caused by omitted variables, simultaneity, and measurement errors all at once. The instrumental variables are not directly correlated with the outcome variable but are correlated with the endogenous explanatory variable.

In the current research context, the chosen instruments should be an exogenous factor that directly affects every user's friends' physical activities but is independent from users' physical activity. For this purpose, the weather of each user's friends' self-identified city is an ideal instrument (Coviello et al., 2014). Our study sample consists of 19,159 users from 5,265 different cities, and all the recorded running activities are outdoor activities. Users experience different weather, which interferes with their running activity but does not perturb their followers (focal users) who are located in other cities. We can then adopt weather as the instrument to estimate the causal contagion effect in exercise churn by fitting a two-stage least squares specification in the form:

$$\text{Stage 1: } \widehat{D}_{it}^{p} = \lambda' W_{it}^{p} + \gamma' Z_{i}^{p} + \theta' N_{i}^{p} + \eta'_{i} + v'_{t} + \varepsilon'_{it} \quad (2)$$

$$\text{Stage 2: } D_{it} = \beta \cdot \widehat{D}_{it}^{p} + \lambda W_{it} + \gamma Z_{i} + \theta N_{i} + \eta_{i} + v_{t} + \varepsilon_{it} \quad (3)$$

where in Stage 1, the average dormant days of focal user *i*'s peers at month t ($\widehat{D}_{it}^{p}$) is regressed on the mean weather experienced by peers (W_{it}^{p}), peers' mean characteristics (Z_{i}^{p}), peers' mean network centralities (N_{i}^{p}), state fixed effect (η'_{i}), time fixed effect (v'_{t}), and an error term (ε'_{it}). In Stage 2, the average dormant days of focal user i at month t (D_{it}) is regressed on *i*'s peers' average dormant days at month t that estimated in Stage 1 ($\widehat{D}_{it}^{p}$), controlling for the weather experienced by i (W_{it}), *i*'s characteristics (Z_i), *i*'s network centralities (N_i), state fixed effect (η_i), time fixed effect (v_t), and an error term ε_{it}.

The predicted values estimated in Stage 1 capture only those changes in a user's friends' churning behavior caused by changes in weather that the focal user does not experience. In Stage 2, only the predicted user's friends' churning behavior is used to estimate the social contagion on physical activity churn. Thus, the IV framework enables causal inference by excluding the focal user's simultaneous effect on her friends and variation created by observable and unobservable confounding factors. Coefficient β measures the contagion effect of the tendency to churn with minimum endogeneity and is our estimate of interest.

4.3. Contagion Fashion

Most existing literature finds that the contagion of active exercises (positive contagion) happens inside-out. However, whether the contagion of exercise churn (negative contagion) follows the same fashion hasn't been explicitly documented. These two types of contagion potentially operate differently, as the positions of active and dormant exercisers in the social network tend to differ. Understanding the contagion fashion of churn is important for the platform to set up interdependent intervention plans to avoid churn with awareness of the churn origin. To investigate such fashion, we consider the following estimating equation:

.

$$C_{it} = \alpha \cdot N_i + W_{it} + Z_i + \eta_i + v_t + \varepsilon_{it} \quad (4)$$

where focal user *i*'s churn day in month t (C_{it}) is regressed on *i*'s network centralities (N_i), mean weather experienced by i in month t (W_{it}), *i*'s demographic characteristics (Z_i), state fixed effect (η_i), time fixed effect (v_t), and an error term ε_{it}. The coefficient α measures the association between the focal user's churn day and her centrality on the social network.

4.4. Robustness Check

We run standard statistical tests for IV framework to confirm the validity of the instruments we choose and the robustness of our results. First, we run the Weak IV test to estimate whether the chosen instruments are highly correlated with the endogenous explanatory variable (i.e., friends' churning behavior). This can be achieved using the F-statistics in the Stage 1 model estimation. Second, we run the Wu-Hausman test to check if the estimation results with and without using instruments are the same. If they are the same, it implies we don't need the instruments to estimate the effect. Finally, we run the Sargan test to evaluate the equivalence of assessing the effect between using all instruments and a subset of the instruments. If they are equivalent, our chosen instruments are all valid.

