

©Copyright 2022

Li Zeng

Social Support on Fitness Behavior in an Online Fitness Community

Li Zeng

A dissertation
submitted in partial fulfillment of the
requirements for the degree of

Doctor of Philosophy

University of Washington

2022

Reading Committee:

Emma Spiro, Chair

Zack Almquist

Nicholas Weber

Program Authorized to Offer Degree:

Information School

University of Washington

Abstract

Social Support on Fitness Behavior in an Online Fitness Community

Li Zeng

Chair of the Supervisory Committee:
Emma Spiro
Information School

Physical activity provides wide-ranging physical and mental health benefits. However, a large proportion of the population fails to achieve recommended levels of physical activity. With the development of pervasive technology such as smartphones and wearable devices, increasing attention has been paid to promoting positive fitness behaviors through self-tracking and participation in online communities. Fitness-based applications allow users to log and share activity-related data, providing opportunities to observe, analyze and reflect on performance and goals. They also serve as a social network that connects users, promoting social interaction features as a means towards behavioral change and healthy life-style promotion. However, limited work has comprehensively investigated these networks; their structural features and association with behavior (and behavior change) may reveal important dimensions of social exercise and its impacts on participants. These platforms capture information on physical activities as well as peer-to-peer interactions at large scale and low cost, allowing researchers to observe how individuals really engage in physical activity and social interaction, rather than how they report on their participation (e.g. in traditional surveys). Utilizing large-scale behavioral traces collected using novel sampling methods, this dissertation aims to advance our understanding of online social interaction in a fitness community, and how it is associated with fitness behaviors. This work provides new insights into pathways for health promotion and the possibility for network-based health interventions.

TABLE OF CONTENTS

	Page
List of Figures	iii
List of Tables	v
Chapter 1: Introduction	1
1.1 Background	1
1.2 Motivation and Aims	3
1.3 Structure and Contributions of the Dissertation	4
Chapter 2: Relevant Literature	5
2.1 Self Tracking	5
2.2 Online Fitness Communities	6
2.3 Social Interaction, Social Support and Physical Activity	7
2.4 Co-presence in Physical Activity	8
Chapter 3: Data Collection: Methods and Data Description	10
3.1 Introduction	10
3.2 Activity-based Sampling with Strava API	12
3.3 Geographically Bounded Sample of Users and Activities	16
3.4 Limitations	20
Chapter 4: Exercise in an Online Fitness Community	23
4.1 Author Preface	23
4.2 Introduction	25
4.3 Related Work	27
4.4 Data	28
4.5 Methods	30

4.6	Results	32
4.7	Discussion	36
4.8	Conclusion	38
Chapter 5:	Network Interactions and Fitness Behavior	39
5.1	Author Preface	39
5.2	Introduction	41
5.3	Related Work	43
5.4	Data	45
5.5	Methods	47
5.6	Results	51
5.7	Discussion	57
5.8	Conclusion	59
Chapter 6:	Social Exercise in Online Fitness Communities	61
6.1	Author Preface	61
6.2	Introduction	62
6.3	Related Work	64
6.4	Data	66
6.5	Methods	67
6.6	Results	69
6.7	Discussion	76
6.8	Limitations	79
6.9	Conclusion	79
Chapter 7:	Conclusion	80
7.1	Summary of Contributions	81
7.2	The Role of Platform and Design Features	83
7.3	Future Directions	84

LIST OF FIGURES

Figure Number	Page
3.1 The framework of rejection sampling for activity-based Strava data	13
3.2 Pseudo codes of rejection sampling	14
3.3 Pseudo codes of web scraping	15
3.4 An overview of important features in the activity-based sample	16
3.5 A screenshot of a posted activity on the Strava platform. We have to scrape the information inside the red rectangle as it is unavailable from the API. . .	17
3.6 Geo-bounded sample: activity counts over time	18
3.7 Geo-bounded sample: following counts over time	19
3.8 Geo-bounded sample: following counts vs activity level. SF stands for San Francisco County, while Suffolk stands for Boston/Suffolk County.	20
3.9 Geo-bounded sample: distributions of mutual, outgoing and incoming ties in egocentric networks by gender and activity level.	21
3.10 Ave. Ratio of Different Types of Network Ties by Gender	22
4.1 Seasonal pattern of physical activity frequency.	33
4.2 Distributions of activity level of workout by gender, training plan, athlete type and age	34
4.3 Distributions of user follower count by gender, training plan, athlete type and age	35
5.1 Network Degree Distributions: subfigures in the first and the second row show degree distributions of male and female personal networks, respectively. . . .	53
6.1 Proportion of different athlete counts by gender. Each rectangle visualizes the proportion of activities with a specified athlete count. The width and height of a rectangle indicate the proportion of men’s activities and the proportion of women’s activities respectively with that count. If the upper right vertex of a rectangle falls on the dashed line ($y = x$), this indicates the cumulative women’s proportion is equal to cumulative men’s proportion at that point. .	71

6.2	Distribution of solo (yellow) and group (green) activities across the seven weekdays. Corresponding lines map trends over the week.	72
6.3	Heat maps visualizing the distribution of activities across local start time and day of the week for solo and group activities by gender.	73
6.4	Activity performance metrics as a function of number of participants. Colored bands around mean line represent bootstrapped 95% confidence intervals. . .	75
6.5	Measure of social feedback by gender and activity (solo versus group).	77

LIST OF TABLES

Table Number		Page
4.1	Data Summary	30
4.2	Modeling activities using cox hazard models	36
5.1	Descriptive Statistics of Strava Dataset	47
5.2	Descriptive Statistics of Personal Networks on Strava	52
5.3	The best fit models for in- and out-degree distributions of male and female users who enrolled in free plan versus paid plan in Strava.	55
5.4	The best fit models for in- and out-degree distributions of male and female users in different age groups in Strava.	56
6.1	Counts and Percentages of Solo/Group Activities by Gender	70

ACKNOWLEDGMENTS

I am deeply grateful to my dissertation committee Emma S. Spiro, Zack W. Almquist, Robert M. Mason and Nicholas Weber for all their help and support throughout the dissertation process. I would like to express my special thanks to my advisor and dissertation chair Emma Spiro for guiding me through my PhD and for her invaluable advice, continuous support, and patience. I am very fortunate to learn from and work with Emma. She taught me how to conduct rigorous research, how to write academic papers, and she afforded me incredible opportunities to grow as a researcher, a teacher, and a person. I could not have imagined having a better advisor and mentor for my PhD study.

I offer special thanks to Zack W. Almquist. Zack has been fully supportive and has provided me with many insightful comments about my research projects. I am very grateful to have collaborated with him on the Strava project and co-written multiple papers. I can never forget his precious support.

Much of the research presented in this dissertation grew out of a larger collaborative project between University of Washington and University of Minnesota. It was supported by, or in part by, the U.S. Army Research Laboratory and the U.S. Army Research Office under YIP awards #W911NF-14-1-0577 and #W911NF-15-1-0270 and the Office of the Vice President for Research, University of Minnesota.

I will always remember my PhD cohort - Niall, Lavi, Mike, Vero, Sonu, Ligaya, Wan-Chen, Stephen, and Travis. They are a group of fun and smart folks with lots of enthusiasm and optimism and they remind me of how I used to be when I first got

to the UW iSchool. I also want to thank all the current and former members of the iSchool DataLab. Your help, support and encouragement have made my PhD study and life a truly unforgettable time.

Finally, more than anyone else, I want to thank my family for their unconditional love and care. My father, my mother and my grandmother to whom I dedicate this dissertation. I love them so much, and I would not have made it this far without them. I always have my family to count on when times are rough. Their unfailing love for me keeps me going.

DEDICATION

To my beloved family - my parents Bin Zeng and Wenxin Li, and my grandmother Shuchun Zeng; their unconditional love motivates me to embrace life and discover day by day.

Chapter 1

INTRODUCTION

1.1 Background

Physical activity is known to provide important and wide-ranging physical and mental benefits, such as lowering the risk of obesity, cardiovascular disease, diabetes and cancer (Penedo and Dahn, 2005), improving mood states (Janisse et al., 2004), reducing symptoms of depression and anxiety (Ross and Hayes, 1988), etc. However, a large proportion of the world population often fails to achieve recommended levels of physical activity (Hallal et al., 2012; Kohl 3rd et al., 2012). Physical inactivity is now one of the top leading factors of death worldwide (Kallio et al., 2021).

Promotion of physical activity is a priority for health agencies. For example, World Health Organization calls for better and fairer opportunities for physical activity to improve health (WHO et al., 2021). Physical activity behaviors are perceived to be affected by personal, social and environmental factors (Kim et al., 2021; Heath et al., 2012). Therefore, a variety of intervention strategies have been designed aiming to promote physical activity through health education (Sharma, 2021; Young et al., 1996), goal setting (Hardeman et al., 2019; Bandura, 2004; Kahn et al., 2002), social support (Treiber et al., 1991; Laird et al., 2018; Chen and Pu, 2014; Ahtinen et al., 2009), and enhanced built environment (Linenger et al., 1991; Cohen et al., 2012), just to name some examples. Interventions to increase social support for physical activity have been shown to be effective across diverse settings (Kahn et al., 2002; Heath et al., 2012). They focus on creating, strengthening and maintaining social networks that provide support for healthy behavior change; for example one intervention organized people into walking clubs within neighbourhoods, providing community physical activity classes, etc. (Chen and Pu, 2014; Ehn et al., 2018).

With the development of pervasive technology, increasing attention has been paid to promoting positive fitness behaviors through social interaction mechanisms in online fitness communities (Stragier et al., 2016; Centola, 2013; Chen and Pu, 2014). Online fitness communities log activity-related data and help users reflect on their performance. Online fitness communities also serve as activity-based networks that connect users, providing them with a series of social interaction features meant to encourage behavioral change and healthy life-style promotion. One example of such a platform is Strava (<https://www.strava.com/>). Strava users may log and share activities they complete and with whom, subscribe to the activity feed of other users around the world, view their own activity statistics and interact with others by making comments, giving kudos, etc.

Recent studies have made great advancement towards understanding how gamification features motivate physical activity in mobile fitness applications. For example, to compete against other application users or collaborate with other users to gain virtual rewards is found to increase users' activity frequency and intensity (Chen and Pu, 2014; Ali-Hasan et al., 2006; Mueller et al., 2012). However, many open questions remain. In particular we don't know how, when and to what degree interactions through online social networking features influence individual engagement in physical activity. In addition to online social interactions, users who exercise with others physically co-present can often log group activities and co-participants are organized under the group activity "thread" in platforms such as Strava. While previous research finds that exercising with others can produce overall positive health-related effects (Plante et al., 2001; Haskell et al., 1985), the results are based on small-scale experiments and in most cases restricted to two-person exercise events, making generalizability in real-world settings a open question. Research that comprehensively explores the dynamics of diverse social interactions in online fitness communities and their relationships to people's fitness behaviors and behavior change is still in its infancy. This dissertation aims at filling some of these gaps.

1.2 Motivation and Aims

Online fitness communities are increasingly attracting attention from researchers from a variety of disciplines as a site for investigation of social and physical behaviors. Some studies revolve around incentives and interventions. For example, studying the potential of health devices and applications for health-related behavioral change (Foster et al., 2010; Chen and Pu, 2014; Cavallo et al., 2012). Another major thread of work focuses on the technical potential of wearable sensors and human-computer interaction aspects of these technologies. For example, examining specific features in designs of fitness applications (Consolvo et al., 2008; Lister et al., 2014; West et al., 2012). Current social interaction schemes include a variety of social interactions among online fitness community users ranging from offline interactions (e.g. exercising together) to online interactions (e.g. connecting to peers, sharing physical activities, giving kudos and comments on peer’s activities, etc.). However, limited work has comprehensively investigated social interaction and peer support on fitness behavior and behavior change in these online communities.

Further, many early studies of social support are based on data that is either self-reported, which are prone to recall inaccuracies and social desirability bias or from expensive wearable sensors, which are limited in terms of number of participants and duration of observation. Fitness tracking applications and platforms capture people’s behavior and interactions on daily routine bases at a larger scale with lower cost. More importantly, such behavioral traces data have higher credibility and reliability compared to self-reported data (Lampe, 2013).

Therefore, utilizing a large-scale behavioral trace dataset collected from the online fitness community Strava, this dissertation examines in the following: (1) the types and characteristics of activities that individuals track and how they share behavioral data online; (2) what the dynamics of social interactions look like; (3) how people’s fitness behavior are potentially shaped by social interaction and peer support. This work offers insights gleaned from large-scale behavioral traces and their consequences for the design of physical activity interventions

via online social interactions, as well as implications that can contribute to reduction in social and health disparities.

1.3 Structure and Contributions of the Dissertation

The purpose of this dissertation is to demonstrate that large-scale behavioral traces collected from an online fitness community can help us better understand online social interaction and peer support on fitness behavior and behavior change. The dissertation develops novel sampling methods and computational techniques on massive digital behavioral traces, and demonstrates how applying these techniques can derive new insights into pathways for health promotion and the possibility for network-based health interventions. The overall structure of this dissertation is as follows. In Chapter 2, I review the recent studies on self-tracking, online fitness communities, social interaction mechanisms and social support for positive behavior change. In Chapter 3, I present a data collection framework that uses rejection sampling and web scraping to collect and archive online digital traces. I apply these methods to the fitness application Strava and describe the resulting dataset. In Chapter 4, I demonstrate how to leverage large-scale fitness data to reveal people's fitness behavior and social interactions in online fitness communities; I additionally model individuals' activity events to quantify the relationship between social subscription and activity occurrence. In Chapter 5, I describe stochastic process models to characterize the empirical network degree distribution of fitness community members. The results of which demonstrate that gender, age and account status are associated with distinct network structure. Lastly, in Chapter 6, I switch focus from online social support to offline behavior and demonstrate that social support from physically co-present peers is associated with intense activities and more feedback to other users; we also discuss interesting gender difference in the number of people males or females prefer to co-exercise with. These findings speak towards the motivational nature of social exercise. Taken together, this series of studies reveals the potential mechanisms of social influence and their affect on individual behavior change. They demonstrate the potential to model and predict fitness behavior in online fitness communities.

Chapter 2

RELEVANT LITERATURE

2.1 *Self Tracking*

Self-tracking is a longstanding human activity. Traditionally, people have been tracking their behaviors and/or lives for centuries using paper and pen. As technologies such as smartwatches, fitness bands or mobile phones become prevalent, much self-tracking is now digital. This expansion brings new opportunities to self-tracking – both for individuals, groups, and platforms – as well as new challenges. In recent years, there has been an increasing body of research examining how digital self-tracking practices combine with the information technologies to lend themselves towards health promotion and behavior change. Here, I review recent academic work on self-tracking especially within the context of health promotion and medical care. This prior work motivates the research questions I investigate in this dissertation and provides a context for interpreting the findings of the empirical studies presented.

Self-tracking, also sometimes referred to as the “quantified self”, life-logging, personal informatics, etc., involves individuals recording (often in real-time) elements of their daily lives. In 2021, there were over 350,000 digital health apps available on the market, according to a report from IQVIA Institute for Human Data Science (IQVIA, 2021). Such health and medical applications are designed to track individuals’ biological processes such as fitness (Jin et al., 2020; Gui et al., 2017; Molina and Sundar, 2018), calories intake (Saleh et al., 2021; Levinson et al., 2017; Franco et al., 2016), sleep (Kang et al., 2017; Robbins et al., 2019), and more. While archiving data about various aspects of a person’s daily life is at the core of these applications, one of the almost universal aims of these self-tracking applications is to assist users with working towards and hopefully achieving health goals.

In prior work, Lupton proposed five modes of self-tracking: private, pushed, communal, imposed and exploited (Lupton, 2017). For private self-tracking, users voluntarily collect data on themselves for self-knowledge and self-optimisation (e.g. Liu et al. (2021); Hassan et al. (2019); Choe et al. (2014).) Pushed self-tracking differs from private self-tracking in that the initial engagement comes from an external actor or encouragement (for instance, a design of persuasive technology for weight management (Purpura et al., 2011) or a clinic self-tracking to encourage patients to monitor their biometrics (Piras and Miele, 2017)). Communal self-tracking involves voluntary sharing of people’s self-tracking data with others typically via social media platforms and/or online communities (such as on Instagram (Kent, 2020) and on Strava (Couture, 2021)). Imposed self-tracking refers to the imposition of self-tracking practices upon individuals by others primarily for these others’ interests and benefits (for example, see Freeman and Neff (2021)), whereas exploited self-tracking often repurposes collected behavioral data for other parties’ benefits or commercial profits (such as in Acar et al. (2015)).

In all, digital self tracking technologies extend the areas of human daily life that can be measured, and they make it possible to keep track with greater frequency and resolution than ever before. Evidence also suggests that exercise practices themselves are being shaped by the development and use of self tracking tools. In response, we examine the revealed social dynamics and try to figure out the mechanisms behind these mutually reinforcing phenomena.

2.2 Online Fitness Communities

With advances in pervasive technologies, activity tracking applications and online fitness communities such as Strava, RunKeeper, and MapMyRun, are attracting more and more users around the world. These online fitness platforms usually have features allowing both activity logging and social networking (Stragier et al., 2016), aiming to build an online community of users. Online fitness communities log activity-related data and help users analyze their performance. Online fitness communities also serve as an activity-based social

network that connect users together and provide them with a series of available social interactions features meant to encourage peer feedback and healthy life-style promotion. For example, on Strava, users can follow both recreational and professional athletes, view their activities and interact with them by making comments, giving kudos, etc.

The behavioral traces archived by online fitness communities record large amounts of data that are generated by users through their physical activity (e.g., running/cycling) and their social interaction online. One might compare this data to the more traditional observations of fitness behaviors, such as self-reported data collected via surveys which document individual's perception of their workout routine and social engagement. These methods are prone to reporting inaccuracies and bias. Behavioral trace data, on the other hand, records the exact physical activity, timing, distance, etc. and precise social interaction (e.g., running in groups or liking someones run activity). This data shows distinct advantages in terms of scale, granularity and observation duration against data collected via expensive sensors (Althoff, 2017). Thus, it provides researchers with new opportunities for understanding the relationship between fitness behavior and social interaction.

Over the recent years, there has been an increasing body of research on online fitness communities. Some studies examine social sharing on such communities and its impact on individual perceptions, as well as implications for behavior (Meng et al., 2021; Alqhatani and Lipford, 2019). Some studies focus on incentives and interventions, such as studying the potential of health devices and applications for supporting health-related behavioral change (Foster et al., 2010; Chen and Pu, 2014; Cavallo et al., 2012); Another major body of work looks at technical potential of wearable sensors and human-computer interaction aspects of these technologies. For example, examining specific features in designs of fitness applications (Consolvo et al., 2008; Lister et al., 2014; West et al., 2012).

2.3 Social Interaction, Social Support and Physical Activity

Social support and physical activity are known to be linked (Dishman et al., 1985; McAuley et al., 2000; Berkman and Glass, 2000). Social interaction among people in social

networks can lead to social support (Lin, 2017). More recently, the relationship between social support and physical activity has been studied by making use of social interaction features designed for mobile fitness applications (Munson and Consolvo, 2012). Some new platforms gamify the element of social support, allowing users to exercise in a virtual group environment so as to motivate them to perform physical activities (Campbell et al., 2008; Consolvo et al., 2006; Chen and Pu, 2014). Stragier et al. (2016) examines how Strava users' social motives predict perceived usefulness of the platform based on survey responses collected from 394 Strava users. Three aspects of social motives are considered: (1) staying informed on friends' activities, (2) viewing progress made by friends, and (3) receiving support from others via kudos and comments. The results show that social motives influence habitual Strava use directly, and when compared to novice users social motives are more important for experienced users. Also, Chen and Pu (2014) designed a mobile application with gamification settings of competition (i.e. two users compete to gain more virtual rewards by exercising), cooperation (i.e. two users contribute equally to win virtual rewards) and hybrid (i.e. weighting the cooperation and competition settings). Even though users are not required to exercise side by side in the physical environment, all three conditions of virtual group activities were found to increase users' activity frequency and intensity.

2.4 Co-presence in Physical Activity

Co-presence is a sociological concept, describing how the physical presence of other people shapes individual behavior (Zhao, 2003; Goffmann, 1963). Recently, studies have extended the concept of co-presence, introducing technologically mediated (e.g. phones, email, online chat, etc.) or parasocial (e.g. listening to music, watching TV, etc.) presence (McLuhan, 1994; Horton and Richard Wohl, 1956). This work will focus on the traditional concept of co-presence - individuals are physically together and interact with one another in person.