5. Results

5.1. Naive Correlation Results

We employ an OLS model as introduced in Equation (1) to estimate the impact of our three explanatory variables of interest, including mean friends dormant days, mean friends distance gap, and the fraction of friends who churn. We find strong evidence of contagion in the tendency of running activity churn, as the regressions with different specifications of control variables and fixed effects show a significant positive relationship ($p < 0.01$) between a focal user's number of dormant days and each of the three variables of interest. Nevertheless, these estimates are subject to the well-known endogeneity biases created by homophily, omitted factors, simultaneity, and measurement error, as introduced in the econometric methodology section. Hence, we focus our analysis based on IV framework, which takes advantage of the natural experiment created by exogenous variation of weather in different cities where users are located

Table 2. Effects of friends' tendency to churn on users' tendency to churn

Variables	Dormant Days								
	Model 1	Model 2	Model 3	Model 4	Model 5	Model 6	Model 7	Model 8	Model 9
Mean friends dorm days	0.386***	0.421***	0.461***	-	-	-	-	-	-
	(0.115)	(0.121)	(0.123)	-	-	-	-	-	-
Mean friends dist gap (km)	-	-	-	0.037***	0.039***	0.040***	-	-	-
	-	-	-	(0.007)	(0.007)	(0.007)	-	-	-
Pct friends who churn	-	-	-	-	-	-	0.058***	0.059***	0.066***
	-	-	-	-	-	-	(0.018)	(0.018)	(0.018)
Mean month temp (°F)	-0.001	-0.000	-0.001	0.002	0.002	0.002	-0.000	-0.000	-0.000
	(0.003)	(0.003)	(0.003)	(0.003)	(0.003)	(0.003)	(0.003)	(0.003)	(0.003)
Mean month prcp (inch)	0.039	0.037	0.040	-0.104	-0.116	-0.119	0.064	0.060	0.067
	(0.292)	(0.293)	(0.295)	(0.288)	(0.288)	(0.289)	(0.295)	(0.296)	(0.299)
Mean month wind (mph)	0.006	0.006	0.006	0.008	0.008	0.008	0.006	0.005	0.006
	(0.009)	(0.009)	(0.009)	(0.009)	(0.009)	(0.009)	(0.009)	(0.009)	(0.009)
Age (years)	-0.012***	-0.012***	-0.012***	-0.013***	-0.014***	-0.014***	-0.011***	-0.011***	-0.012***
	(0.001)	(0.002)	(0.002)	(0.002)	(0.002)	(0.002)	(0.002)	(0.002)	(0.002)
Weight (kg)	0.011***	0.011***	0.011***	0.010***	0.010***	0.010***	0.010***	0.011***	0.010***
	(0.001)	(0.002)	(0.002)	(0.001)	(0.001)	(0.001)	(0.002)	(0.002)	(0.002)
Gender (male)	-0.221***	-0.234***	-0.224***	-0.165***	-0.176***	-0.168***	-0.210***	-0.225***	-0.213***
	(0.049)	(0.050)	(0.050)	(0.051)	(0.051)	(0.051)	(0.051)	(0.051)	(0.052)
Degree centrality	-26.546***	-	-	-25.327***	-	-	-26.145***	-	-
	(3.766)	-	-	(3.622)	-	-	(3.780)	-	-
Eigenvector centrality	-	-7.426***	-	-	-7.405***	-	-	-7.887***	-
	-	(2.678)	-	-	(2.649)	-	-	(2.688)	-
Betweenness centrality	-	-	-9.038**	-	-	-8.890**	-	-	-9.136**
	-	-	(3.564)	-	-	(3.528)	-	-	(3.579)
Month FE	Yes								
State FE	Yes								
Observations	76,636	76,636	76,636	76,636	76,636	76,636	76,636	76,636	76,636

Notes: The dependent variables of all models are users' dormant days. All models use linear regression based on IV framework for estimation. Robust standard errors in parentheses and clustered at the user level. *** p<0.01, ** p<0.05, * p<0.1.

5.2. IV Framework Results

We estimate the social contagion in the tendency of running activity churn using the IV framework introduced in Equations (2) and (3). The results confirm a strong contagion effect: the average number of dormant days among friends, the average distance gap among friends, and the fraction of friends who churn during the month all exert a significant positive effect on users' dormant days. I.e., when any of these three variables become larger, the focal user is more likely to churn in the running challenge. We show the results in Table 2.

5.3. Contagion Fashion Results

As shown in Table 2, we find a strong negative relationship between the number of dormant days and individuals' network centralities. I.e., the more central a user is on the social network, the less likely she will churn the exercise challenge. This insight suggests that the contagion of churn is initiated by less central users who are more tend to churn. To verify if this speculation is true and understand how the contagion of physical activity churn spreads in the social network, we evaluate the correlation between the extent to which a focal user is connected with other users in the social network and the timing of the focal user's churn.

Table 3. Association between users' churn day and their network centralities

Variables	Churn Day Model 1	Model 2	Model 3	Model 4	Model 5	Model 6	Model 7	Model 8	Model 9
Degree	20.890**	-	-	21.094**	-	-	17.398*	-	-
centrality	(9.122)	-	-	(9.107)	-	-	(9.155)	-	-
Eigenvector	-	6.336	-	-	6.485	-	-	0.473	-
centrality	-	(8.981)	-	-	(8.960)	-	-	(9.219)	-
Betweenness	-	-	10.308**	-	-	10.553**	-	-	8.500**
centrality	-	-	(4.154)	-	-	(4.166)	-	-	(4.053)
Mean month	-	-	-	0.025*	0.025*	0.025*	0.022	0.022	0.023
Temp (°F)	-	-	-	(0.015)	(0.015)	(0.015)	(0.015)	(0.015)	(0.015)
Mean month	-	-	-	0.146	0.131	0.136	0.140	0.119	0.132
prcp (inch)	-	-	-	(1.311)	(1.310)	(1.311)	(1.308)	(1.307)	(1.308)
Mean month	-	-	-	0.018	0.018	0.017	0.017	0.017	0.017
wind (mph)	-	-	-	(0.032)	(0.032)	(0.032)	(0.032)	(0.032)	(0.032)
Age	-	-	-	-	-	-	0.030***	0.031***	0.030***
(years)	-	-	-	-	-	-	(0.006)	(0.006)	(0.006)
Weight	-	-	-	-	-	-	0.002	0.002	0.002
(kg)	-	-	-	-	-	-	(0.005)	(0.005)	(0.005)
Gender	-	-	-	-	-	-	0.133	0.140	0.137
(male)	-	-	-	-	-	-	(0.169)	(0.169)	(0.169)
Month FE	Yes								
State FE	Yes								
Observations	9,321	9,321	9,321	9,321	9,321	9,321	9,321	9,321	9,321

Notes: The dependent variables of all models are users' churn day. All models use linear regression based on IV framework for estimation. Robust standard errors in parentheses and clustered at the user level. *** p<0.01, ** p<0.05, * p<0.1.

Following Equation (4), we approach this evaluation by modeling a focal user's churn day (introduced in the Dependent Variables section) as a function of her different types of centralities on the social network. Users who don't churn (number of dormant days less than five) are excluded from this analysis. The results are reported in Table 3. We find that a user's churn day is positively related to her degree centrality and betweenness centrality, but no significant relationship is found between churn day and eigenvector centrality. The results suggest that the greater the extent to which a user is directly connected with other users or serves as a bridge between other users, the later she will churn in the challenge. In other words, churn happened outside-in, i.e., i.e., spreading from the sub-central or peripheral individuals to the more central individuals on the social network.

We further demonstrate this correlation with four graphs of the largest component in the social network, where each graph represents a different phase in a month. Figure 2. depicts the evolvement of users' running activity churn in February 2021. The graphs show that in the first five days of the month, only a small number of users churned, most of whom are scattered around the network's periphery with low degrees. Moving towards the end of the month, more users churned, and the churned users occupied more space in the network center with high degrees. We obtain similar results when we use seven dormant days as the cutoff to define churn.

5.4. Robustness Check Results

Our chosen instruments and final results are robust to multiple standard tests. In unreported studies, we find a significant relationship between egos' monthly weather conditions and egos' monthly dormant days, suggesting strong predicting power of weather on dormant days. Additionally, the result of Stage 1 in our IV framework shows strong relationships between peers' average weather conditions and the three explanatory variables, indicating the significant effect of weather on peers' behaviors. Further, the Weak IV test confirmes a high correlation between the chosen instruments (weather) and the endogenous explanatory variable (friends' churning behavior), with F-statistics in Stage 1 excessively larger than the critical threshold of 10 as suggested by (Bound et al., 1995; Staiger & Stock, 1997), implying strong instruments (Stage 1 F-statistics range from 25.33 to 354.38, $p < 0.01$, N = 76,636). The coefficients of weather variables are statistically insignificant in Stage 2 (Table 2) because the effect of weather has been absorbed by egos' peers' behavior, which is represented by the three endogenous explanatory variables.

The Wu-Hausman test rejects the null hypothesis that estimates yielded from IV framework are equivalent to OLS, implying the contagion in running activity churn is endogenous, and using IV framework helps rule out the endogenous effect (F-statistics range from 7.52 to 23.82, $p < 0.01$, N = 76,636). Finally, the Sargan test fails to reject the null hypothesis that the effect estimated using all instruments is equivalent to using a portion of the instruments (p values range from 0.08 to 0.87, N = 76,636), implying the instruments we use are all valid.