Previous work suggests that factors such as pleasant surroundings, an enthusiastic exercise leader, and sympathetic co-exercisers during leisure-time activities are all likely to relieve negative emotions associated with exercise (Haskell et al., 1985). Interestingly, in one study

that assigned mothers and daughters to university- or home-based physical activity groups found both improved family relations and increased physical activity as a result of group participation (Ransdell et al., 2003). Research by (Lyons et al., 2016; Winters-Stone et al., 2012) recruited couples into a project aimed to understanding how exercising together may help fight against prostate cancer. The results show significant increases in engagement in affectionate behaviors for wives in the project compared to wives in a usual care control group, but no significant intervention effects were found for husbands. In the above studies, co-present exercise partners are family members and it is well-established that family support is an important component for increasing physical activity in adolescents and adults (Biddle and Goudas, 1996; Field et al., 2001; Baranowski et al., 1990). Beyond family support, one laboratory-based study found that exercising with randomly assigned buddies helped to reduce stress and produce overall positive effects on energy, calmness and tiredness, compared with a control group exercising alone (Plante et al., 2001). However, this experiment was conducted in a laboratory setting and participants (recruited from a college student population) were fairly homogeneous in terms of age and fitness levels, making its applicability in real-world settings a open question. This work aims at filling this gap by using behavioral trace data to examine the relationship between co-presence and fitness behavior.

Chapter 3

DATA COLLECTION: METHODS AND DATA DESCRIPTION

3.1 Introduction

This chapter discusses two data collection methods utilized in this work. A key feature of technology-mediated group interactions is the ability of these systems to archive a history of social interaction happening within online communities as a byproduct of their design. Archived digital traces can then be made available to researchers and data analysts – both internally and externally to the platform company. Strava provides a publicly available Application Programming Interface (referred to as the Strava API in the following). Developers and Strava users are allowed to access and interact with the rich Strava dataset via Strava API. The API, however, does not afford access to all the data, the Strava platform records on their users; public data includes but is not limited to, user profile data, activities data, geographical segments data, etc. Other data remains private, for example, GPS traces of activities and exact timing of social relationship creation. The Strava API is continually being improved, but this description captures the functionality at the time of this research.

Traditional methods of collecting behavioral data often involve asking participants to recall their behaviors in surveys and interviews. Compared to traditional data collection, behavioral trace data has three major advantages: scale, validity and granularity (Althoff et al., 2017b; Kosinski et al., 2015). In terms of scale, many of the social media platforms record users' online interactions in large scale – capturing the behaviors of millions of users across the globe and generating a large amount of data potentially available to researchers. Behavioral traces show the actual behaviors of online users and groups, not just those reported by people. The high granularity of behavioral traces ensures the records of detailed behavioral data such as demographic, temporal and performance measurements. There are

multiple methods for getting access to the behavioral trace data of users in these sites. For example, collecting via API as discussed above. Another common method to collect behavioral traces is web scraping which has its benefits and deficits.

Firstly, researchers can use computer programs to access a website and ‘scrape’ the data into an external database. This method can be an effective way to collect data when the researcher does not have access to the databases underlying the online community, but there are many limitations to this approach. Due to concerns about user rights, intellectual property and server efficiency, many sites have “terms of use” that prohibit people from using automated scripts to download user data. This prohibition raises an ethical concern for researchers regarding accessing data “owned” by a private entity. Most online communities in the modern environment are commercial entities and privately owned. Does the researcher have a right to make a copy of data owned by that third party? Ethical issues aside, there are legal issues, as well as technical issues since sites often block computers that use automated scraping from accessing their servers (more discussion of data limitations can be found in Chapter 3.4).

An additional limitation of web scraping is that not all behaviors of interest to the researcher may be visibly posted to the platform, limiting which data can be automatically retrieved with this method. For example, individual user visits are usually not displayed on a page, so would not be accessible. Many sites require users to authenticate and on social network sites many users have privacy setting options that make them invisible to anyone but people they have “friended”. This makes them effectively invisible to researchers scraping data from online platforms, affecting overall sampling and analysis possibilities.

The first sample of behavioral traces collected for this dissertation - *the activity-based sample* was obtained by combining the techniques of data collection via the Strava API and via web scraping. This dataset records individuals’ fitness behavior and social interactions on the platform. The second collection of data utilized in this work was acquired via internal research collaborations with the Strava company (referred to as *geographically bounded sample*). The most unique feature of the second dataset is that it contains a historical record of

posted activities and network formation of the sampled Strava users in two representative areas in the United States: Boston and San Francisco. With the second dataset, we are able to not only further implement event analysis on Strava users' posted activities, but also closely investigate differences in network structures among different groups of Strava users. The following two sections introduce data collection frameworks and/or data processing procedures for these two samples. Moreover, we will also provide an overview of these two raw datasets without further data cleaning and processing for specific research questions in the later chapters.

3.2 Activity-based Sampling with Strava API

3.2.1 A Framework of Collecting Activity-based Sample with Rejection Sampling

We aim at collecting fitness-related data along with user information via Strava API. However, as stated in the previous Chapter 3.1, data from the platform can only be accessed via certain portals. Therefore, one of the major challenges to activity-based sampling is to construct an appropriate sampling framework for user-based research questions (Almquist et al., 2019). Ideally, we would like to draw a random sample from Strava users. However, this method does not work in practice. First, it is unfeasible to determine the unique space of all users. Moreover, since new users might join the platform and current users may quit, this user set is subject to constant change. Like many others, Strava does not allow to retrieve data based on unique userID. Therefore, we need to propose an alternative sampling strategy to collect an approximate random sample of users. In this work, we develop a data collection framework using rejection sampling based on the dense space of activity IDs on the platform.

Rejection sampling is a Monte Carlo algorithm to generate a sample of observations – in the context of this work, this means generating a set of activity identifiers (activityID) that can be used to collect data from the Strava API. Utilizing a rejection sampling strategy, we are able to obtain a sample of user data with the help of a proxy distribution of posted

activities (identified by random activityID) on the Strava platform. Rejection sampling has been successfully applied in the past as a strategy to draw a simple random sample from online platforms where the set of all possible IDs of interest is enumerable (Nazi et al., 2014; Gjoka et al., 2010).

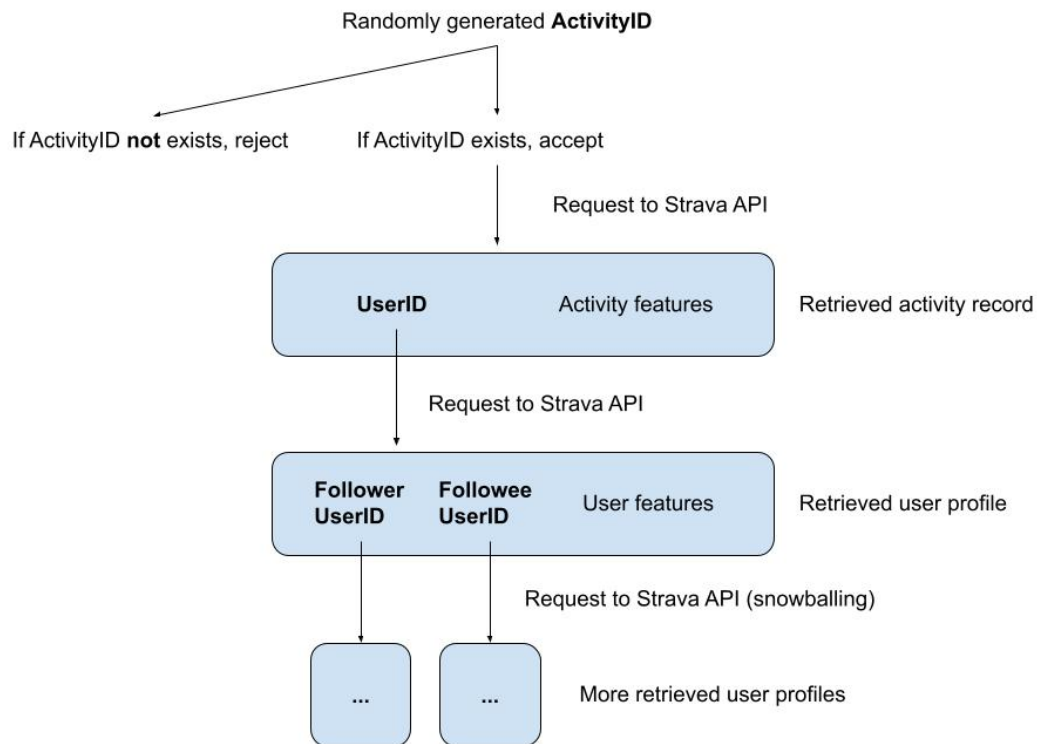


Figure 3.1: The framework of rejection sampling for activity-based Strava data

Figure 3.1 shows the proposed framework of activity-based data collection with rejection sampling. Under this rejection sampling framework, we consider the set of all theoretically possible activity identifiers as our sampling frame and then randomly generate a set of activityIDs as sampling seeds via the Strava API. For each activityID in this set, we execute the API query to see whether the given activityID exists on the platform. In cases where an

activity is found (i.e., that activityID exists), we sample it and collect data. If the activity does not exist, we reject the generated ID from the sample.

Given an existing activityID, it is then possible to utilize the Strava API to retrieve various data related to the activity itself – activity features (e.g., performance, location, timestamp). As each activityID is associated with a particular user, one can further trace additional user-level data. This includes user profiles as well as their network ties. In this way, the existing activityID serves as an entry point through which other data or features, in particular user-related features, can be collected.

Next, we found that only a part of behavioral traces collected by the Strava platform are made available through the API. Other useful data/features, although are displayed on the platform, lack of access through the API, such as a user’s summary stats of total runs/bikings/swims, a list of user’s followers and followees, a list of user’s co-exercises, etc. In this case, web scraping can serve as a supplement to the data that are unavailable through the Strava API but shown on the platform web-pages.

Figure 3.2 and Figure 3.3 summarize this entire process of activity-based data collection in pseudo codes.

```

1 # Step.1 Rejection Sampling
2 Create a set of all possible activity IDs (eg. 000000-999999)
3
4 For each activityID:
5     Make a query to the Strava API using this ID
6     if ID found:
7         #ID found, activity-related features succesfully collected)
8         Save the returned data including userID
9         # to collect user-related features
10        Make another query to the API using the userID and save the data
11    else:
12        #ID not found - reject
13        continue

```

Figure 3.2: Pseudo codes of rejection sampling

```

15 # Step.2 Scraping more user-related data
16 For each userID:
17
18     Construct the URL using this userID to access his/her Strava webpage
19
20     Send a request to the Strava server to retrieve the page info
21
22     # to scrape summary stats of user's activities
23     activity_stats_scraper();
24
25     # to scrape summary stats of user's online network
26     network_stats_scraper();
27
28     # to scrape the list of user's online followers and followees
29     subscription_list_scraper();
30
31     # to scrape the list of user's offline co-exercisers
32     coexerciser_list_scraper();

```

Figure 3.3: Pseudo codes of web scraping

3.2.2 Dataset Description

The activity-based sample contains 888,093 activity records with a rich set of features which can be categorized into four types of characteristics as shown in Figure 3.4. Chapter 6 utilizes this dataset to examine offline social support when people exercise physically co-present.

Note that data acquired from the API only contains summary information. For example, one can collect how many peers physically co-exercised with this user using the API, but detailed information on peers themselves are not included; the number of comments is available from the API, but the textual content of the comments is not. Therefore, we continue to apply web scraping techniques to supplement the API dataset. Figure 3.5 presents a screenshot of a posted activity on the Strava platform.

Activity characteristics	<ul style="list-style-type: none"> •Activity type: run, bike, walk, swim •Activity location: start lat/long, end lat/long •Names, description, moving time, achievements, etc.
User characteristics	<ul style="list-style-type: none"> •Demographics: name, sex •Location •Physiological measurements: height, weight, etc. •Equipment
Subscription relationships	<ul style="list-style-type: none"> •Following: out-ties •Friends: in-ties
Activity-based relationships	<ul style="list-style-type: none"> •Kudos or "likes" •Comments

Figure 3.4: An overview of important features in the activity-based sample

3.3 Geographically Bounded Sample of Users and Activities

3.3.1 Introduction

The second dataset utilized in this work is the geographically bounded sample shared by the Strava company through internal research collaboration. Except for information on co-exercise, this proprietary dataset covers almost all features available from the Strava API. This dataset is a complete sample of active Strava users who live in Boston or San Francisco (at the time of collection). Note that the dataset construction was done by the Strava team and the shared dataset contains around 5% noisy observations where we were unable to tell which regions the users were considered to belong in.

One of the unique features of the geographically bounded sample is historical records of

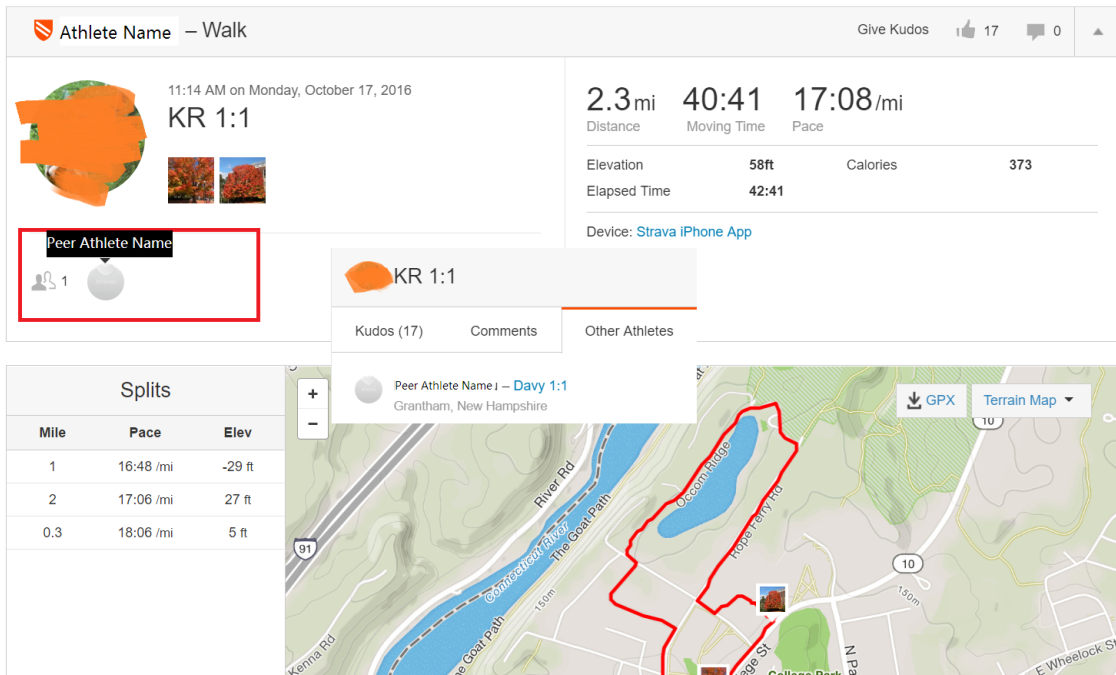


Figure 3.5: A screenshot of a posted activity on the Strava platform. We have to scrape the information inside the red rectangle as it is unavailable from the API.

each user's reported activity since sign-up. We leverage this unique feature to implement an event analysis to characterize the occurrence of the repeated events (i.e. physical activities in our context) given a time-dependent variable follower count in order to simulate how one's activity level changes over time when receiving more network subscription (and potentially more social interaction and peer support).

This dataset also contains detailed formation of network ties (both incoming- and outgoing-ties). Chapter 5 utilizes this dataset to examine how the size of one's personal network is structured by demographics (such as gender and age) and an economic indicator (if an account type is regular or premium); moreover, we model the degree distributions of these personal networks to reveal underlying social mechanisms of friend formation for different groups of users.

3.3.2 Data Description

After the removal of noisy observations, this dataset contains 11,245 anonymized Strava users from two major metro areas - San Francisco (2,312,669 activities) and Boston (293,478 activities). The work in Chapter 5 compares fitness behavior of users in these two areas. Here, I conduct some exploratory analysis in order to describe the data used.

First, we examine the time series of reported activities in the sample. As seen in Figure 3.6, we see a clear seasonal pattern of activity counts. Not surprisingly, summer is the most favorable season for physical activities while winter is the least.

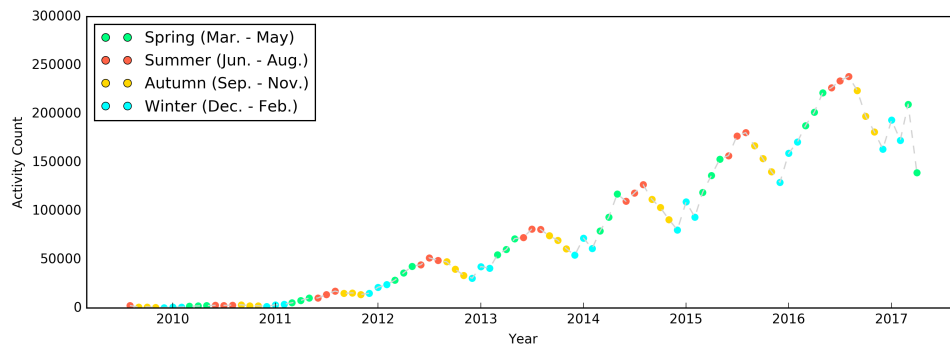


Figure 3.6: Geo-bounded sample: activity counts over time

Next, Figure 3.7 shows the time series of following activities. Again, we observe peaks in summer and less active following in winter. This may also imply a co-relation between users' usage of networking features on the platform and activity level. A line plot to examine the relationship with one's follower count and activity level is also shown in Figure 3.8.

Moreover, leveraging this dataset allows us to further investigate the structure of egocentric networks on Strava and examine whether these patterns are structured along gender and activeness lines. The exploratory analysis examines three types of network ties: mutual ties, out-going ties (i.e. following) and in-coming ties (i.e. followers). Figure 3.9 presents the distributions of different network ties by gender and activity level. Interestingly, active users

Following Over Time

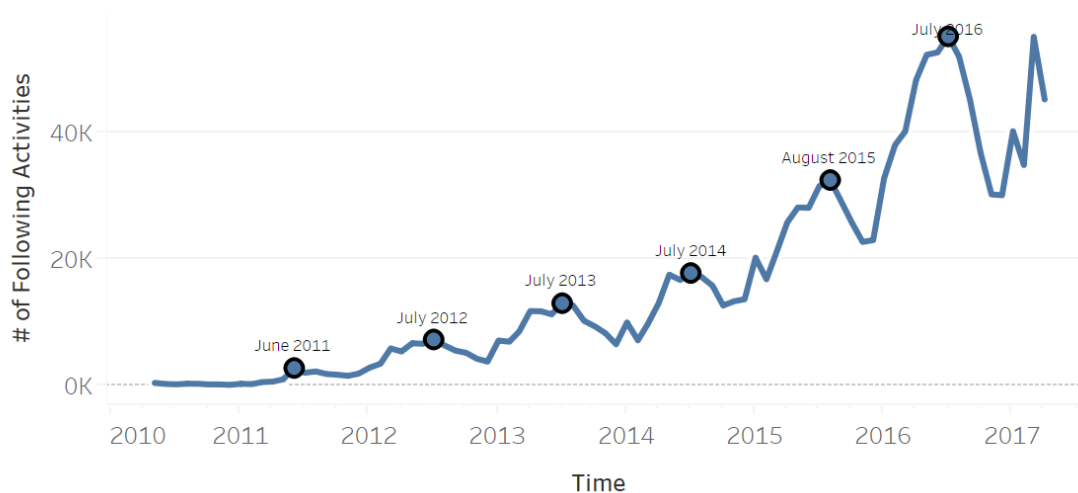


Figure 3.7: Geo-bounded sample: following counts over time

are active in online social networking too. Active users tend to explore the functions of the Strava application more actively. One significant gender difference is that for both active ($p < 0.05$) and inactive categories ($p < 0.01$), female users averagely have less followers than male users. Since distributions of mutual connections for female and gender follow similar patterns, greater follower count on average for men indicates that men may have more asymmetric ties.

Figure 3.10 further looks at the average ratio of mutual to total connections and of followers to following. Overall, inactive users have significantly lower ratio of mutual to total connections than active users (t -statistic = 37.32, $p < 0.001$), indicating that networks of inactive users tend to have more asymmetric connections. Also, it is interesting to find gender differences in the ratio too. Both active and inactive female users are found to have significantly (both $p < 0.001$) higher ratio of mutual to total connections compared to their male counterpart. These exploratory results drive this work to a deeper analysis of network structure in Chapter 5.