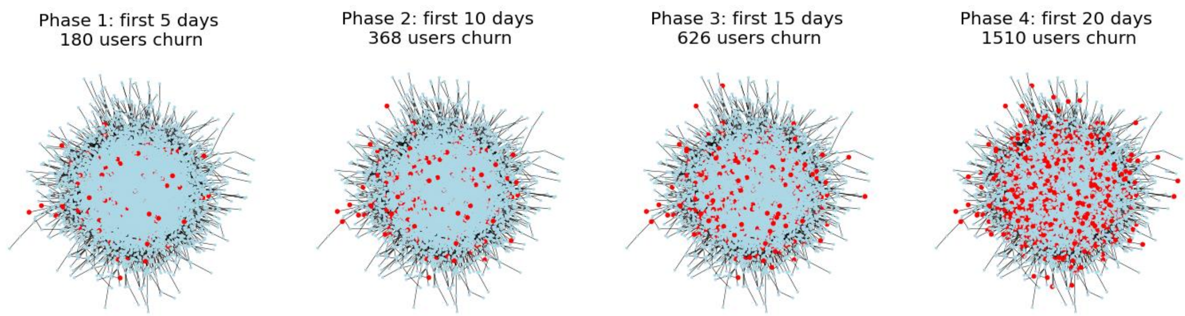

Figure 2. The evolvement of users' running activity churns in February 2021. Notes: Each graph shows the largest component of the network under study. The more central a node is positioned in the network, the more other nodes it is connected to (higher degree). Each red node represents a churning user, each blue node represents a not-churning user, and the blue lines are edges that connect these users.

6. Discussion and Future Work

In this work, we leverage a large online social network and a series of online running distance challenges to estimate the social contagion in physical activity churn. We find a strong contagion effect in physical activity churn and show that the contagion happens outside-in. These findings confirm the social contagion in physical activities from the flip side and uncover the differences in spreading fashions between the contagion of active exercises (positive contagion) and exercise churn (negative contagion).

Our analysis results imply several implications for individuals and online fitness platforms that host virtual challenges. For the former, our findings suggest that following active and popular exercisers and being posted on their updates is beneficial for individuals' adherence to regular exercises. For the latter, our analysis results indicate that interdependent intervention plans (plans accounting for the structure of social networks) may reduce churn more effectively than independent intervention plans (plans only focus on individuals' characteristics and behaviors).

In addition, the insight that negative contagion happens outside-in suggests that to prevent a physical activity social network from unraveling (i.e., having many individuals churn regular exercise), platforms should start the intervention from off-central individuals on the network to stop the churn contagion from its origin. The intervention plans could be sending notifications to remind individuals of their regular physical activities or help them connect with more active and central exercisers to receive more positive contagion.

In future work, our research could benefit from several further analyses. First, our current analyses are based on monthly churning behaviors, which cannot capture the dynamics of the tendency to churn during the challenge process. Since individuals' churn decisions may change during the challenge (i.e., they may decide to churn now but resume later), it would be interesting to bring the analyses to the daily level and investigate how friends' short-term behavior influences individuals' churn decisions. Second, churning one challenge does not necessarily mean the user churns all physical activities or stops using the fitness platform. It is of high interest and value to find out if the users who churned the running challenge also paused other physical activities or stopped using the platform. With such evidence, we can further confirm the robustness of our results. Finally, causal inference should base on the premise that the cause happens before the effect, and the interference between treatment and control group is minimum. Therefore, further falsification tests are needed to rule out the possibility that friends' future physical activity influence the focal user and unconnected users influence each other.

Our work also opens several directions for extended research. First, virtual challenge of different activities (e.g., exercising, reading, dieting) is an emerging marketing strategy being explored by an increasing number of digital platforms. Compared with traditional one-way marketing that attracts traffic by offering customers benefits, this kind of strategy draws and retains customers by strengthening their self-motivation. What value virtual challenges can bring to business and what type of business can benefit from virtual challenges are understudied, but the importance cannot be overstated. Second, now that we find evidence of social contagion in physical activity churn, an interesting question is how individuals' health benefits from such contagion, as regular physical activity is beneficial to people's health. If the social contagion discussed in this research is found to have a spillover effect in shaping individuals' health, it may imply a network-based way of thinking for reducing obesity- and sedentariness-induced diseases.

7. References

Angrist, J. D., & Krueger, A. B. (2001). Instrumental Variables and the Search for Identification: From Supply and Demand to Natural Experiments. Journal of Economic Perspectives, 15(4), 69–85.

Aral, S., & Nicolaides, C. (2017). Exercise contagion in a global social network. Nature Communications, 8(1), 14753.