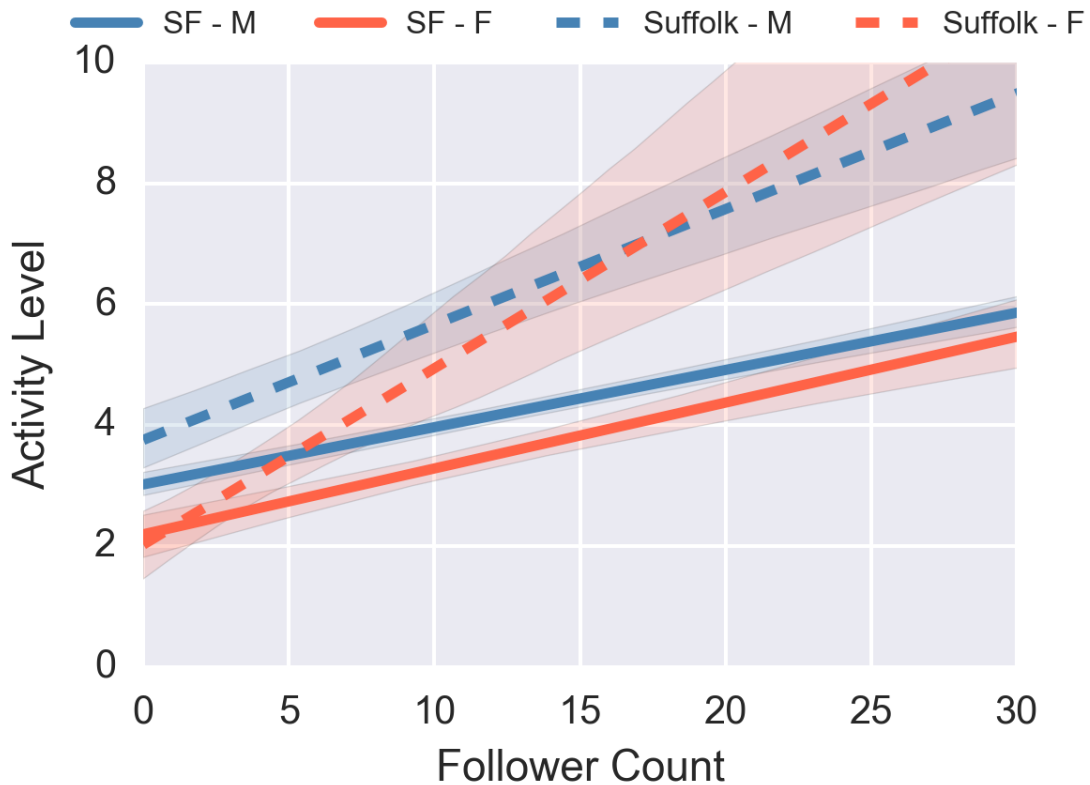


Figure 3.8: Geo-bounded sample: following counts vs activity level. SF stands for San Francisco County, while Suffolk stands for Boston/Suffolk County.

3.4 Limitations

Large-scale digital behavioral traces provide researchers with a great channel to understand both social dynamics and online interactions. However, such data often have limitations in terms of what they can reveal. Behavioral traces record detailed information about what people do with activity tracking applications on a daily basis, but they can rarely explain the reasons and motivations behind people's behavior in the online platforms. For example, Strava behavioral traces tell the patterns of social exercise, but it is difficult to derive causal claims by utilizing this type of data alone.

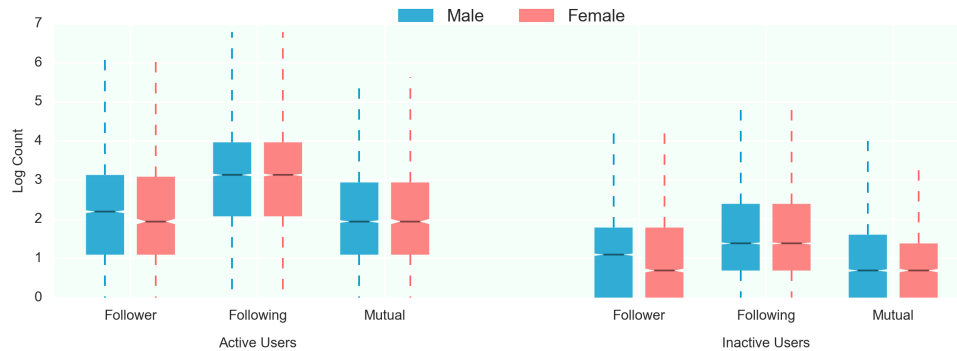
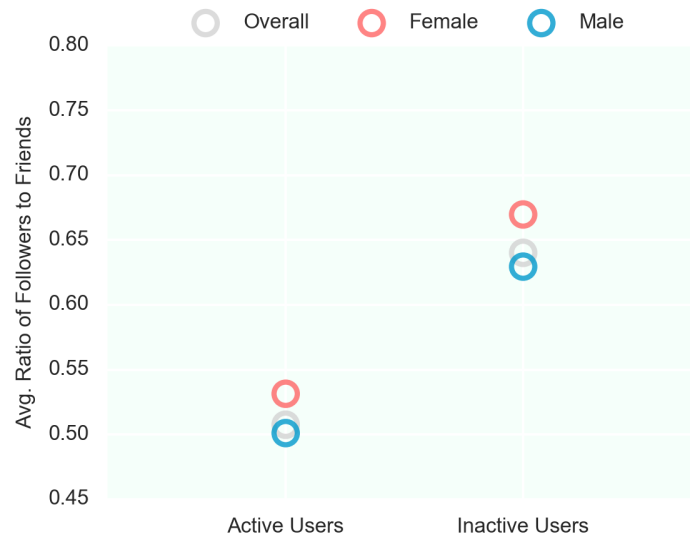


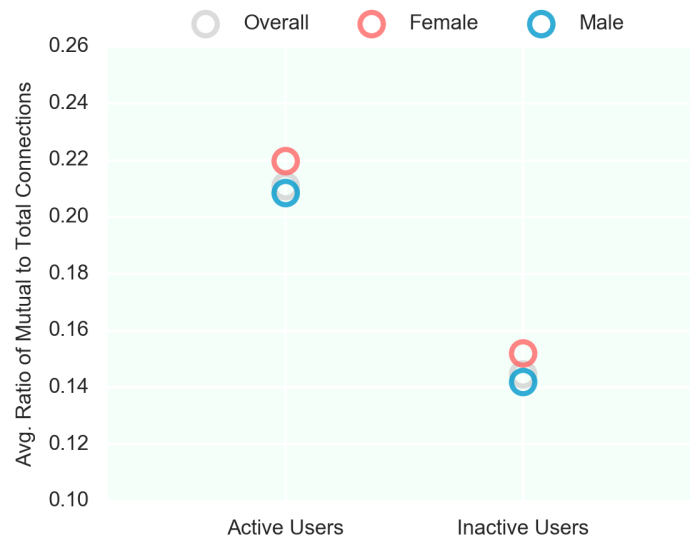
Figure 3.9: Geo-bounded sample: distributions of mutual, outgoing and incoming ties in egocentric networks by gender and activity level.

Several studies have attempted to validate online behaviors with offline beliefs and opinions regarding use, but more multi-methodological work needs to be conducted to triangulate behavioral data with psycho-social mechanisms that underlie the behavior. Multi-methodological work or more multi-dimensional data involves experiments, surveys, etc.

Additionally, behavioral traces collected in the Strava platform are mostly limited to the activities taken place in this platform. Therefore, it has limitations in terms of examining user behaviors in multiple contexts. For instance, active users of fitness applications may use several fitness applications at the same time; they might record and upload their digital behavioral traces to RunKeeper, Google Fit, Nike+ Running, etc. Such fitness platforms maintain separate databases and follow different privacy policies. Indeed, some platforms ask for user's approval of data access, whereas others may allow for anonymous or pseudonymous posting, which makes it extremely difficult to compare individual users across multiple platforms. This, to some extent, limits the level of generalizability that the findings can achieve and the types of claims that researchers can make. However, such data are still able to provide insights into behaviors within individual platforms that generalize to overall online (fitness) communities by connecting these insights to broader theories of mediated interaction.



(a) Avg. Ratio of Followers to Following



(b) Avg. Ratio of Mutual to Total Connections

Figure 3.10: Ave. Ratio of Different Types of Network Ties by Gender

Chapter 4

EXERCISE IN AN ONLINE FITNESS COMMUNITY

4.1 *Author Preface*

The widespread use of sensor-equipped mobile phones and other wearable devices allows researchers to leverage their capacity to collect and analyze digital footprints to reveal human behavior patterns and community dynamics. The breath, depth and scale of such behavioral traces provides an opportunity to better understand the exercise patterns of users. Understanding such exercise patterns would help reveal people’s motivations and barriers to participate or not to participate in physical exercise. This is important because this can lead to improvements of the existing online fitness communities as well as development of effective health intervention strategies. Therefore, the first component of this dissertation aims at answering the major research question: how do people perform physical exercises in online fitness communities? do we observe differences in exercise patterns by users’ identity characteristics? This is one of the fundamental questions studied in this dissertation; not only do the answers to this question provide fundamental insight into how exercise happens, but they also provide the foundation for the following chapters to then examine how social interaction and peer support lead to fitness behavior diversity and behavior change for different types of users. To answer this question, this chapter begins with analysis of exercise dynamics in the Strava platform using a large-scale behavioral trace sample in Boston and San Francisco. We then analyze the network subscription in the Strava platform to understand the correlations between people’s personal network ties and their exercise activity level. The results of an event history analysis quantify the “power” of gaining one follower for users to exercise more and demonstrate a positive relationship between social subscription and activity occurrence.

The following paper is published in proceedings of the 10th International Conference on Social Computing and Social Media, SCSM 2018 held as part of HCI International 2018 in Las Vegas, United States. This work won the Best Paper Award at this venue.

Reference: Zeng, L., Almquist, Z.W., and Spiro, E.S. (2018). Stay Connected and Keep Motivated: Modeling Activity Level of Exercise in an Online Fitness Community. In: Meiselwitz G. (eds) Social Computing and Social Media. Technologies and Analytics. SCSM 2018. Lecture Notes in Computer Science, vol 10914. Springer, Cham. https://doi.org/10.1007/978-3-319-91485-5_10

<<THE PUBLISHED PAPER BEGINS BELOW>>

Abstract

Recent years have witnessed a growing popularity of activity tracking applications. Previously work has focused on three major types of social interaction features in such applications: cooperation, competition and community. Such features motivate users to be more active in exercise and stay within the track of positive behavior change. Online fitness communities such as Strava encourage users to connect to peers and provide a rich set of social interaction features. Utilizing a large-scale behavioral trace data set, this work aims to analyze the dynamics of online fitness behaviors and network subscription as well as the relationship between them. Our results indicate that activeness of fitness behaviors not only has seasonal variations, but also vary by user group and how well users are connected in an online fitness community. These results provide important implications for studies on network-based health and design of application features for health promotion.

Keywords: fitness behaviors; social interaction; event history analysis; online fitness communities; social media; behavioral traces

4.2 Introduction

Peer influence and social interaction have been found to have positive health-related effects, such as helping people lose weight and increasing physical activities (Ahtinen et al., 2009; Chen and Pu, 2014; Wing and Jeffery, 1999; Kulik and Mahler, 1989; Dishman et al., 1985). Recently, increased attention has been paid to promoting health habits through online social interaction in online fitness communities (e.g., running or cycling online groups) (Centola and van de Rijt, 2015; King et al., 2015). Studies show that social interaction features such as cooperation and competition provide participants with a group of peers and help motivate them to reach their fitness or health goals (Chen and Pu, 2014; Hamari and Koivisto, 2015). In this work, we study the relationship between fitness behaviors and online social behavior via a subscription service where users can post and follow peers' activity feeds.

Traditionally, studies on physical activity and social interaction have relied on information that is self-reported (e.g., diary studies) or measured via (expensive) wearable sensors, which have been limited in terms of scale, granularity of activity and duration of observation period (Stragier et al., 2016; Althoff, 2017). Recent developments in smart phone GPS tracking and accessibility has provided for an increasing large adoption of mobile devices which track everyday physical activities. These resultant behavioral trace data allow for precise measurement of individual’s activities and online social action in scale and cost that traditional survey-based methods for collecting data on physical activities cannot match, and provide a good alternative to classic sensor studies (Prince et al., 2008; Bernard et al., 1984). Our work leverages these new data sources (e.g., behavioral trace data) to explore the relationship between fitness behaviors and online social interaction over time. We employ this novel data set along with event history methods to understand the relationship between online social interaction and activity levels within one of these app-based activity communities. Specifically, we focus on the large app-based community known as Strava where users have covered over 12 billion miles worldwide (Strava, 2017). In this work we focus on two comparable major metropolitan areas in the US.

This work analyzes the dynamics of online fitness behaviors and network subscription as well as the relationship between them. We ask the following research questions: (1) how do users’ activity levels of exercise change over time? (2) how do exercise activeness and network subscription vary among users? (3) how subscription magnitude is associated with activity occurrence? We find clear seasonal patterns of users’ fitness behaviors and discuss the implications for fitness application designs and health prevention. We also show that paid-plan users exercise more actively and attract more followers than free-plan users. Last, our analysis shows a positive relationship between social subscription and physical activities, supporting the claim – “stay connected, keep motivated.”

The remainder of this paper is organized as follows. We start by reviewing existing studies on online fitness communities and roles of social interaction in physical activities in online fitness communities. Next, we describe the behavioral trace data collected from Strava for

this work. We then describe the methodology for this work, describing our analysis on activity levels and online social interaction via network subscription, as well as the methodology used to model physical activities given user characteristics and network subscription. We finally discuss and summarize our findings.

4.3 Related Work

With advances in pervasive technologies, activity tracking applications and online fitness communities such as Strava, RunKeeper, MapMyRun, etc. are attracting more and more users around the world. Online fitness communities usually have features of both activity logging and social networking (Stragier et al., 2016). Online fitness communities log activity-related data and help users analyze their performance. Online fitness communities also serve as an activity-based network that connect users and provide users with a series of social interactions features meant to encourage behavioral change and healthy life-style promotion. For example, on Strava, users can follow both recreational and professional athletes, view their activities and interact with them by making comments, giving kudos, etc.

Behavioral trace data archived by online fitness communities record large amount of data that is generated by users throughout their physical activity (e.g., running/cycling) and through their social interaction online. One can compare this data to the more often collected self-report surveys which acquire individual’s perception of their workout routine and social engagement, rather than the behavioral trace data which records their exact physical activity, timing, distance, etc. and precise social interaction (e.g., running in groups or liking someones run activity). This data source also shows advantages in terms of scale, granularity and observation duration against data collected via expensive sensors (Althoff, 2017). Thus, it provides researchers with new opportunities for understanding the relationship between fitness behavior and social interaction.

Online fitness communities are increasingly attracting researchers from a variety of areas. Some studies revolve around incentives and interventions. For example, studying potential of health devices and applications for health-related behavioral change (Foster et al., 2010;

Chen and Pu, 2014; Cavallo et al., 2012). Another major body of work focuses on the technical potential of wearable sensors and human-computer interaction aspects of these technologies. For example, examining specific features in designs of fitness applications (Consolvo et al., 2008; Lister et al., 2014; West et al., 2012). However, there are few studies on how online social network structure influence activity engagement, leaving a gap in our current understanding of the social dynamics in these settings.

Recent work have studied the relationship between social interaction and physical activity using fitness applications. Social interaction in online fitness communities may include cooperation and competition and sharing physical activities. Studies by (Campbell et al., 2008; Choi and Chung, 2013; Consolvo et al., 2006; Chen and Pu, 2014) suggest that social interaction is essential to motivate users to perform physical activities. As mentioned above, online fitness communities enable users to connect and interact with a group of peers online. In the case of Strava, users can follow other athletes, view their profile and activities, and receive activity feed once their peer post a completed activity. They may also compare workout and network-based stats with each other, “like” others’ posted activities and make comments under posts. The work by Stragier et al. (2016) examines how Strava users’ social motives predict perceived usefulness of the platform based on survey responses collected from 394 Strava users. Three aspects of social motives are considered: staying informed on friends’ activities, viewing progress made by friends, and receiving support from others via kudos and comments. The results show that social motives influence habitual Strava use directly, and when compared to novice users social motives are more important for experienced users.

4.4 Data

This work utilizes behavioral trace data collected from the activity-based network, Strava. Strava sits at the intersection of social media and activity tracking applications and is known colloquially as the “Facebook” of activity-based apps; users have the ability to not only track and log their activities, but also connect to and interact with a group of peers online. The platform continues to grow in popularity among cyclists and runners in recent years around

the world.

In this work, we study 2,605,147 cycling and running activities from 11,245 anonymized users from an activity-based tracking platform known as Strava. Our data includes but is not limit to the following three main components: (1) user profile information, including gender, date of birth, location, user account status (e.g. free or paid plan), sign-up date, etc.; (2) logs of posted activity for users in the sample over time, including activity timestamp, location, type, performance stats, etc.; and (3) social subscription: who followed whom and the corresponding timestamp.

We focus on two major metro areas within the continental US, which have a large active set of Strava users. We have chosen to focus on San Francisco City/County, CA which is where Strava started and continues to be headquartered, and Boston/Suffolk County, MA. Both cities represent similar size metropolitan areas within the US. Boston metro is ranked 10th with about 4.8 million residents, and San Francisco metro is ranked 11th with about 4.7 million residents. Thus, these two areas represent comparable cities on the coasts within the US context, but with wildly different weather patterns. For example, Boston has an average high of 36 degrees Fahrenheit in January, and San Francisco has an average high of 58 degrees Fahrenheit in same month. Our analysis examines and compares fitness behaviors and social subscription behaviors of users from these two counties. The key difference in these two areas is their weather patterns, and so we expect differences in community activities to stem from these seasonal differences.

Table 5.1 presents the summary of our data in terms of user group and activities. Male users are proportionally greater than female users in San Francisco and Suffolk County, and gender proportion is relatively similar in both locations. There exist more cyclists than runners in both groups – this is to be expected as Strava was originally developed by a group of cyclists. San Francisco users have 69.5% of self-reported cyclists and 30.5% of runners, while the proportion of cyclists and runners of Suffolk County users are 62.3% and 37.7%, respectively. Lastly, we observe that the major age group of users is between 18 and 35 for both locations, and the second majority age group is between 36-49.

Table 4.1: Data Summary

	San Francisco County	Suffolk County
# of Activities	2312669	292478
# of Users	9831	1414
Gender Proportion		
- Male	0.777	0.760
- Female	0.223	0.240
User Type Proportion		
- Cyclist	0.695	0.623
- Runner	0.305	0.377
Age Proportion		
- 0-17	0.005	0.008
- 18-35	0.524	0.590
- 36-49	0.380	0.303
- 50+	0.091	0.099

4.5 Methods

First, we examine seasonal patterns of workouts done by users in our sample. Specifically, we look at total numbers of posted activities across all users ranging from September, 2009 to April, 2017. Theoretically, seasonality of physical activities may vary by activity type and location. Hence, this analysis focuses on the two major activity types – cycling and running, which accounts for 53% and 40% of total activities, respectively. Moreover, we do so for users from San Francisco County and Suffolk County separately in order to have a simple control setting for weather in the analysis.

As we are interested in the relationship between fitness behaviors and online social interaction over time, we examine users’ activity level and network subscription by different

user groups. We begin with a simple metric *activity level* measuring to what extent a user engages Strava to track/log exercise actively:

$$\text{Activity level} = \frac{\# \text{ of activities by a user}}{\# \text{ of days a user in Strava service}}$$

We focus on how users' activity level and number of their followers varies by gender, age group, training plan enrolled (i.e. free or paid plan) and athlete type (i.e. cyclist or runner). For gender differences in physical activity, prior work argues that physical inactivity is more prevalent among female. Indeed, male users are proportionally greater than female users in Strava. However, we are also interested in exploring gender differences in activeness in the online fitness community. Next, we observe a disproportionate distribution of users age in our sample data (i.e., the majority of Strava users aged between 18 and 49). Therefore, we want to examine how actively each user age group engages these activities. Further, we are interested in finding whether users from paid plan workout more than free-plan users. We are also interested understanding how workout activity differ by athlete type.

Last, we model user workout frequency over time with a particular interest in examining whether a user who is followed by more peers tend to exercise more. In this work, we perform an event history analysis in order to characterize the occurrence of the repeated events – physical activities given a time-dependent variable – follower count along with time-independent variables – gender, age, training plan enrolled and athlete type (i.e. cyclist or runner). To control for seasonality of location we include a dummy variable to indicate whether a user comes from San Francisco County (indicator = 0) or Suffolk County (indicator = 1). We use the popular cox proportional hazards model where the unique effect of a unit increase in a covariate is multiplicative with respect to the hazard rate. In this model, the dependent variable is $h(t)$ – the hazard of an event at time t . Roughly speaking, $h(t)$ can be interpreted as the instantaneous probability that an event will occur at time t . The hazard function follows the form:

$$\lambda(t|X_i) = \lambda_0(t)\exp(\beta_1 X_{i1} + \cdots + \beta_p X_{ip} + \beta_{p+1} Z_i(t))$$

where X_{ip}, \dots, X_{ip} are the 1st to the p^{th} time-independent covariate and $Z_i(t)$ is the time-dependent covariate at time t for the observation i . Users who churned the Strava service and hence deleted the account are censored in this analysis.

4.6 Results

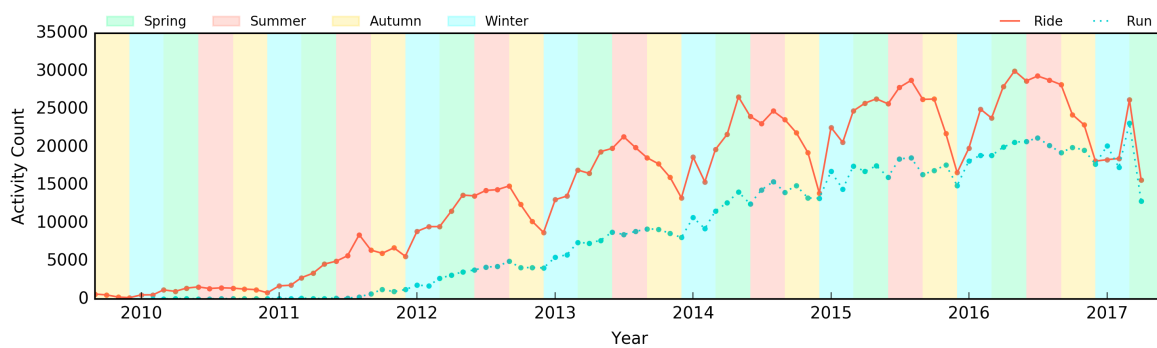
4.6.1 Seasonality of Physical Activity

Figure 4.1 shows a time series of activity frequency in Strava ranging from 2009 to 2017. For both runs and rides, we first observe a clear seasonal pattern for both locations; number of activities done per year usually peak at late spring or summer and then drop to the local lowest point at winter. When comparing between the two locations, we observe Suffolk users' activity behaviors are more greatly influenced by the time of year. Unsurprisingly, this indicates that weather is indeed a great factor for exercise activeness especially for outdoor exercise. Further, it is interesting to notice that cycling is more subject to seasonal change with larger difference in frequency during different seasons when compared to running. Moreover, it is important to note an overall increasing trend for both running and cycling, which indicates an increasing popularity of exercise within the Strava app over time. This analysis suggests that it is important to consider temporal dimension in the analysis and modeling of user fitness behaviors.

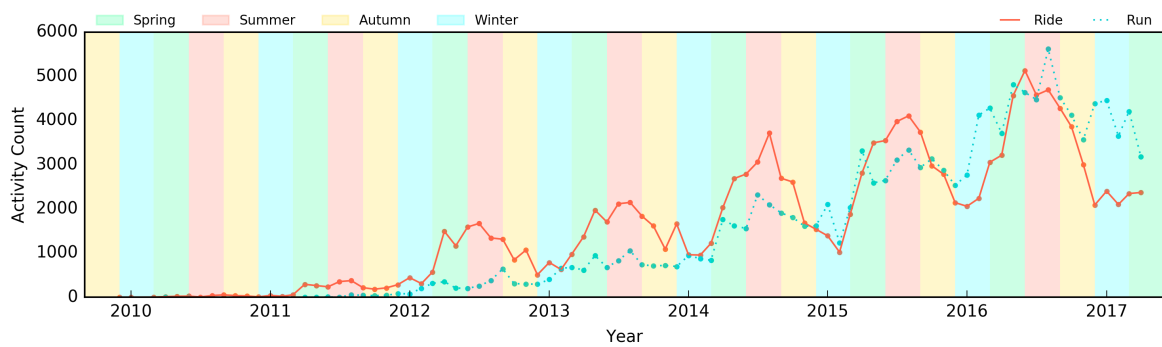
4.6.2 Exercise Activeness and Network Subscription

Figure 4.2 and Figure 4.3 show how users' activity level and follower count vary in terms of gender, training plan, athlete type and age. We compare the two locations with the entire set to see if there is any spatial difference. Overall, we find that distributions of activity level and follower count follow similar patterns for both locations.

One significant difference is between users with free plan and users with paid plan. We observe that paid-plan users tend to exercise more actively and have more followers. It is worth pointing out that on average female paid-plan users tend to have more followers than



(a) San Francisco County



(b) Suffolk County

Figure 4.1: Seasonal pattern of physical activity frequency.

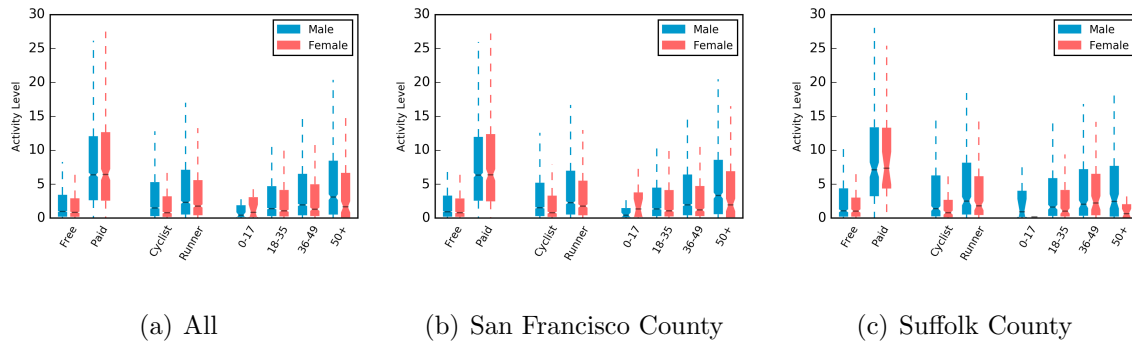


Figure 4.2: Distributions of activity level of workout by gender, training plan, athlete type and age

their male paid-plan counterparts, while we do not see significant gender difference in follower count in free-plan user group.

We find that the activity level of cyclists is on average higher than runners, and male cyclists and runners tend to be more active than their female counterpart. Future study might examine users and their activities from other areas and countries for a generalization of these findings.

Last, we observe that activity level differs by age group and gender. Men who are over 50 tend to be more active than men in other age groups, and the pattern seems consistent across two locations. However, for female users in Suffolk County differences in activity level by age group is larger; women who are between 36 and 49 tend to be more active than women who are younger or older. For gender differences, we find that men tend to be exercise more than women across all age groups. While mid-aged or older-aged users tend to exercise more actively, the results show that younger users tend to receive more follower counts. However, note that the age group of 0-17 contains only a small number of observations and may not be representative of the larger sub-population.

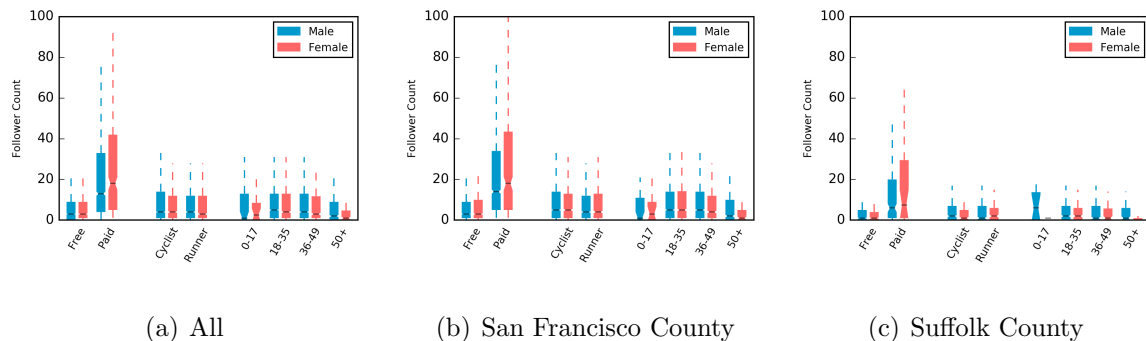


Figure 4.3: Distributions of user follower count by gender, training plan, athlete type and age

4.6.3 Modeling Activity Occurrence

Overall, in the Figure 4.2 and Figure 4.3 we observe that activity level does not always align with social subscription level. For example, users who are over 50 exercise quite actively but have far fewer followers compared to other age groups. This suggests that there might exist a more complex relationship between activity level and online social interaction. Therefore, we move on to discuss the results of modeling users' physical activities over time, given variables of interest that are explored in our previous analysis.

Table 4.2 presents the results of this event history analysis. To recall, the variables in the model are follower count which is time dependent as well as age, gender, athlete type and location which are time independent. Time independent variables except for age are categorical variables in this analysis. The results of cox model show that gender of male, runner and paid plan are significantly related to increase in possibility of activity occurrence, supporting the findings from our previous exploratory data analysis.

Moreover, our cox model reveals that every one unit increase in follower count results in an increased 2.1% probability of an exercise occurrence (P-value < 0.001). For instance, the model suggests that a user who has 50 followers has an approximately 100% increase in the probability of performing a physical activity. Therefore, even though 2.1% appears to be a modest boost for activity occurrence, this could be a relatively large boost given that

the followers is a count variable. This suggests that a greater follower count that a user has is thus correlated to a higher probability that the user exercises. Strava users who have more followers are experiencing more exposure of their posted activities to their followers and likely receiving more social feedback (i.e. comments, kudos) from them.

Table 4.2: Modeling activities using cox hazard models

Variable	Exponential	p-value
Age	1.014	< 0.001
Gender (Male)	1.407	< 0.001
Athlete Type (Runner)	1.807	< 0.001
Plan (Paid)	4.345	< 0.001
Follower count	1.021	< 0.001
Location (San Francisco)	1.217	0.038

4.7 Discussion

In online fitness community such as Strava, a rich set of social interaction features starts with following other athletes and hence building users' activity-based social network. Therefore, our work aims to analyze the dynamics of online fitness behaviors and network subscription as well as the relationship between them. Specifically, we ask how users' activity levels of exercise change over time, how exercise activeness and network subscription vary among users, and how subscription magnitude is associated with activity occurrence. We utilize behavioral trace data from the online community Strava to answer these research questions. Data focus on two major U.S. metro areas that have a large number of active Strava users - San Francisco, CA and Suffolk, MA; data contain profile information of sampled users, user activity logs as well as network subscription logs.

We find that users' fitness behaviors display clear seasonal patterns. In general, late spring and summer are more attractive seasons for rides and runs, whereas winter appears

to be less attractive. We also observe that compared to running, cycling is more sensitive to seasonality. Although strong seasonal patterns of physical activities (especially outdoors activities) are unsurprising in human behavior, results demonstrate that individual physical inactivity is likely to be aligned with seasonality in a systematic way. For designers of fitness applications, an implication of the analysis may be to take into account both individual exercise preference and optimal seasons for certain activity types. For example, Strava supports a great variety of activity type, but current practices in using and advertising the application are limited to outdoors activities, (mostly cycling and running).

Paid-plan users exercise more actively and attract more followers than free-plan users. We also observe significant gender differences in follower counts among paid-plan users; while activity levels of paid-plan users do not vary much by gender, active female users tend to have more followers than male users do. However, reasons behind the findings require a further investigation. It could be that active female users tend to connect to more users and hence receive more followers in return. Future work might also examine gender differences in the way that networks are structured in terms of symmetric and asymmetric ties for free-plan and paid-plan users.

The results demonstrate a positive relationship between social subscription and activity occurrence. Modeling individual activity occurrence using event history analysis enables us to quantify the “power” of gaining one follower for users to exercise more. In this work, we focus on characteristics of egos (eg. gender, training plan, age group, etc). One analysis that may be worth to perform next is to take into account nodal covariates for both egos and alters. For example, users who have many active followers versus users who have many inactive followers; or users who are mostly followed by the same gender versus users who are mostly followed by users whose gender differs from them. Also, built upon the findings of this work, future work may further compare one-way connections with mutual connections to see which type of connections has a stronger association with user activity levels of exercise.

4.8 *Conclusion*

Our work analyzes the dynamics of online fitness behaviors and network subscription as well as the relationship between them. We utilize a large-scale behavioral trace data set from an online fitness community Strava. Our results indicate that fitness activity levels not only has seasonal variations, but also vary by user group. The results of event history analysis suggest that individual activity levels are significantly associated with how well users are connected in an online fitness community. The implications of these results for studies on network-based health and design of application features for health promotion are also discussed.

<<END OF PUBLICATION>>

Chapter 5

NETWORK INTERACTIONS AND FITNESS BEHAVIOR

5.1 *Author Preface*

As suggested in Chapter 4, the frequency with which a user exercises, i.e. their offline fitness behavior, is closely related to personal network size in the online fitness community. This association could have important implications, especially as extensive prior work has documented the association between social networks and individuals' decisions and behavior, as well as on social and material support. In this chapter, I focus on the question of network interaction, describing both the structure of personal networks in online fitness communities, as well as the ways in which there are systematic differences in this structure by identity characteristics.

Online fitness communities are often designed to encourage users to make connections to online peers through suggesting connection or other features. With a click of a button, users can reach out to others who might live down the block or may live on the other side of the world; one might hypothesize then that individuals form their online social networks differently than traditional social network formation processes. On the other hand, prior work has shown that traditional opportunity structure still effect online social networks (Spiro et al., 2016). Demographics and socio-economic status characteristics are known to lead to differences in social network structure; does this also hold in online fitness communities? To address this gap, we further examine how people's online personal network size varies. Specifically, we utilize the geographically bounded dataset to study how demographics, socio-economic status and locations may structure people's social networks in online fitness communities.

The large-scale analysis in this chapter found a distinct gender difference in people's online

network structure. Men who pay for the service, and in the age group 30-49 tend to have larger networks, i.e. more peers with whom they engage on the platform. Women, along with the youngest and oldest users have smaller personal networks – fewer individuals to turn to for social support and motivation in reaching fitness goals. We also model the degree distribution of personal networks to study the underlying mechanisms of tie formations. The findings show that gender may also play an important role in the mechanisms for relation formation.

This paper is published in the Hawaii International Conference on System Sciences (HICSS). I presented this work in Maui, Hawaii in 2019.

Reference: Zeng, L., Almquist, Z. and Spiro, E., 2019, January. “Friending’ in Online Fitness Communities: Exploring Activity-Based Online Network Structure.” *In Proceedings of the 52nd Hawaii International Conference on System Sciences*.

<<THE PUBLISHED PAPER BEGINS BELOW>>

Abstract

Individuals are influenced by both direct and indirect interaction with their social contacts. While peer influence is known to affect health-related outcomes such as exercise, limited work has fully explored how social networks are structured to support (or inhibit) interaction that could lead to positive health behaviors. With the development of pervasive technology and rise of personal health and wellness tracking, increasing attention has been paid to promoting positive fitness behaviors through social interaction mechanisms in online fitness communities. This trend offers a unique opportunity to understand the opportunity structures for personal health and wellness support. Utilizing a large-scale behavioral trace dataset from the online fitness community Strava, we examine how the size of people’s personal network is structured by demographics (e.g. gender and age) and an economic indicator (i.e. if they pay for a premium account). We employ stochastic process models to characterize the empirical network degree distributions in this population of fitness community members. We find that gender, age and account status are associated with distinct network structure. Results have implications in the analysis and the design of health interventions that make use of network relationships in online settings.

5.2 Introduction

People are influenced by the social networks in which they are embedded. It is known that direct and indirect social interaction affect not only people’s opinions and attitudes, but also the actions they take and decisions they make (Marsden and Friedkin, 1993; Aral and Nicolaides, 2017). Social interaction among individuals in a social network leads to social integration, social cohesion and social support (Lin, 2017; Miyata, 2002). A number of prior studies have demonstrated that social support and peer influence have positive effects on health outcomes, for example motivating people to exercise more, lose weight and quit smoking (Chen and Pu, 2014; Pechmann et al., 2015). However, research into the social dynamics of health behaviors is still in its infancy. Indeed, while much previous research has

studied the function of social support (Miyata, 2002; Chen and Pu, 2014), limited work has focused on the structural mechanisms that encourage or suppress certain social interactions, and therefore pathways for peer influence and social support.

Recent years have seen a rapidly growing body of research on promoting health and fitness through social interaction in online fitness communities (Centola, 2013; Chen et al., 2018; Zeng et al., 2018). Online fitness communities not only track and archive activities, but also adopt social networking features meant for increasing user engagement and retention (Zeng et al., 2018; Althoff et al., 2017a). One example of such a platform is Strava. Strava allows users to post activity updates, interact with others by commenting and liking others' activities, and designate subscription relationships to keep up to date on the activities posted by friends and other users of interest. Prior research demonstrates that people think carefully when deciding with whom to interact in online health communities (Newman et al., 2011). We build on this work and ask the question: what do people's personal networks look like in online fitness communities, and can we understand underlying "friending" processes from the aggregate characteristics of networks across the population? Understanding of the underlying mechanisms of relationship construction in this context is an essential step in analysis and design of health interventions that aim to utilize network structure and processes to promote specific behaviors.

Online activity tracking platforms record digital traces of human behavior as people use the service, offering a new data source and an unprecedented opportunity to study social networks at scale and with finer temporal granularity. More importantly, behavioral trace data allows researchers to observe how individuals really engage in physical activity and social interaction, rather than their reported participation as in much prior work in this area. Researchers can test existing theories on the large-scale data, as well as study new forms of activity-related behaviors in online communities.

In this work we examine network structure, exploring potential underlying mechanisms for the formation of personal networks. We utilize a large-scale collection of behavioral traces of social relationships in the online fitness community Strava over eight years. Specifically,

we aim to address the following research questions: (1) *How are activity-based egocentric networks patterned (in terms of in-coming and out-going ties), and are these patterns structured along demographic and economic lines?*, and (2) *Does observed network structure match specific underlying mechanisms for the formation of social relationships? If so, what differences are observed between social and demographic groups?*

5.3 Related Work

5.3.1 Personal Networks

Demographic and socioeconomic factors are known to structure real-world networks; for example, Hill and Dunbar (2003) analyzed a network of Christmas card giving and found that age, household type, and the relationship to the card giver are associated with the resulting network structure. Homophily is observed time and time again in real-world networks. Individuals associate with others who have similar characteristics, interests, opinions and behaviors. Homophilous relations can be formed based on shared demographics including race, ethnicity, age, religion, education (Kandel, 1978; McPherson et al., 2001). Geographic propinquity has also been found to structure human interaction and hence social relationships (McPherson et al., 2001). Newman et al. (2011) found that users in online health communities prefer to interact with those who are in similar situations whether it's losing weight, managing diabetes or having similar health goals.

Prior work has studied not only the structure of social networks, but also the processes that could construct them. The implications of understanding the underlying social processes vary from case to case. For example, Handcock and Jones (2004, 2006) proposed a set of stochastic/generative models for personal networks. They demonstrate how understanding the potential underlying mechanism of relationship construction in this type of network is vital for public health interventions - attempts to control and eventually eradicate sexually-transmitted diseases. Another work (Mislove et al., 2007) analyzed the structure of multiple online social networks (e.g. Flickr, YouTube, LiveJournal, and Orkut) and found a much

larger proportion of symmetric ties and much higher level of local clustering in these social networking sites than other networks. In this case, the authors argue that understanding the structure of online social networks helps explain the robustness and security of such networks. In each of these cases, researchers aim to understand how opportunities for interaction lead to a specific pattern of social relationships, which in turn affects processes that occur via the network ties.

5.3.2 Online Fitness Communities

Online fitness communities are web and mobile application-based environments allowing users to interact in order to support and encourage others to achieve fitness goals. Recently, we have seen the release of many online fitness communities, paralleling the explosion of personal health and wellness tracking wearable devices. Strava, RunKeeper, Nike+, and MapMyRun are just some of the popular new platforms. Online fitness communities usually combine both activity tracking features and social networking features (Stragier et al., 2016; Althoff et al., 2017a). Strava illustrates this trend. It has marketed itself as an online fitness community for runners and cyclists, allowing users to post activities and “prove it” to their peers and friends. The platform also allows users to follow others - both recreational and professional athletes, view their activities and interact with others by making comments, giving kudos, etc. on specific activities posted.

Online fitness communities record large amounts of behavioral trace data generated by users - both via their exercise behaviors as well as through their online social interaction. Compared to self-reported data, which measures individuals’ perceptions of their workout routine and social engagement, behavioral trace data has many advantages including the ability to record people’s exact activity performance and precise social interaction (e.g., running in groups, comments and likes given to someone’s activity). Behavioral trace data is not subject to the same issues of accuracy and recall error as more traditional data collection methods. This data source also has advantages in terms of scale, granularity and observation duration against data collected via expensive sensors (Althoff, 2017). Thus it offers a new

window into how social interaction unfolds in large-scale online social networks.

Most research on online fitness communities focuses on design of incentives and interventions (Chen and Pu, 2014; Foster et al., 2010; Cavallo et al., 2012), or on the technical potential of wearable sensors and other tracking technologies (Consolvo et al., 2008; Lister et al., 2014; West et al., 2012). A much less explored aspect of this phenomena is the dynamics of social interaction and how the resulting networks offer opportunities and constraints on peer influence and health outcomes. Zeng et al. (2017), for example, found that exercising with peers physically co-present is associated with greater exercise performance. Clearly more work is needed to characterize and explore the consequences of social network structure in these settings.