Brassington, G. S., Atienza, A. A., Perczek, R. E., DiLorenzo, T. M., & King, A. C. (2002). Intervention-related cognitive versus social mediators of exercise adherence in the elderly. American Journal of Preventive Medicine, 23(2, Supplement 1), 80–86.

CDC (2012). Preventing Relapse.
https://www.cdc.gov/diabetes/prevention/pdf/postcurriculum_session11.pdf

Coviello, L., Sohn, Y., Kramer, A. D. I., Marlow, C., Franceschetti, M., Christakis, N. A., & Fowler, J. H. (2014). Detecting Emotional Contagion in Massive Social Networks. PLOS ONE, 9(3), e90315.

De La Cámara, M. Á., Jiménez-Fuente, A., & Pardos, A. I. (2020). Falls in older adults: The new pandemic in the post COVID-19 era? Medical Hypotheses, 145, 110321.

Dijk, M., & Treur, J. (2018). Physical Activity Contagion and Homophily in an Adaptive Social Network Model. In N. T. Nguyen, E. Pimenidis, Z. Khan, & B. Trawiński (Eds.), Computational Collective Intelligence (pp. 87–98). Springer International Publishing.

Eys, M. A., & Brawley, L. R. (2018). Reflections on cohesion research with sport and exercise groups. Social and Personality Psychology Compass, 12(4), e12379.

Glaros, N. M., & Janelle, C. M. (2001). Varying the Mode of Cardiovascular Exercise to Increase Adherence. Journal of Sport Behavior, 24(1), 42.

Hallam, J. S., & Petosa, R. (2004). The Long-Term Impact of a Four-Session Work-Site Intervention on Selected Social Cognitive Theory Variables Linked to Adult Exercise Adherence. Health Education & Behavior, 31(1), 88–100.

Lee, C., & Owen, N. (1986). Exercise persistence: Contributions of psychology to the promotion of regular physical activity. Australian Psychologist, 21(3), 427–466.

Linke, S. E., Gallo, L. C., & Norman, G. J. (2011). Attrition and Adherence Rates of Sustained vs. Intermittent Exercise Interventions. Annals of Behavioral Medicine, 42(2), 197–209.

Maturo, C. C., & Cunningham, S. A. (2013). Influence of Friends on Children's Physical Activity: A Review. American Journal of Public Health, 103(7), e23–e38.

Mcauley, E., Courneya, K. S., Rudolph, D. L., & Lox, C. L. (1994). Enhancing Exercise Adherence in Middle-Aged Males and Females. Preventive Medicine, 23(4), 498–506.

Mcauley, E., & Blissmer, B. (2000). Self-Efficacy Determinants and Consequences of Physical Activity. Exercise and Sport Sciences Reviews, 28(2), 85–88.

Mohseni, F., Rahimi, K., Niroumand Sarvandani, M., Jamali, Z., Seyedhosseini Tamijani, S. M., & Rafaiee, R. (2022). Lapse and Relapse Rates in Narcotics Anonymous versus Methadone Maintenance Treatment: A 12-Month Prospective Study. Iranian Journal of Psychiatry, 17(1), 1–13.

Morey, M. C., Dubbert, P. M., Doyle, M. E., MacAller, H., Crowley, G. M., Kuchibhatla, M., Schenkman, M., & Horner, R. D. (2003). From Supervised to Unsupervised Exercise: Factors Associated With Exercise Adherence. Journal of Aging & Physical Activity, 11(3), 351–368.

Newman, M. E. J., Strogatz, S. H., & Watts, D. J. (2001). Random graphs with arbitrary degree distributions and their applications. Physical Review E, 64(2), 026118.

Proestakis, A., di Sorrentino, E. P., Brown, H. E., van Sluijs, E., Mani, A., Caldeira, S., & Herrmann, B. (2018). Network interventions for changing physical activity behaviour in preadolescents. Nature Human Behaviour, 2(10), 778–787.

Rashotte, L. (2007). Social Influence. In The Blackwell Encyclopedia of Sociology. John Wiley & Sons, Ltd.

Rhodes, R. E., Martin, A. D., & Taunton, J. E. (2001). Temporal Relationships of Self-Efficacy and Social Support as Predictors of Adherence in a 6-Month Strength-Training Program for Older Women. Perceptual and Motor Skills, 93(3), 693–703.

White, J. L., Randsdell, L. B., Vener, J., & Flohr, J. A. (2005). Factors Related to Physical Activity Adherence in Women: Review and Suggestions for Future Research. Women & Health, 41(4), 123–148.

World Health Organization. (2014). Global Status Report on Noncommunicable Diseases. https://apps.who.int/iris/bitstream/handle/10665/148114/9789241564854_eng.pdf