5.4 Data

This work utilizes behavioral trace data from an online activity-tracking platform, Strava. Since its launch in 2009, Strava has been a rapidly growing social network for athletes around the world. Strava incorporates features from both activity tracking applications and social media. Users can track, record and share physical activities (e.g. running, cycling) via the Strava website and/or mobile application. The social media side of Strava is similar to social networking sites such as Facebook or Twitter. The platform allows users to explore and search for other users. Before connecting to peers on the platform, users can view their profiles (e.g. name, photo, location), activity posts, achievements, and their personal networks. Users can create out-going connections to their peers by simply clicking a “Follow” button (as on Twitter) unless follow requests are required by the target user. Users will be notified if someone has followed them and can create mutual connections if interested. As discussed subsequently, many relationships remain asymmetric in this setting.

This work studies personal networks of 69,658 anonymized Strava users in the community. The dataset used was obtained directly from the research team at Strava through a data sharing agreement. Data focuses on all active users located in two major metropolitan areas – San Francisco, CA and Boston, MA. A geographic-based sampling strategy was

chosen to allow for geographic comparison, in this case East versus West coast of the United States. San Francisco and Boston are very similar metro areas in terms of their mix of jobs, though separated by over 2500 miles (Times, 2018). They also have similar land area and population. Boston is slightly less dense and younger in terms of media age. San Francisco, CA is where Strava started its business; then the company has gained increasing popularity in the U.S. East coast and globally. Both locations have a large set of active Strava users; in fact, coastal regions and metropolitan areas have much higher rates of usage than elsewhere¹. While sampling may limit the generalizability of any geographic-specific findings, this dataset enables us to take the first step to study potential patterns of personal network structure and how these patterns are consistent or different between these two metro areas. In the remainder of this paper, we refer to these datasets as the CA and MA samples.

Compared to public Strava data available on the Strava Application Programming Interface (API), this collection includes more details for anonymized Strava users (e.g. age and number of years on the service). Importantly, this data records actions of personal networks' tie formation for each user and historical logs of their posted activities. The data consists of but is not limited to the following three major components: (1) user profile information, such as gender, age, location, sign-up date, enrolled plan (i.e. free or paid plan); (2) network subscription records: who followed whom and the time-stamp at which the connection was made; (3) detailed representation of activities posted by users over time, e.g. activity description, activity performance, etc.

Table 5.1 presents a descriptive summary of the dataset used in this work. Since Strava started and continues to be headquartered in CA, it is unsurprising that athletes in CA adopted this new fitness application earlier; CA enjoys a larger body of users. Strava has a much higher proportion of male users, reflected in both location-specific datasets. Around 40% of users reported their age information. Among them the largest age group is the 30-39 group for both locations.

¹<https://www.strava.com/heatmap#2.29/-45.89087/0.00000/hot/all>

Table 5.1: Descriptive Statistics of Strava Dataset

	California	Massachusetts
# of Users	46,181	23,477
Gender Proportion		
Male	0.757	0.744
Female	0.243	0.256
Age Proportion		
18-29	0.215	0.235
30-39	0.449	0.369
40-49	0.236	0.243
50-60	0.100	0.153
Enrolled Plan Proportion		
Free	0.923	0.927
Paid	0.077	0.073

5.5 Methods

5.5.1 Degree Distribution Models

To characterize and understand the process by which users intentionally build their personal networks in online fitness communities, we focus on the distribution of network size across the population of users in the study. The degree of a node in a social network is the number of ties attached to it. The degree distribution then characterizes the frequency of degree among network participants and is one of the fundamental of network properties (Newman, 2018). While degree distributions do not give complete information about network structure, their form and features communicate very important information, including possible mechanisms by which ties could have been constructed. Modeling degree distributions is an area of ongoing work across many domains in the field of social network analysis. Most

networks in the real world have highly skewed degree distributions (Newman, 2018). In particular, a wide range of networks approximately follow a power-law degree distribution including but not limited to, the world wide web, protein-protein interaction and some social networks (Newman, 2003; Barabási and Bonabeau, 2003; Eagle et al., 2009). Such networks are referred to as scale-free networks (Barabási and Bonabeau, 2003). This “scale-free” property implicitly implies that there exists underlying stochastic mechanisms for the formation of network structure.

Power-law degree distributions have been observed in many cases, however, they are not the only form observed in real-world social networks (Clauset et al., 2009). There are many theoretical models for degree distributions and each one implies a type of mechanism of the formation of networks. Thus, these models are important statistical tools for analyzing and understanding the structure and formation of social relationships. In this work, we employ the methods developed by Handcock and Jones (Handcock and Jones, 2004). Their work discusses three general classes of stochastic process models for network formation: (1) non-homogeneous Poisson; (2) preferential attachment; and (3) “vetting” models. While originally developed to capture potential mechanisms for choice of sexual partners, these models are general partner choice models. In particular, they are likely to apply in cases where individuals have to make decisions about which relationships to invest in, given a limited capacity for investment. In the case of online fitness communities, individuals may choose partners on the basis of social support, informational resources, potential for co-presence and social exercise, etc. In each case, users are forming social ties in a deliberate and intentional way. Next, we briefly introduce these three types of degree distribution models. An expanded discussion of the models can be found in (Handcock and Jones, 2004). All models used here have been implemented in the *R* package *degreenet* (Handcock, 2003).

Non-homogeneous Poisson: Consider the population K of individuals with at least one tie in a given time period. Suppose that the number of additional ties $K - 1$ that the person has in the time period follows a Poisson distribution with expected value λ . This process (also referred to as homogeneous Poisson process) often fails to describe data

realistically because it assumes all individuals have identical propensities to form ties in networks. Thus, more general non-homogeneous models, or extensions thereof, which account for within-population heterogeneity are more commonly used to study real-world problems (e.g. Sakaki et al. (2010); Mahmud et al. (2013)). For example, if the structural mechanism differs by user demographics or socioeconomic status, non-homogeneous Poisson process is a better choice than homogeneous Poisson process.

As with Handcock and Jones (2004), we model within population heterogeneity by representing individual expected values of λ as independent draws from a Gamma distribution with mean μ and standard deviation σ . The resulting distribution can have an extreme long tail but finite variance. It is called *shifted negative binomial distribution* in Handcock and Jones (2004).

Preferential Attachment: Preferential attachment indicates a process in which new nodes prefer to attach to well-connected nodes over less-well connected nodes. This process is also referred to under the names “cumulative advantages”, the “Matthew effect” and more colloquially, “the rich get richer.” Handcock and Jones (2004) proposes two distribution models that result from preferential attachment processes; they are the Yule distribution (Simon, 1955) and the Waring distribution (Irwin et al., 1963).

Beginning with a network of r connections, the Yule distribution is motivated by assuming that (1) there is a constant probability $\frac{\rho-2}{\rho-1}$ that the $r + 1$ st relationship in the population will be initiated from a randomly chosen person to a previously inactive person, and (2) otherwise the probability that $r+1$ st relationship will be to a person with exactly k relations is proportional to $k \cdot f(k|r)$. Here, $f(k|r)$ is the frequency of persons with exactly k connections out of the r total connections in the population (Simon, 1955). The Yule distribution will exhibit power-law behavior.

The Waring distribution is motivated similarly as the Yule distribution, but in this case the constant probability that the $r + 1$ st relationship in the population will be initiated from a randomly chosen person to a previously inactive person is $\frac{\rho-2}{\rho+\alpha-1}$. The Waring distribution is a natural generalization of the Yule distribution (Irwin et al., 1963).

Vetting Models: This class of models involve a two-stage process. In general, the process first generates a list of potential contacts and then forms relationship based on this list. The vetting models are very flexible, allowing to specify any distribution for both of the two processes. In this work, we consider the Yule-vetting models (see Handcock and Jones (2004) for more detail).

For the Yule-vetting models, the process first generates a number of A potential contacts from a distribution $P(A = a)$ which may exhibit power-law behavior. Next, the process assumes the potential number of friends L that a person has at the time follows a distribution that is short-tailed in most cases (e.g. geometric or negative binomial distribution). However, the *actual* number of friends that a person has K is bounded by the number of potential contacts A that a person has - that is $K = \min(A, L)$. One example of Yule-vetting models is the negative binomial Yule. The number of potential contacts are formed from a Yule distribution and the potential friendship distribution is negative binomial.

5.5.2 Estimating Model Parameters

We adopt a likelihood framework to estimate model parameters and to compare between candidate models. There are substantial differences when the process fits low and high network degree. Therefore, the proposed likelihood model includes extra parameters to fit low-degree observations. Here we apply non-parametric fitting and standard model selection techniques.

Denote y as the data - a vector of degree counts and define k as the threshold above which the parametric model is fit, π as a vector of lower-tail probabilities for the non-parametric component, and p as a vector of upper tail probabilities for the parametric component which is $p_i = Pr(d = i | d > k_{min})$.

The log-likelihood of the data y is:

$$\log Pr(y|k_{min}, \pi, p) = \sum_{i=0}^{k_{min}} y_i \log \pi_i + \left(N - \sum_{i=0}^{k_{min}} y_i\right) \log \left(1 - \sum_{i=0}^{k_{min}} \pi_i\right) + \sum_{i=k_{min}+1}^{\infty} y_i \log p_i$$

5.5.3 Model Selection

We select the best fitted model among a set of candidate models implying certain types of social mechanisms of friend formation as discussed above. A reasonable model selection technique will balance goodness of fit with model complexity. Here we employ the model selection process stated in the work (Handcock and Jones, 2004; Almquist and Butts, 2014). Due to the finite nature of social networks (Burnham and Anderson, 2003), we use the Akaike information criterion with a correction for finite sample sizes (AICc) as the indicator of the best performing models (Hurvich and Tsai, 1989). For convenience, BIC (Kass and Wasserman, 1995) is also provided in later results. A model with smaller AICc or BIC is judged better than another model. Under the assumptions of MLE, the AICc is given by: $AICc = AIC + \frac{2p(p+1)}{n-p-1}$, where n is the number of observations and p is the number of model parameters.

5.6 Results

5.6.1 Empirical Egocentric Network Characteristics

Table 5.2 shows descriptive statistics for the personal networks of users. Personal network size is closely related to the length of time since an individual joined the Strava platform. As seen in the Table 5.2, men consistently have longer tenure on the site than women. Therefore, rather than comparing the raw counts of incoming and outgoing ties, we divide an individual count of both types of ties by a user's tenure and compare the resulting normalized numbers by gender, age, location and plan enrolled groups.

First, we compare network characteristics between free-plan and paid-plan users. It is not surprising to see that paid-plan users, on average, have a much greater number of network

Table 5.2: Descriptive Statistics of Personal Networks on Strava

California							
Men				Women			
	Tenure	Avg. $\frac{\text{In-Degree}}{\text{Tenure}}$	Avg. $\frac{\text{Out-Degree}}{\text{Tenure}}$		Tenure	Avg. $\frac{\text{In-Degree}}{\text{Tenure}}$	Avg. $\frac{\text{Out-Degree}}{\text{Tenure}}$
Plan				Plan			
Free	27.818	0.076	0.358	Free	26.219	0.098	0.323
Paid	50.821	0.404	1.187	Paid	45.331	0.590	1.369
Age				Age			
18-29	29.049	0.174	0.702	18-29	26.512	0.242	0.614
30-39	39.790	0.194	0.615	30-39	36.646	0.217	0.559
40-49	44.979	0.155	0.575	40-49	40.168	0.137	0.498
50-60	39.100	0.096	0.418	50-60	36.388	0.060	0.242
Massachusetts							
Men				Women			
	Tenure	Avg. $\frac{\text{In-Degree}}{\text{Tenure}}$	Avg. $\frac{\text{Out-Degree}}{\text{Tenure}}$		Tenure	Avg. $\frac{\text{In-Degree}}{\text{Tenure}}$	Avg. $\frac{\text{Out-Degree}}{\text{Tenure}}$
Plan				Plan			
Free	27.827	0.100	0.309	Free	22.468	0.107	0.324
Paid	43.661	0.376	1.021	Paid	33.790	0.485	1.382
Age				Age			
18-29	29.414	0.176	0.554	18-29	24.313	0.216	0.590
30-39	37.698	0.173	0.527	30-39	31.447	0.172	0.498
40-49	40.640	0.159	0.465	40-49	31.597	0.165	0.460
50-60	39.536	0.167	0.412	50-60	31.332	0.090	0.306

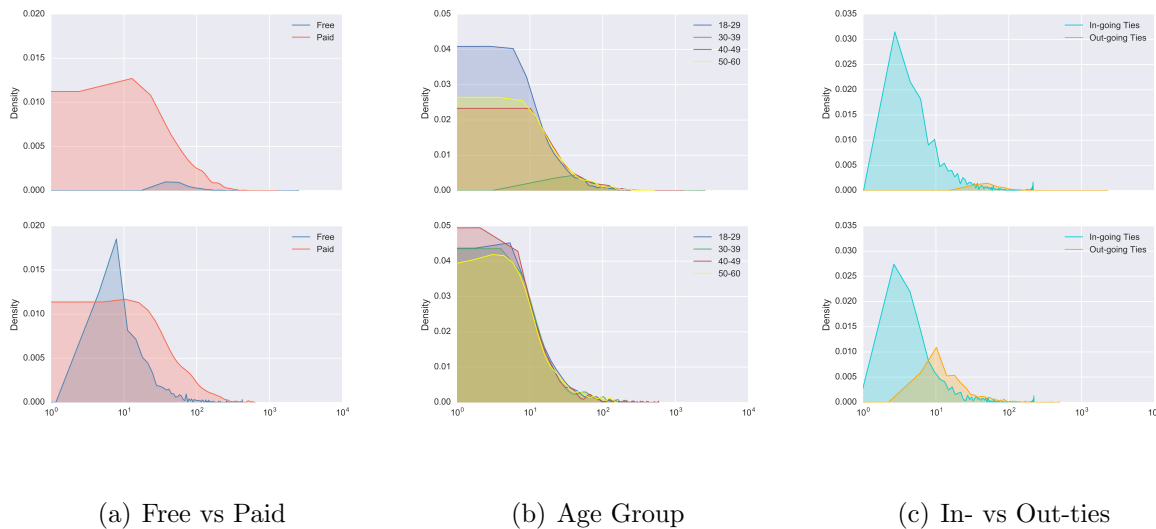


Figure 5.1: Network Degree Distributions: subfigures in the first and the second row show degree distributions of male and female personal networks, respectively.

connections than free users do. However, it is interesting to observe that both average incoming and outgoing ties for women who enrolled in paid plans are greater than paid-plan men – the pattern is consistent for both locations. Moreover, in the MA data women with free plan also tend to have larger in- and out-degree than counterpart men.

Recall that only about 40% of users report age information in their profiles. Therefore, the network statistics by age group in the Table 5.2 are obtained based on the personal networks whose age information is available. We find that a large gender difference exists in the group 50-60 years of age for both CA and MA; men tend to have larger counts of both in-coming and out-going ties than women do. For the mid-aged groups (e.g. 30-39 and 40-49), we see less substantial gender differences in averages of both in- and out-degree.

These results indicate gender and age differences in personal network structure. This may imply that underlying mechanisms of relationship formation differ by plan enrolled and demographics.

Further, Figure 5.1 shows the degree distributions for egocentric networks. Specifically,

we compare distributions of network size (i.e. total degree) along gender, age and economic lines; we also compare how in- and out-degree are distributed by gender. Three sub-figures in the first row present degree distributions of male personal networks, whereas the second row represents that of female personal networks.

Figure 5.1(a) compares the distributions of personal network size for free-plan and paid-plan users by gender. Consistent with the results in the Table 5.2, paid-plan users on average have larger networks than free-plan users. Interestingly, we see a salient gender difference in network size among free-plan users; overall, female free-plan users have smaller network size than counterpart males. Figure 5.1(b) shows the distributions of network size for users in different age groups. We observe that the degree distribution varies most by gender for the age group 30-39. Figure 5.1(c) presents the in- and out-degree distributions for male and female users. The average in-coming degrees for both men and women are lower than the average out-going degrees. Overall, in-degrees of female users' networks are slightly greater than male users'. But this gender difference is not quite noticeable as in the out-degree distributions. The out-degree distributions present a gender difference that women on average have fewer out-going ties than men do. Note that here we only show the figures based on the MA dataset due to space constraints, but the patterns remain consistent across two areas.

5.6.2 Modeling Degree Distributions

Table 5.3 and Table 5.4 present model selection results for the degree distribution modeling. We consider all stochastic models discussed in the Section 5. Here we show only the best fit model for each user group in both tables, but our model selection process went up to 10 for the non-parametric component.

Table 5.3 shows the most “likely” degree distribution model for users in free and paid plan; we also model men and women separately. For in-degree distributions which can be viewed as number of followers, we find that Strava users with free plans have an in-degree distribution most consistent with a Negative Binomial Yule distribution, while users with paid plans are best fit by a Negative Binomial distribution. This suggests that free-plan users

Table 5.3: The best fit models for in- and out-degree distributions of male and female users who enrolled in free plan versus paid plan in Strava.

In-Degree Distribution Models							
Plan	Gender	Model	k	np	log-lik	AICc	BIC
Free	M	NegBinYule	1	5	-19712.35	39434.7	39471.66
	F	NegBinYule	2	5	-6865.921	13741.86	13773.78
Paid	M	NegBinom	1	3	-5318.866	10643.75	10659.56
	F	NegBinom	1	3	-860.3571	1726.815	1737.193

Out-Degree Distribution Models							
Plan	Gender	Model	k	np	log-lik	AICc	BIC
Free	M	NegBinom	1	3	-27185.63	54377.27	54399.45
	F	NegBinom	1	3	-9205.993	18417.99	18437.15
Paid	M	NegBinom	1	3	-6706.873	13419.76	13435.57
	F	NegBinom	1	3	-1091.877	2189.855	2200.234

Table 5.4: The best fit models for in- and out-degree distributions of male and female users in different age groups in Strava.

In-Degree Distribution Models							
Age	Gender	Model	k	np	log-lik	AICc	BIC
18-29	M	Waring	1	3	-3157.574	6321.164	6336.88
	F	Waring	1	3	-1358.552	2723.145	2736.28
30-39	M	NegBinom	1	3	-6127.021	12260.05	12277.39
	F	Waring	1	3	-1793.552	3593.136	3606.98
40-49	M	NegBinom	1	3	-4400.432	8806.879	8823.196
	F	Waring	1	3	-770.2596	1546.587	1558.177
50-60	M	NegBinom	1	3	-2719.809	5445.64	5460.594
	F	Yule	2	3	-385.2882	776.689	786.716

Out-Degree Distribution Models							
Age	Gender	Model	k	np	log-lik	AICc	BIC
18-29	M	NegBinom	1	3	-4330.269	8666.556	8682.272
	F	Waring	1	3	-1759.673	3525.386	3538.521
30-39	M	NegBinom	1	3	-8147.217	16300.44	16317.78
	F	NegBinom	1	3	-2294.011	4594.053	4607.898
40-49	M	NegBinom	1	3	-5882.915	11771.84	11788.16
	F	Waring	1	3	-1049.921	2105.91	2117.501
50-60	M	NegBinom	1	3	-3426.347	6858.716	6873.67
	F	NegBinom	1	3	-528.0239	1062.16	1072.188

are employing a vetting process in which first there is a list of potential users who may become ego’s followers and then people in the list have equal chance to follow ego. On the other hand, paid-plan users are employing a non-homogeneous Poisson process. One interpretation in this context is that each user in the paid-plan group has a constant probability of being followed by other users, but this probability varies by users. This process would be consistent with an environment in which athletes have different levels of popularity or desirability to be targets of attention. The results of out-degree distribution modeling indicate that users with both free and paid plans are employing a non-homogeneous Poisson process when following other users. These results do not show substantial gender differences in each plan group.

Table 5.4 shows the best fit for personal networks of Strava users categorized into four age groups. Except for the 18-29 years of age group, we find that the formation of in-coming relations for men most consistent with a Negative Binomial distribution and that female’s in-degree formation most likely follows a preferential attachment process (e.g. the Yule and Waring distribution). We observe some age differences in out-degree distribution modeling. While men across all age groups employ a Negative Binomial process, women in the age group of 18-29 and 40-49 employ a preferential attachment process. All best-fitted degree distribution models have very low cut-offs (i.e. $k = 1$ and 2 as seen in the Table 5.3 and 5.4). Therefore, these models exhibit distinct processes at low degree; users with just one or two followers may be new or only recently active. They might also be interested in self tracking more than social networking features of the platform.

5.7 Discussion

Previous literature points out that women and men often have distinct social networks (Granovetter, 1977; Centola and Macy, 2007). These documented differences have important implications for information access, social support and health. Utilizing large-scale behavioral trace data, our work presents empirical evidence for variability in network size across users of the online fitness community Strava; importantly these differences are associated with demographic and economic categories. Men, users who pay for the service, and 30-49

year old users tend to have larger networks, i.e. more peers with whom they engage on the platform. Women, along with the youngest and oldest users have smaller personal networks – fewer individuals to turn to for social support and motivation in reaching fitness goals. These results are not just for a small sample of participants, but instead represent large-scale analyses of users in two metro areas on both coasts of the United States. Beyond characterizing empirical patterns in network size, this study also explores the mechanisms of the observed empirical distribution of network size in the population.

This investigation reinforces longstanding empirical patterns of inequality in network positions. Strava users have different opportunities for social interaction. They make different choices about which relationships to initiate, invest in, and pursue. These choices, in turn, have important consequences for receiving social support and resources, looking to others for motivation or reinforcement in achieving goals, and long term engagement versus drop out. Our findings suggest that men and women, for example, may approach choices about social interaction and investment in social relationships in a fitness setting in different ways. Initiating relationships based on popularity (e.g. a preferential attachment process) versus a two-stage vetting process implies different underlying process considerations. Not only does this imply inequality in terms of network position, but it also suggests behavioral differences in the ways in which individuals make use of the platform itself. Further explorations, perhaps participant interview for example, might offer insight into the approaches that Strava users consider when constructing social networks to support fitness goals.

Previous literature on homophily and social support suggests that the individuals with certain health conditions (e.g. obesity) or with lower social status may be more dependent on the composition of their social networks than healthier or high-status individuals (Centola, 2011; Rogers, 2010). Our work suggests that network-based interventions for health should take structural differences, and their associated implications about mechanisms for tie formation, into account. In fact, accounting for differences might be especially vital for people who are most in need of support to achieve healthier behavior. Moreover, this study points to the need for a better understanding of the dynamics of relationship construction and maintenance.

For new participants to an online fitness community, with whom are their first relationships forged? Are they with friends from offline settings? Are they with high status and high visibility professional athletes? Do they look for others like themselves, in terms of social and demographic characteristics? Answering these questions could not only improve the design of these systems and their affordances, but they could also have important implications for health behavior change.

Previous work suggests that social motives are vital for persistent use of activity-tracking platforms (Stragier et al., 2016; Consolvo et al., 2006). For current users of online fitness communities, stronger social motives and peer support may motivate them to stay with the fitness service for a longer time and be more likely to develop long-term healthy habits. Being embedded in an active and supportive network of peers within online fitness communities may be related to persistence of use. Understanding the processes individual platform users behave in accordance with may inform the design of tools and/or systems to support healthy personal networks. In fact, users may not behave in the ways that would lead to an optimally supportive mix of personal contacts. In such cases, platforms themselves should be able to suggest missing connections or social contacts that can support certain goals. Many of these important questions could be addressed in a follow-up study, perhaps one that also compiles participant interview or survey responses to explore motivations behind specific relationship construction choices.

5.8 Conclusion

Modeling the distribution of personal networks in this online community demonstrates that individuals are embedded such that they experience systematic differences in the social environment based on demographic and economic factors such as age groups, genders, and economic brackets on Strava and likely other activity based platforms. These differences have numerous implications for network-based interventions that leverage peer influence, social support, and other applications in efforts to change behavior in this setting.

The curve fits of network degree distributions presented in this study are just the first

step in exploring and understanding the social dynamics at play in these new social settings. Future research should directly explore the stochastic mechanisms which generate empirical networks in these settings through complex dynamic models or careful observational studies and/or follow up surveys.

<<END OF PUBLICATION>>

Chapter 6

SOCIAL EXERCISE IN ONLINE FITNESS COMMUNITIES

6.1 *Author Preface*

Previous chapters discuss how online interaction and personal network structure relate to offline physical activities. In this chapter, we examine how online connections may interact with offline physical activities in online fitness communities. Specifically, we look at social exercise - fitness activities with at least one Strava peer physically co-present. Some existing studies demonstrate that peer influence and social support have positive health-related effects. However, most of these studies are based on surveys or experiments which has limitations in terms of scale and validity. Therefore, we use a large-scale dataset collected from Strava API to analyze how exercise performance may vary by the number of people participating in an exercise and whether gender plays a role in these effects. The findings demonstrate that social exercise tends to be more intense and last for a longer time. Gender difference is also observed - women and men have their preferences to choose the gender and the number of peers they would love to exercise with.

This work is published in the iConference and selected as Lee Dirks Best Paper Award Finalist.

Reference: Zeng, L., Almquist, Z.W. and Spiro, E.S., 2017. Let's Workout! Exploring Social Exercise in an Online Fitness Community. iConference 2017 Proceedings Vol. 2.

<<THE PUBLISHED PAPER BEGINS BELOW>>

Abstract

Increasing attention has been paid to promoting certain healthy habits through social interaction in online communities. At the intersection of social media and activity tracking applications, these platforms capture information on physical activities as well as peer-to-peer interactions. Importantly, they also offer researchers a novel opportunity to understand health behaviors by utilizing the large-scale behavioral trace data they archive. In this study we explore the characteristics and dynamics of social exercise (i.e. fitness activities with at least one peer physically co-present) using data collected from an online fitness community popular with cyclists and runners. In particular, we ask if factors such as temporal seasonality, activity performance and social feedback vary by the number of people participating in an activity; we do so by comparing associations for both men and women. Our results indicate that when peers are physically co-present for fitness activities (i.e. group workouts), exercise tends to be more intense and receive more feedback from other users, across both genders. Findings also suggest gender differences in the observed tendency to complete activities with others. These results have important implications for health and wellness interventions.

Keywords: social exercise; health behaviors; online fitness communities; social media; behavioral traces

6.2 Introduction

Exercising is known to be associated with numerous physical and mental benefits such as controlling weight (Blair, 1993), decreasing the risks of cardiovascular diseases and reducing stress (Fletcher et al., 1996). Although aware that exercising is good for health, not everyone engages in physical activity on a regular basis. The Center for Disease Control in the United States estimates that only 20% of adults meet exercise guidelines (Jaslow, 2013).¹ Indeed, the World Health Organization continues to combat the global obesity epidemic (WHO, 2000). Health, and more specifically physical activity, is a complex issue and as

¹<http://www.cdc.gov/nchs/fastats/exercise.htm>

such has been studied in many fields. Beyond the medical community, issues of health are studied in the social sciences, where research has examined how social, racial, emotional, and socioeconomic factors influence health promotion and likewise health disparities (Walker et al., 1987; Frankish et al., 1998; Lee et al., 2014; Dishman et al., 1985).

A number of studies demonstrate that peer influence and social support have positive health-related effects, such as helping people lose weight and participate in more physical activities (Ahtinen et al., 2009; Chen and Pu, 2014; Wing and Jeffery, 1999; Kulik and Mahler, 1989; Dishman et al., 1985). Prior research also points out that exercising with others can improve psychological functioning (Plante et al., 2001). However, many of these early studies of peer effects are based on surveys or experiments involving a small number of participants. Recently, social media and online fitness communities are gaining scholarly attention as a new research environment in which to study health behaviors (Centola, 2013).

Social media have been used to explore health communication and promotion (Paul and Dredze, 2011; Morris et al., 2011; Vaterlaus et al., 2015; Pechmann et al., 2015) and physical activity (Teodoro and Naaman, 2013; Munson et al., 2013). Studies have even used randomized experiments to establish causal peer influence effects (Zhang et al., 2015). Online fitness communities offer even more promising directions for work in this area (Centola and van de Rijt, 2015). These new platforms are specifically designed to provide participants with a group of peers and social support in reaching their fitness or health goals. Moreover, users are able to use wearable devices to track personal activities, including exact Global Positioning System (GPS) traces of routes, and upload them to their online profiles within in the community.

For researchers, an important feature of online platforms is their ability to archive a large volume of behavioral trace data, including fitness statistics, user profile data, and potentially users' social networks, as part of their normal operation. This enables users to explore and compare their own activity efforts to others in their "fitness circles", but it also presents novel opportunities to analyze health behaviors. Researchers now have the opportunity to

observe not what people *say* they do, but what they actually do.²

In this study, we employ behavioral trace data from one such fitness community to study social exercise. In particular, this work analyzes how factors including gender, temporal seasonality, activity performance and social feedback may vary by number of people participating in exercise. Our work aims to answer the following research questions: (1) *When do individuals choose to exercise alone and in groups, and are these dynamics gender dependent?* (2) *How do peoples' fitness behaviors differ, in terms of performance and social feedback, when exercising alone compared to activities with others physically co-present?* In answering these questions, this study has implications for health promotion and social network-based health interventions.

6.3 Related Work

6.3.1 Social Support and Physical Activity

Previous work suggests that factors such as pleasant surroundings, an enthusiastic exercise leader, and sympathetic co-exercisers during leisure-time activities are all likely to relieve negative emotions associated with exercise (Haskell et al., 1985; Pelphrey et al., 2003; Flaherty, 2005). A laboratory-based study found that exercising with others helped to reduce stress and produce overall positive effects on energy, calmness and tiredness, compared with a control group exercising alone (Plante et al., 2001). However, this experiment was conducted in a laboratory setting and participants (recruited from a college student population) were fairly homogeneous in terms of age and fitness levels, making its applicability in real-world settings a open question. Despite limitations of prior studies, it is well-established that social support and physical activity are linked (Dishman et al., 1985; McAuley et al., 2000; Berkman and Glass, 2000).

More recently, the relationship between social support and physical activity has been studied using mobile fitness applications (Munson and Consolvo, 2012). Some new platforms

²This is not to say behavior trace data from online fitness communities is not without limitations (as we discuss at the end of this paper).

explicitly include the element of social support, allowing users to exercise in a virtual group environment so as to motivate them to perform physical activities (Campbell et al., 2008; Consolvo et al., 2006; Chen and Pu, 2014). For example, Chen and Pu (2014) designed a mobile application with gamification settings of competition (i.e. two users compete to gain more virtual rewards by exercising), cooperation (i.e. two users contribute equally to win virtual rewards) and hybrid (i.e. weighting the cooperation and competition settings). Even though users are not required to exercise side by side in the physical environment, all three conditions of virtual group activities were found to increase users' activity frequency and intensity.

6.3.2 Online Fitness Communities

The past few years have seen an explosion of new online fitness communities (e.g. RunKeeper, MapMyRun, Strava, etc.) where users' natural, everyday activity can be tracked and explored with a rapidly expanding collection of tools and technologies (Centola, 2013). These fitness communities sit at the intersection of social media and activity tracking applications; users can not only track/log their activities, but also interact with a group of peers and posted activities. One example of such a platform is Strava. Promotional content on the site's home page³ says: "The social network for athletes. Connect with friends and make the most of every run and ride." The proliferation of similar platforms, coupled with recent research indicating that just over 20% of adults use some form of technology to track their health data (Fox, 2011), signals new opportunities for understanding the social dimensions of health-oriented behaviors. In particular, behavioral traces of human behavior and interaction collected from these online sources offer novel data and strategies for understanding social dynamics and peer influence.

Online fitness communities have attracted researchers from many disciplines. Some scholars are interested in the technical potential of sensors and human-computer interaction

³Strava homepage is at <https://www.strava.com/>

aspects of these technologies (Consolvo et al., 2008), others have focused on play, incentives, and user engagement (Chen and Pu, 2014). A growing body of work concerns social media as a potential tool in medicine (Centola, 2013). All of these approaches promise insight into the social aspects of health, however, limited work has specifically explored the effects of peer co-presence, leaving a gap in our current understanding of the social dynamics in these settings.

6.4 Data

Data were collected for the online fitness community Strava. Strava continues to grow in popularity among cyclists and runners in recent years. Strava provides two main competitive, gamifying features to motivate users to reach their fitness goals. The one is the ability to compare users' activity efforts against their history efforts or compete with other athletes. Another important feature is to accomplish challenges and earn achievement badges. As such, millions of people upload their rides and runs to Strava every week via their smartphones or GPS devices.

Strava was also chosen as the research environment because it has a number of desirable characteristics: (1) it attracts an increasing number of users around the world who upload millions of activities to the platform every week; (2) the Strava Application Programming Interface (API)⁴ provides access to public Strava behavioral trace data; and (3) data include activity characteristics, user characteristics and user social network characteristics, providing rich data for research that seeks to understand health behaviors.

To retrieve data about an activity posted to a user's profile on Strava, a valid `activityID` is required. We randomly generate a list of `activityIDs` from a previously-built (and ideally exhaustive) ID space. We then query each `activityID` so as to check if this ID exists. If that `activityID` does not exist on the platform, we discard this one. If it exists, we retrieve the data associated with that `activityID` (i.e. information and metadata about a particular

⁴Strava V3 API Documentation is at <https://strava.github.io/api/>

activity). As each activity is associated with a particular user, this data includes a summary representation of the posting athlete which allows us to query for a detailed representation of that user. Study procedures were reviewed by the Institutional Review Board at the authors' university.

As stated previously that Strava API provides access to its rich metadata, our data include but not limit to the following main components: (1) activity characteristics such as activity type (e.g. cycling, running, swimming, etc.), activity location, activity names, activity-related stats (e.g. distance, moving time, elevation, etc.); (2) user characteristics such as user demographics, user location, physiological measurements (e.g. height, weight, etc.), equipment; (3) social interaction characteristics such as following and followed relationship, comments, kudos (or "likes").

We collected 888,093 sampled activities posted during 2011-2016 from 514,362 unique users. 81.33% of users report their gender as male. 14.64% of users report their gender as female and 4.03% of users' self-reported gender are unknown. 93.98% of data represent rides (cycling) and runs, but data also represents workout types such as swimming, walking, hiking, skiing, etc. It is important to note that this data does not represent a random sample of platform users, but is instead more likely to capture the behaviors of highly active users. While this limits the generalizability of any user-specific findings, it also means our analysis is conducted on regular users who are more likely to fully utilize platform features. This is important for this case, because many of our research questions focus on group activities - fitness events where users specify in the application that they are physically co-present with others. Moreover, we do obtain a random sample of activities, and much of the following analysis uses the activity itself as the unit of analysis.

6.5 Methods

To explore the characteristics and dynamics of solo and group exercise, we begin by classifying observed data based on the number of peers co-present for the activity. Each activity record has an attribute that represents the number of athletes/users taking part in

that activity. We identify activities involving only one athlete as *solo activities* and activities having more than one athletes as *group activities*. Our data have 696,856 (78.47%) solo activities and 191,237 (21.53%) group activities. In the analysis that follows, we evaluate the characteristics and dynamics of each type of activity, comparing solo exercise against group exercise. We do so for men and women separately, in order to tease apart any gender differences in these results. Comparing between genders is also important because the Strava platform is heavily male dominated and prior work has noted that individuals within this environment may have different experiences and social networks (Spiro and Almquist, 2016).

When exploring the temporal dynamics of exercise, we consider: (1) the day of the week (i.e. Monday to Sunday) and (2) the hour of the day (i.e. 0:00 to 23:00) during which activities are observed. Each activity has a local timestamp indicating the start time of the event. An analysis of temporal features could help understand the seasonal patterns of solo and group activities. We hypothesize that people prefer to do more solo activities on weekdays because it is possibly harder to coordinate time among multiple people for group activities during these days. Hence, we also hypothesize that group activities take place more frequently on weekends. We also want to find out the “busy” time periods for solo activities and group activities.

Next, we want to understand how performance, that is physical exertion, is related to peer co-presence. Prior work on peer influence and social support suggests that peer co-presence motivates individuals to engage in physical activity more regularly and more often, but it might also motivate them to work harder, exerting themselves more throughout their physical activity. Measuring performance is challenging and likely involves a multifaceted approach. Strava also applies a diversity of measurements for physical activities. Therefore we consider five distinct measures to operationalize activity performance. These include:

- a. Distance: total distance of an activity (available on all data)
- b. Elevation: total elevation gain of an activity (available on all data)

- c. Duration: total moving time of an activity (available on all data)
- d. Activity effort: average watts (available on cycling activities - around 60% data) and average speed (available on all data)⁵
- e. Physical challenge level: Strava provides its users with a computed “suffer score” which it calculates based on estimated heart rate intensity (available on premium users’ activities - around 10.5% of data)⁶

Finally, our work aims to compare social feedback for activities that are complete solo versus in groups. Strava users can post comments and kudos (i.e. “likes” or +1s) on an activity. Related work suggests that these social interaction functions have motivating effects on physical activities (Chen and Pu, 2014). Hence, we are interested in analyzing whether workouts as a group receive more comments and/or kudos from activity participants or other Strava users.

6.6 Results

In this section, we present findings to address the research questions outlined previously. First, we discuss overall gender differences in posting behavior and exercising alone or with a group. We then continue, considering the temporal dynamics of when individuals choose to exercise alone and in groups, and how these dynamics are gender dependent. Next we evaluate how fitness behaviors differ, in terms of performance and social feedback, when individuals exercise alone compared to when they exercise with others physically co-present. Again, we consider the differential effect of these activity features by gender.

⁵Average watts measures the rate of energy conversion with respect to time. Since it is available on partial data, we use the measure of average speed that is available on activities of other types including running, hiking so as to avoid possible biased sampling merely from rides.

⁶Solo activities and group activities account for 76.25% and 23.75% of premium users’ activities, respectively, roughly matching the proportions in the entire data sample. Therefore, we believe that the measure of `suffer score` is still representative and not likely to introduce large bias to the analysis.

6.6.1 Gender Differences in Activity

	Activity counts	Percentage
Female Alone	91,248	0.768
Female with Others	27,576	0.232
Male Alone	580,832	0.786
Male with Others	158,448	0.214

Table 6.1: Counts and Percentages of Solo/Group Activities by Gender

Strava is a male-dominated platform, where the large majority of platform users are males, and this is reflected in our dataset. To begin our analysis, we count the number of activities of both types for each gender, and calculate the corresponding percentages. Table 6.1 shows a greater percentage of activities posted by females are group activities than among males, indicating that there are proportionally more women who are involved in group activities.⁷

Figure 6.1 shows how each gender is observed to participate in activities that involve a specified number of participants. For example, in the Figure 6.1(a), the largest rectangle represents the proportion of activities posted that include only one user across gender. The width (0.7857) and height (0.7679) of the rectangle correspond to the proportion for men and women, respectively. Since, in this case, the width is greater than the height, we know that solo activities enjoy a larger proportion among men than among women. However, for activities having two participants, we find a much larger proportion among women, indicating exercising with a single peer is observed more frequently among women. As the number of activity participants increases, rectangles gradually deviate from the dashed line, indicating that female users tend to engage in more activities in smaller groups (mostly in a group of size two); as participants continue to increase, we find that the trajectory of the rectangle

⁷A chi-squared test was performed to determine if significant difference exists in these counts. Unsurprisingly, given the dataset size, we find a highly significant ($p < 0.001$) relationship between gender and exercising alone versus with others.

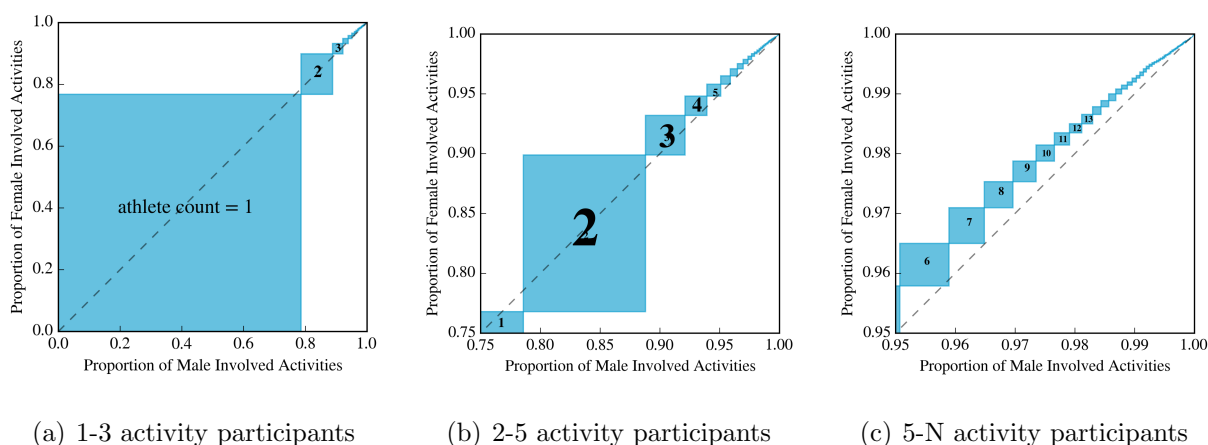


Figure 6.1: Proportion of different athlete counts by gender. Each rectangle visualizes the proportion of activities with a specified athlete count. The width and height of a rectangle indicate the proportion of men’s activities and the proportion of women’s activities respectively with that count. If the upper right vertex of a rectangle falls on the dashed line ($y = x$), this indicates the cumulative women’s proportion is equal to cumulative men’s proportion at that point.

position goes back, closer to the dashed line, indicating that men (in proportion) are involved in larger group activities.

6.6.2 Temporal Dynamics of Activity

One of our primary research questions aims to identify seasonality patterns for solo and group exercise. Figure 6.2 shows the proportion of solo and group activities posted to Strava across the week. We observe that proportions of group activities are smaller on weekdays compared to that of solo activities. Moreover, about 50% group activities occur over weekends, greatly exceeding the proportion of solo activities. For solo activities, we see a relatively stable and consistent pattern throughout the week. Exercising on Mondays and Fridays appears to be less attractive to athletes, as fewer activities - both solo and group activities - are posted on these days.

We further examine the seasonality patterns of solo and group activities in terms of

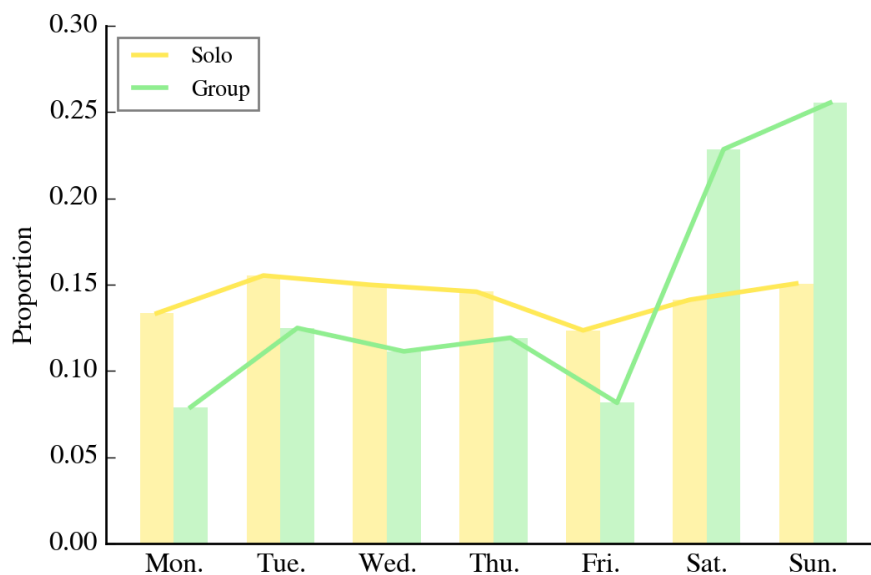
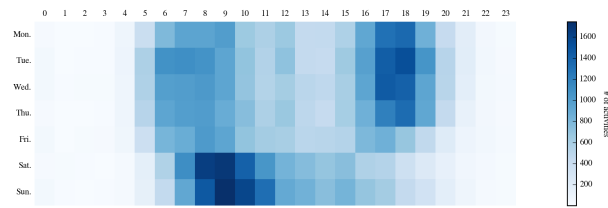


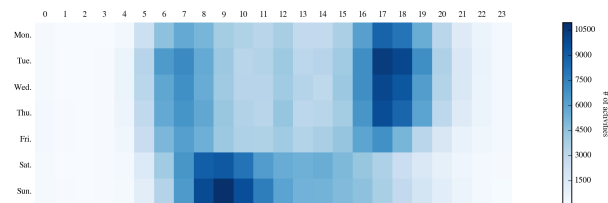
Figure 6.2: Distribution of solo (yellow) and group (green) activities across the seven weekdays. Corresponding lines map trends over the week.

the hour of the day and the day of the week. Figure 6.3 visualizes solo/group activities occurring during a specific hour on a specific day of the week. Greater numbers of activities are represented by darker blue squares; lighter squares indicate smaller numbers of activities.

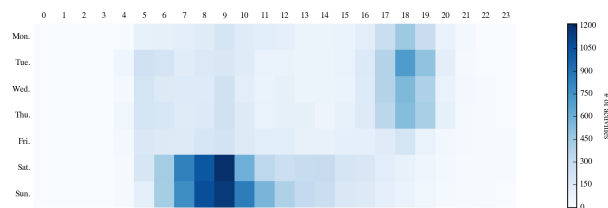
We observe that group activities, in general, are not frequent during regular work hours as we see clusters of dark blue squares positioned at the hours after work (nearly 5pm - 7pm) on weekdays or in the mornings (nearly 6am - 11am) on weekends. This pattern is consistent across gender, however males exercising with others tend to do so slightly early on weekends. We find that solo activities occur most often early in the morning or after work on weekdays, and in the mornings on weekends. Again, this pattern is consistent across gender with males exercising slightly earlier in the morning on weekdays and weekends.



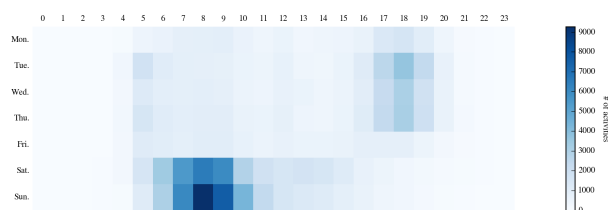
(a) Female - Solo Activities



(b) Male - Solo Activities



(c) Female - Group Activities



(d) Male - Group Activities

Figure 6.3: Heat maps visualizing the distribution of activities across local start time and day of the week for solo and group activities by gender.

6.6.3 Co-Present Peers and Activity Performance

Our second research question considers the relationship between the number of activity participants (i.e. peer co-presence) and activity performance. We measure performance

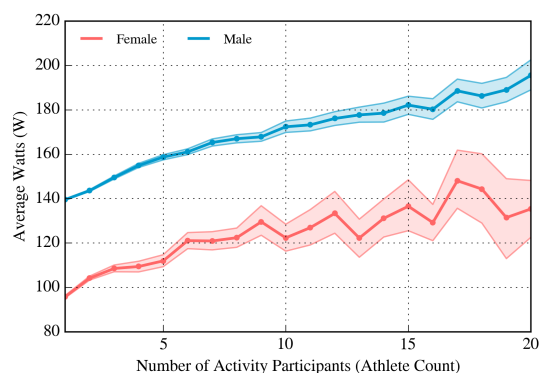
in terms of multiple dimensions, including, *distance*, *elevation*, *duration*, *activity effort*, and *physical challenge level*. These measures were discussed in detail in Section 6.5. For each measure, we take the average of its value across all activities given a specific **athlete count**. Then, we examine trends as the number of activity participants increases. We find a consistent pattern across all measures: performance increases sharply as the number of participants increases up to five, when performance shows diminishing returns – increasing but at a much smaller rate.

Figure 6.4 shows this result for average watts, as well as moving time, as a function of athlete count. Other performance metrics show similar results, and are available in the appendix. It should be noted that we have few activities with very large group size, so we expect greater noise (e.g. greater uncertainty about estimates and larger confidence intervals) as group size goes up. In order to present readable visualizations, we truncate athlete count at 20. Our preliminary analysis shows some gender differences in baseline activity effort, such as average watts for solo activities, so visualizations show results by gender. However, the observed relationships between performance and group size are consistent across both gender groups, as seen. Moving time is one of the few performance metrics that shows a negative relationship with group size, and only for groups larger than 5-10 athletes.

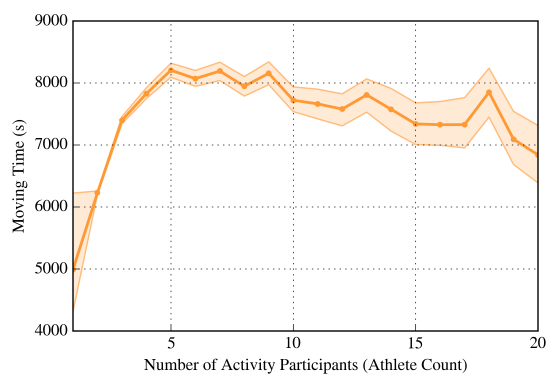
6.6.4 *Social Feedback for Activities*

Finally, we analyze how the number of athletes involved in a particular activity is associated with subsequent feedback received by the posting athlete by examining three main interaction types. Each of these types of social interaction enables Strava users to provide peers with social feedback about their posted activities. Users can make comments and give kudos (or “likes”). In addition, the original author of the activity can post associated photos.

Figure 6.5(a) shows what percentage of activities that were ever commented on by other platform users for type of activity, across genders and then combined. We find that group activities are much more likely to be commented on; percentages of activities commented involving groups are twice as high as those of solo activities in across all cases. Additionally,



(a) Average watts of activities posted by female users (red) and male users (blue) as a function of number of activity participants.



(b) Average moving time of activities as a function of number of activity participants.

Figure 6.4: Activity performance metrics as a function of number of participants. Colored bands around mean line represent bootstrapped 95% confidence intervals.

we do not see significant gender differences here. We do not show results for kudoed activities because it has roughly the same pattern as commented activities.

Figure 6.5(b) shows the percentage of activities with accompanying photos posted in Strava or Instagram. We see that users are more likely to post photos to group activities. Moreover, we observe that female users tend to post more photos to both solo and group

activities than male users.

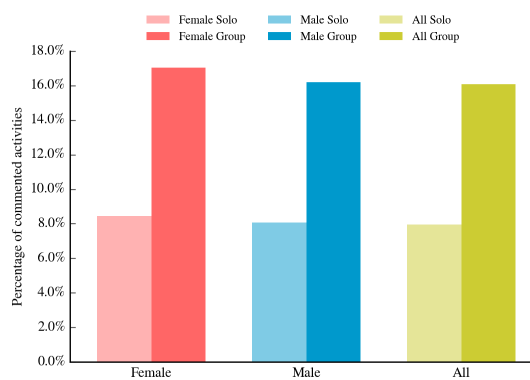
Overall, group activities and solo activities differ in terms of the proportion of received social feedback. We find that group activities are much more likely to attract social feedback including comments and likes from peers. Moreover, group activities tend to motivate users' behavior of content sharing by posting activity-related photos, which in turn is likely to gain more attention among users' online social circle.

6.7 Discussion

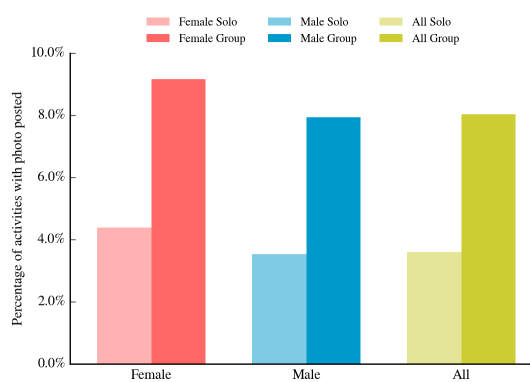
This study compares activities posted by users in an online fitness community. In particular, we focus analysis on how the characteristics of these activities – when they occur, how intense they seem to be, and how much social feedback they receive – may be associated with the number of co-participants. To do so we make use of a unique dataset collected from the online community Strava, utilizing application features that allow users to specify who they are exercising with; data comprise not only the behavior of individual athletes (users), but also detailed records of who is physically co-present with these users. Our analysis demonstrates a number of significant findings.

First, we observe that female users tend to post activities that involve a single peer – exercise events where the number of total participants is two. Males, on the other hand, tend to post solo activities or activities that involve larger groups. These results hint at specific gender preferences in group exercise and have numerous implications for peer effects on motivation and health promotion. Importantly, for any network-based intervention or behavior change, the social networks (and as a consequence influential peers) for men and women look very different (Granovetter, 1973; Centola and Macy, 2007; Bakshy et al., 2012; Lewis et al., 2012). Results suggest that females might have a single or small set of influential strong ties (i.e. exercise partners), while men may have a large, diverse set of peers who could be influential.

This study finds evidence for strong seasonal effects on group exercise. Group activities usually take place after work during the week or early in the day on weekends, whereas many



(a) Percentage of activities earning at least one comment by another Strava user



(b) Percentage of activities with photos available on Strava or Instagram

Figure 6.5: Measure of social feedback by gender and activity (solo versus group).

solo activities also take place in the early mornings on weekdays. Building from the previous discussion, while strong diurnal patterns are unsurprising in human behavior (Golder and Macy, 2011), results demonstrate that opportunities for peer influence on health behaviors are likely to be restricted or constrained in systematic ways. For example, designers of application features might suggest that exercise partners should take into account optimal times for group exercise and individuals preferences for when to work out. While outside the scope of this study, there are many interesting directions for future work that considers mechanisms to affect behavior change.

Group activities differ from solo activities, in terms of effort, exertion and performance, our analysis indicates. When exercising with others, even just a single peer, athletes see notable gains in workout intensity and energy expended - increases in average power output (measured in average watts for cycling events), moving time and distance. Interestingly, these gains continue to increase for every additional activity participant (though primary gains are seen for the first 5 additional co-present peers). Activities with co-present peers might be informal (organized by the participants themselves) or formal (group rides perhaps organized by a local club or other organizational entity). Further work might tease apart these different conditions to offer insight into peer effects and impact of institutional structure on exercise (Vilhjalmsson and Kristjansdottir, 2003).

Co-presence and social interaction are distinct but related concepts. In the final component of the analysis presented here, we consider observations of social exchange among athletes. In particular, we consider social feedback behaviors - platform users commenting and liking each others' activities. Findings demonstrate that group activities are associated with higher levels of social feedback than solo activities. Moreover, group activities are significantly more likely to include multimedia (photos). In the latter case, data also reveals gender differences, where females are more likely to post group activities with photos than males. Increased engagement and social feedback may also be related to motivation and future activity, suggesting more promising directions for further work.

6.8 Limitations

While the study presented here offers novel insight into the characteristics of physical activities where peers are co-present, it is not without limitations. One notable concern is the ability of Strava application users to restrict their activities to be private, shared only with pre-screened peers. Private accounts, and likewise private activities, cannot be accessed from the Strava API, and therefore are excluded from the data used in this study. Athletes who choose to restrict access to their data may systematically differ from users who make their data public. Unfortunately, we are unable to assess the impact of this bias because of lack of data. Instead, one should be careful about generalizing these results beyond the population of study.

6.9 Conclusion

As social fitness mobile applications become widely used for personal activity tracking, social support and health promotion, opportunities for understanding the effects of social networks and peer influence on behavior change and health expand. Drawing on features of social media and activity tracking applications, many of these new platforms capture rich data about physical activities as well as peer-to-peer interactions. The behavioral trace data they archive have the potential to significantly alter understanding of health and well-being. In this study, we explore the characteristics and dynamics of social exercise, that is fitness activities with at least one peer physically co-present. Our research focuses on quantifying diurnal patterns, activity performance and social feedback as they vary by the number of people participating in an activity; we also compare associations by gender. Our results indicate that when peers are physically co-present exercise tends to be more intense and receive more feedback from other users. Findings also suggest gender differences in the observed tendency to engage in physical activity with others. The implications of these results for network-based health and wellness interventions are also discussed.

<<END OF PUBLICATION>>

Chapter 7

CONCLUSION

The goal of this dissertation is to demonstrate that computational approaches can derive valuable insights from large-scale behavioral traces to better understand how fitness tracking applications and social support mechanisms work together to shape people's fitness behavior and social networks. How do social features and underlying mechanisms designed in fitness platforms facilitate people's behavior change? How are online social interactions tied to offline social support, if at all? Answers to these questions can lead to actionable insights into human behavior and health promotion, improving the health and well-being of communities.

To better understand fitness behavior of users in online fitness communities, it is necessary to first ask, and answer, a basic set of questions: how much and when do people exercise? how much do people interact with each other online fitness platforms, and in what capacity? The empirical chapters in this dissertation begin with these analyses, and build from this foundation to explore the intersection of social engagement online and social exercise in offline environments. Beginning first with some of the data collection challenges, this dissertation responds to platform restrictions imposed on data access and visibility by implementing a novel data sampling approach that allows one to obtain a representative dataset of users and their activities. Using this, along with a dataset obtained through collaboration with the a fitness tracking platform, this dissertation tackles a series of questions related to exercise and social interaction. Chapter 4 investigates the dynamics of people's physical activities and the social network of relationships with other users, whose activities they subscribe to in the fitness community, as well as their association. We learn that network interactions are playing important roles in supporting and encourage more active physical exercises. Next, to advance our understanding of online social support, Chapter 5 investigates the

underlying processes of how people’s online networks may be formed and demonstrates how online personal network structure varies by demographic and economic factors. Next, we ask about how such differences may also persist in physical exercise, by investigating social exercise – exercise that takes place with others. Chapter 6 analyzes the characteristics and dynamics of social exercise with at least one peer physically co-present. This work shows that social exercise has overall advantages over solo exercise, particularly in terms of exercise performance and social feedback. In all, this dissertation provides methodological and empirical contributions to better understand the complex mechanisms of social support on individual fitness behavior.

7.1 Summary of Contributions

In the preceding chapters, this work developed and utilized techniques to address core research questions at the intersection of information science, social science and computer science. It is well established in prior work that physical activity brings mental and physical benefits. Research on health intervention mechanisms to motivate behavior change and promote health has been a critical topic in social science especially in disciplines such as public health and sociology. Social support has been linked to many physical and mental benefits and has been demonstrated to be an effective intervention strategy. As pervasive technologies advance, activity tracking applications and online fitness communities are on the rise. Recently, increasing attention has been paid to promoting positive fitness behaviors through social interaction in online fitness communities. These platforms archive digital behavioral traces of users’ physical activity and online social interaction behaviors offering researchers a new data source – one that provides researchers opportunities to overcome limitations in generalizability of previous research on social support and health promotion. However, large-scale data also brings challenges to data collection, data processing, data analysis and data communication, which are of primary interest for researchers in information science. This work collects and leverages large-scale behavioral trace data, attempting to deepen current understanding of human behaviors and social dynamics as well as provide

new insights into health promotion and intervention. We build computational models using machine learning and social network-based analyses to investigate empirical patterns in self tracking of exercise.

This dissertation investigates the ways in which the tools of self-tracking and social support are designed and as a result shape people's fitness behavior. The research site for this work is Strava. Strava is an exemplar of many online platforms designed to allow tracking of exercise in combination with building a community of users to engage with each other around these fitness activities; we expect to see the methods developed and empirical learnings presented in this work to be broadly applicable.

Chapter 2 reviews studies on self-tracking, online fitness communities, social interaction mechanisms and social support for positive behavior change. Chapter 3 presents a framework of collecting behavioral traces from online fitness communities. We demonstrate these methods in the case of Strava, but one can easily adapt the data collection framework to collect behavioral traces in other online fitness communities within similar properties and access points. Next, in Chapter 4, we showed how to leverage large-scale fitness data to reveal people's fitness behavior and social interactions in online fitness communities; also, we modeled individuals' activity level - not only we found a strong co-relation between social subscription and activity occurrence, but also we quantified the "power" of gaining one follower to exercise more. We then, in Chapter 5, further proposed stochastic process models to characterize the empirical network degree distribution of fitness community members. The results demonstrate that gender, age and account status are associated with distinct network structure. Lastly, in Chapter 6, our focus switched from online social support to offline where we showed social support from physically co-present peers provides motivation to more intense activities and more feedback to other users; we also found interesting gender difference in the number of people males or females prefer to co-exercise with. These studies form a first step towards developing scalable computational methods to measure, understand and predict individual fitness behavior and the mechanism of social support in behavior change.

7.2 *The Role of Platform and Design Features*

While the preceding chapters focus on individuals and groups – people, as they navigate online fitness community – it is also important to recognize the role that these platforms themselves play in this domain. Previous research has found that content and features in online platforms can effectively nudge users towards certain attitudes and behavior (Gupta et al., 2021; Dessart and Duclou, 2019). For example, sharing features or a “share button” to encourage social sharing (Malhotra et al., 2013); gamification elements to nudge for more active behavior (Althoff et al., 2016); content (eg. images, quotes, advice) on exercise and healthy eating to facilitate positive perceptions of health (Raggatt et al., 2018). Indeed, such online platforms are constantly adding features to engage users in behavior change and health promotion (Couture, 2021; Wang et al., 2016). It is important to recognize that the empirical results discussed in this dissertation result from the joint influence of individual actions or preferences along with these platform nudges. For example, as individuals join the Strava platform they may be suggested social contacts based on other social media accounts or local users. Hence their activity behavior and social networks as observed in our datasets is not only an outcome from their individuals choices, but likely a joint outcome from intrinsic sources of motivation and algorithmic decisions made by the Strava platform which help to establish or strengthen their social networks. This is one dimension of why we choose Strava as an exemplar online fitness community to study social support on individual fitness behavior.

Platforms are also constantly in flux; new features are introduced, algorithms that govern recommendations are updated and expanded, and different experiences are tested against each other in live experiments. This environment is much different from prior work on health behaviors, most of which was confined to laboratory settings. Data collected from lab experiments are often expensive and at limited scale. For researchers, online platforms also offer pathways for experiment design and the ability to examine causal effects on human behavior (at much larger scale and lower cost than previously possible). For

example, natural experiments can analyze individual behavior before and after introducing new features in platforms, estimating the effect of the feature on how individual users behave (Althoff et al., 2017a); it is also possible to conduct carefully devised controlled experiments in online platforms (Aral and Walker, 2009; Bapna and Umyarov, 2015). For example, this dissertation points out people’s preference for choosing exercise partners with specific genders. One potential future work might be able to establish a causal link for this finding by suggesting exercise partners – those whose gender matches the subject as well as cross-gender suggestions. Observing how these suggestions translate into offline exercise could help tease apart the choice of partner, especially if the researchers are able to control for other factors such as physical distance apart, workout schedules, etc. There are many exciting possibilities for future work here.

7.3 Future Directions

The results presented in this dissertation point to several interesting future directions, some of which we shall outline here.

7.3.1 Social support to promote health-related behavior in marginalized groups

The society we live in is marked by social inequalities, many of which related to class, gender and age (Skauge and Seippel, 2020; Ulseth and Seippel, 2011). This dissertation showed that gender difference exists in people’s fitness behavior, the processes governing the formation of online network ties, perceptions of social support after participating in physical activities, etc.

However, academic research on how marginalized or stigmatized groups engage in or react to digital self-tracking for fitness well-being is still in its infancy. In addition to gender inequality, how are elderly people, people from minority ethnic or racial groups, people with poor literacy skills or people with disabilities receive and/or perceive social support via the engagement of digital self-tracking technologies and online communities? Is the ever-increasing popularity of digital self-tracking practices and online fitness communities

widening or narrowing social inequality in fitness and health? A better understanding of the mechanisms of social support on these populations, particularly with respect to exercise-related behavior is ever-more important as it can potentially decrease social inequality in fitness, and therefore health.

7.3.2 Fitness behavior change over a long period of time

This dissertation examined large-scale trajectories of users' posted activities to analyze how their activity level may change over time. This study analyzed users in two metropolitan areas - Boston and San Francisco. The ability to use large-scale behavioral traces to understand behavior change over a long period of time is crucial to sustainably improve the overall social well-being.

Currently, the world is suffering under the Covid-19 pandemic. How does fitness behavior change as a result of social distancing and/or lockdown at population scale during the Covid-19 pandemic? How should we develop strategies that still motivate positive change of health-related behavior during crisis? To answer these questions can be a bigger challenge but lead to meaningful outcomes for social good.

7.3.3 Supportive social networks

Online fitness communities are primarily designed to promote positive fitness behavior. Millions of social interactions among users are taking place during their daily usage of the services. In this dissertation, we leveraged large-scale digital traces of online network interactions and offline co-exercise interactions to better understand how to support each other to better achieve one's fitness goal. We have also shown that social support both online or offline can have significant relationship on someone's activity level and activity performance.

However, we have yet to fully understand supportive social networks that optimize for the users' mental and physical health. In what ways do social interactions in such networks

provide the members with comfort or reassurance? In contrast, when do such interactions frustrate or disappoint people? These questions may point to an exciting future direction.

BIBLIOGRAPHY

- G. Acar, B. Van Alsenoy, F. Piessens, C. Diaz, and B. Preneel. Facebook tracking through social plug-ins. *Belgian Privacy Commission, Ver*, 1(2), 2015.
- A. Ahtinen, M. Isomursu, M. Mukhtar, J. Mäntyjärvi, J. Häkkinä, and J. Blom. Designing social features for mobile and ubiquitous wellness applications. In *Proceedings of the 8th international Conference on Mobile and Ubiquitous Multimedia*, page 12. ACM, 2009.
- H. Akaike. Factor analysis and aic. In *Selected Papers of Hirotugu Akaike*, pages 371–386. Springer, 1987.
- N. Ali-Hasan, D. Gavales, A. Peterson, and M. Raw. Fitster: social fitness information visualizer. In *CHI'06 Extended Abstracts on Human Factors in Computing Systems*, pages 1795–1800. ACM, 2006.
- Z. W. Almquist and C. T. Butts. Logistic network regression for scalable analysis of networks with joint edge/vertex dynamics. *Sociological methodology*, 44(1):273–321, 2014.
- Z. W. Almquist, S. Arya, L. Zeng, and E. Spiro. Unbiased sampling of users from (online) activity data. *Field Methods*, 31(1):23–38, 2019.
- A. Alqhatani and H. R. Lipford. {“There”} is nothing that i need to keep {secret”}: Sharing practices and concerns of wearable fitness data. In *Fifteenth Symposium on Usable Privacy and Security (SOUPS 2019)*, pages 421–434, 2019.
- T. Althoff. Population-scale pervasive health. *IEEE Pervasive Computing*, 16(4):75–79, 2017.

- T. Althoff, R. W. White, E. Horvitz, et al. Influence of pokémon go on physical activity: study and implications. *Journal of medical Internet research*, 18(12):e6759, 2016.
- T. Althoff, P. Jindal, and J. Leskovec. Online actions with offline impact: How online social networks influence online and offline user behavior. In *Proc. of the Tenth ACM International Conference on Web Search and Data Mining*, pages 537–546. ACM, 2017a.
- T. Althoff, R. Sosič, J. L. Hicks, A. C. King, S. L. Delp, and J. Leskovec. Large-scale physical activity data reveal worldwide activity inequality. *Nature*, 547(7663):336–339, 2017b.
- L. A. N. Amaral, A. Scala, M. Barthelemy, and H. E. Stanley. Classes of small-world networks. *Proc. of the national academy of sciences*, 97(21):11149–11152, 2000.
- S. Aral and C. Nicolaides. Exercise contagion in a global social network. *Nature communications*, 8:14753, 2017.
- S. Aral and D. Walker. Identifying peer influence in massive online social networks: A platform for randomized experimentation on facebook. In *Workshop on Information Systems Economics (WISE)*, page 1, 2009.
- E. Bakshy, I. Rosenn, C. Marlow, and L. Adamic. The role of social networks in information diffusion. In *Proceedings of the 21st international conference on World Wide Web*, pages 519–528. ACM, 2012.
- A. Bandura. Health promotion by social cognitive means. *Health education & behavior*, 31(2):143–164, 2004.
- R. Bapna and A. Umyarov. Do your online friends make you pay? a randomized field experiment on peer influence in online social networks. *Management Science*, 61(8):1902–1920, 2015.
- A.-L. Barabási and R. Albert. Emergence of scaling in random networks. *science*, 286(5439):509–512, 1999.

- A.-L. Barabási and E. Bonabeau. Scale-free networks. *Scientific american*, 288(5):60–69, 2003.
- T. Baranowski, B. Simons-Morton, P. Hooks, J. Henske, K. Tiernan, J. Kay Dunn, H. Burkhalter, J. Harper, and J. Palmer. A center-based program for exercise change among black-american families. *Health Education Quarterly*, 17(2):179–196, 1990.
- D. Belle. Gender differences in the social moderators of stress. *Stress and coping: An anthology*, 1991.
- L. F. Berkman and T. Glass. Social integration, social networks, social support, and health. *Social epidemiology*, 1:137–173, 2000.
- H. R. Bernard, P. Killworth, D. Kronenfeld, and L. Sailer. The problem of informant accuracy: The validity of retrospective data. *Annual review of anthropology*, pages 495–517, 1984.
- S. Biddle and M. Goudas. Analysis of children’s physical activity and its association with adult encouragement and social cognitive variables. *Journal of School Health*, 66(2):75–78, 1996.
- S. N. Blair. Evidence for success of exercise in weight loss and control. *Annals of Internal Medicine*, 119(7_Part_2):702–706, 1993.
- T. G. P. R. Bogdan. Psychological benefits of exercising with another. *Journal of Human Movement Studies*, 44:093–106, 2003.
- H. Bozdogan. Model selection and akaike’s information criterion (aic): The general theory and its analytical extensions. *Psychometrika*, 52(3):345–370, 1987.
- K. P. Burnham and D. R. Anderson. *Model selection and multimodel inference: a practical information-theoretic approach*. Springer Science & Business Media, 2003.

- T. Campbell, B. Ngo, and J. Fogarty. Game design principles in everyday fitness applications. In *Proceedings of the 2008 ACM conference on Computer supported cooperative work*, pages 249–252. ACM, 2008.
- D. N. Cavallo, D. F. Tate, A. V. Ries, J. D. Brown, R. F. DeVellis, and A. S. Ammerman. A social media-based physical activity intervention: a randomized controlled trial. *American journal of preventive medicine*, 43(5):527–532, 2012.
- D. Centola. An experimental study of homophily in the adoption of health behavior. *Science*, 334(6060):1269–1272, 2011.
- D. Centola. Social media and the science of health behavior. *Circulation*, 127(21):2135–2144, 2013.
- D. Centola and M. Macy. Complex contagions and the weakness of long ties¹. *American journal of Sociology*, 113(3):702–734, 2007.
- D. Centola and A. van de Rijt. Choosing your network: Social preferences in an online health community. *Social Science & Medicine*, 125:19–31, 2015.
- X. Chen, Z. Zhu, M. Chen, and Y. Li. Large-scale mobile fitness app usage analysis for smart health. *IEEE Communications Magazine*, 56(4):46–52, 2018.
- Y. Chen and P. Pu. Healthytogether: exploring social incentives for mobile fitness applications. In *Proceedings of the Second International Symposium of Chinese CHI*, pages 25–34. ACM, 2014.
- E. K. Choe, N. B. Lee, B. Lee, W. Pratt, and J. A. Kientz. Understanding quantified-selfers’ practices in collecting and exploring personal data. In *Proceedings of the SIGCHI conference on human factors in computing systems*, pages 1143–1152, 2014.
- G. Choi and H. Chung. Applying the technology acceptance model to social networking sites

- (sns): Impact of subjective norm and social capital on the acceptance of sns. *International Journal of Human-Computer Interaction*, 29(10):619–628, 2013.
- A. Clauset, C. R. Shalizi, and M. E. Newman. Power-law distributions in empirical data. *SIAM review*, 51(4):661–703, 2009.
- D. A. Cohen, T. Marsh, S. Williamson, D. Golinelli, and T. L. McKenzie. Impact and cost-effectiveness of family fitness zones: a natural experiment in urban public parks. *Health & place*, 18(1):39–45, 2012.
- S. Consolvo, K. Everitt, I. Smith, and J. A. Landay. Design requirements for technologies that encourage physical activity. In *Proceedings of the SIGCHI conference on Human Factors in computing systems*, pages 457–466. ACM, 2006.
- S. Consolvo, D. W. McDonald, T. Toscos, M. Y. Chen, J. Froehlich, B. Harrison, P. Klasnja, A. LaMarca, L. LeGrand, R. Libby, et al. Activity sensing in the wild: a field trial of ubifit garden. In *Proceedings of the SIGCHI Conference on Human Factors in Computing Systems*, pages 1797–1806. ACM, 2008.
- J. Couture. Reflections from the ‘strava-sphere’: Kudos, community, and (self-) surveillance on a social network for athletes. *Qualitative Research in Sport, Exercise and Health*, 13(1):184–200, 2021.
- L. De Vries, S. Gensler, and P. S. Leeflang. Popularity of brand posts on brand fan pages: An investigation of the effects of social media marketing. *Journal of interactive marketing*, 26(2):83–91, 2012.
- L. Dessart and M. Duclou. Health and fitness online communities and product behaviour. *Journal of Product & Brand Management*, 2019.
- R. K. Dishman, J. F. Sallis, and D. R. Orenstein. The determinants of physical activity and exercise. *Public health reports*, 100(2):158, 1985.

- S. N. Dorogovtsev and J. F. Mendes. Evolution of networks. *Advances in physics*, 51(4):1079–1187, 2002.
- P. Drentea and J. L. Moren-Cross. Social capital and social support on the web: the case of an internet mother site. *Sociology of health & illness*, 27(7):920–943, 2005.
- R. Dubos. *Social capital: Theory and research*. Routledge, 2017.
- N. Eagle, A. S. Pentland, and D. Lazer. Inferring friendship network structure by using mobile phone data. *Proc. of the national academy of sciences*, 106(36):15274–15278, 2009.
- M. Ehn, L. C. Eriksson, N. Åkerberg, A.-C. Johansson, et al. Activity monitors as support for older persons’ physical activity in daily life: qualitative study of the users’ experiences. *JMIR mHealth and uHealth*, 6(2):e8345, 2018.
- T. Field, D. Miguel, and C. E. Sanders. Exercise is positively related to adolescents’ relationships and academics. *Adolescence*, 36(141):105, 2001.
- A. W. Flaherty. Frontotemporal and dopaminergic control of idea generation and creative drive. *Journal of Comparative Neurology*, 493(1):147–153, 2005.
- G. F. Fletcher, G. Balady, S. N. Blair, J. Blumenthal, C. Caspersen, B. Chaitman, S. Epstein, E. S. S. Froelicher, V. F. Froelicher, I. L. Pina, et al. Statement on exercise: Benefits and recommendations for physical activity programs for all americans a statement for health professionals by the committee on exercise and cardiac rehabilitation of the council on clinical cardiology, american heart association. *Circulation*, 94(4):857–862, 1996.
- D. Foster, C. Linehan, B. Kirman, S. Lawson, and G. James. Motivating physical activity at work: using persuasive social media for competitive step counting. In *Proceedings of the 14th International Academic MindTrek Conference: Envisioning Future Media Environments*, pages 111–116. ACM, 2010.
- S. Fox. The social life of health information, 2011, 2011.

- R. Z. Franco, R. Fallaize, J. A. Lovegrove, and F. Hwang. Popular nutrition-related mobile apps: a feature assessment. *JMIR mHealth and uHealth*, 4(3):e5846, 2016.
- C. J. Frankish, C. D. Milligan, and C. Reid. A review of relationships between active living and determinants of health. *Social Science & Medicine*, 47(3):287–301, 1998.
- J. L. Freeman and G. Neff. The challenge of repurposed technologies for youth: Understanding the unique affordances of digital self-tracking for adolescents. *new media & society*, page 14614448211040266, 2021.
- D. Freitag. Machine learning for information extraction in informal domains. *Machine learning*, 39(2-3):169–202, 2000.
- W. Gerin, C. Pieper, R. Levy, and T. G. Pickering. Social support in social interaction: a moderator of cardiovascular reactivity. *Psychosomatic medicine*, 54(3):324–336, 1992.
- A. Giachanou and F. Crestani. Like it or not: A survey of twitter sentiment analysis methods. *ACM Computing Surveys (CSUR)*, 49(2):28, 2016.
- K. Gimpel, N. Schneider, B. O’Connor, D. Das, D. Mills, J. Eisenstein, M. Heilman, D. Yogatama, J. Flanigan, and N. A. Smith. Part-of-speech tagging for twitter: Annotation, features, and experiments. In *Proceedings of the 49th Annual Meeting of the Association for Computational Linguistics: Human Language Technologies: short papers-Volume 2*, pages 42–47. Association for Computational Linguistics, 2011.
- M. Gjoka, M. Kurant, C. T. Butts, and A. Markopoulou. Walking in facebook: A case study of unbiased sampling of osns. In *2010 Proceedings IEEE Infocom*, pages 1–9. Ieee, 2010.
- E. Goffmann. Behavior in public places. *New York*, (s 8), 1963.
- S. A. Golder and M. W. Macy. Diurnal and seasonal mood vary with work, sleep, and daylength across diverse cultures. *Science*, 333(6051):1878–1881, 2011.

- M. S. Granovetter. The strength of weak ties. *American journal of sociology*, pages 1360–1380, 1973.
- M. S. Granovetter. The strength of weak ties. In *Social networks*, pages 347–367. Elsevier, 1977.
- X. Gui, Y. Chen, C. Caldeira, D. Xiao, and Y. Chen. When fitness meets social networks: Investigating fitness tracking and social practices on werun. In *Proceedings of the 2017 CHI Conference on Human Factors in Computing Systems*, pages 1647–1659, 2017.
- J. Gummerus, V. Liljander, E. Weman, and M. Pihlström. Customer engagement in a facebook brand community. *Management Research Review*, 35(9):857–877, 2012.
- A. Gupta, N. Dhiman, A. Yousaf, and N. Arora. Social comparison and continuance intention of smart fitness wearables: An extended expectation confirmation theory perspective. *Behaviour & Information Technology*, 40(13):1341–1354, 2021.
- P. C. Hallal, L. B. Andersen, F. C. Bull, R. Guthold, W. Haskell, U. Ekelund, L. P. A. S. W. Group, et al. Global physical activity levels: surveillance progress, pitfalls, and prospects. *The lancet*, 380(9838):247–257, 2012.
- J. Hamari and J. Koivisto. “working out for likes”: An empirical study on social influence in exercise gamification. *Computers in Human Behavior*, 50:333–347, 2015.
- M. S. Handcock. Degreenet: Models for skewed count distributions relevant to networks. *Statnet Project, Seattle, WA. Version*, 1, 2003.
- M. S. Handcock and J. H. Jones. Likelihood-based inference for stochastic models of sexual network formation. *Theoretical population biology*, 65(4):413–422, 2004.
- M. S. Handcock and J. H. Jones. Interval estimates for epidemic thresholds in two-sex network models. *Theoretical Population Biology*, 70(2):125–134, 2006.

- W. Hardeman, J. Houghton, K. Lane, A. Jones, and F. Naughton. A systematic review of just-in-time adaptive interventions (jitais) to promote physical activity. *International Journal of Behavioral Nutrition and Physical Activity*, 16(1):1–21, 2019.
- W. L. Haskell, H. J. Montoye, and D. Orenstein. Physical activity and exercise to achieve health-related physical fitness components. *Public Health Reports*, 100(2):202, 1985.
- L. Hassan, A. Dias, and J. Hamari. How motivational feedback increases user’s benefits and continued use: A study on gamification, quantified-self and social networking. *International Journal of Information Management*, 46:151–162, 2019.
- G. W. Heath, D. C. Parra, O. L. Sarmiento, L. B. Andersen, N. Owen, S. Goenka, F. Montes, R. C. Brownson, L. P. A. S. W. Group, et al. Evidence-based intervention in physical activity: lessons from around the world. *The lancet*, 380(9838):272–281, 2012.
- R. A. Hill and R. I. Dunbar. Social network size in humans. *Human nature*, 14(1):53–72, 2003.
- A. Holton, N. Lee, and R. Coleman. Commenting on health: A framing analysis of user comments in response to health articles online. *Journal of Health Communication*, 19(7):825–837, 2014.
- A. Holzinger, G. Searle, and A. Nischelwitzer. On some aspects of improving mobile applications for the elderly. In *International Conference on Universal Access in Human-Computer Interaction*, pages 923–932. Springer, 2007.
- D. Horton and R. Richard Wohl. Mass communication and para-social interaction: Observations on intimacy at a distance. *Psychiatry*, 19(3):215–229, 1956.
- C. M. Hurvich and C.-L. Tsai. Regression and time series model selection in small samples. *Biometrika*, 76(2):297–307, 1989.

- K. O. Hwang, A. J. Ottenbacher, A. P. Green, M. R. Cannon-Diehl, O. Richardson, E. V. Bernstam, and E. J. Thomas. Social support in an internet weight loss community. *International journal of medical informatics*, 79(1):5–13, 2010.
- IQVIA. Digital health trends, 2021.
- J. O. Irwin et al. The place of mathematics in medical and biological statistics. *Journal of the Royal Statistical Society*, 126(Pt. 1):1–41, 1963.
- H. C. Janisse, D. Nedd, S. Escamilla, and M. A. Nies. Physical activity, social support, and family structure as determinants of mood among european-american and african-american women. *Women & health*, 39(1):101–116, 2004.
- R. Jaslow. Cdc: 80 percent of american adults don't get recommended exercise. cbc news may 3, 2013, 2013.
- D. Jin, H. Halvari, N. Maehle, and A. H. Olafsen. Self-tracking behaviour in physical activity: a systematic review of drivers and outcomes of fitness tracking. *Behaviour & Information Technology*, pages 1–20, 2020.
- S. Kabadayi and K. Price. Consumer–brand engagement on facebook: liking and commenting behaviors. *Journal of Research in Interactive Marketing*, 8(3):203–223, 2014.
- E. B. Kahn, L. T. Ramsey, R. C. Brownson, G. W. Heath, E. H. Howze, K. E. Powell, E. J. Stone, M. W. Rajab, and P. Corso. The effectiveness of interventions to increase physical activity: A systematic review^{1, 2}. *American journal of preventive medicine*, 22(4):73–107, 2002.
- P. Kallio, K. Pahkala, O. J. Heinonen, T. H. Tammelin, K. Pälve, M. Hirvensalo, M. Juonala, B.-M. Loo, C. G. Magnussen, S. Rovio, et al. Physical inactivity from youth to adulthood and adult cardiometabolic risk profile. *Preventive Medicine*, 145:106433, 2021.

- D. B. Kandel. Homophily, selection, and socialization in adolescent friendships. *American journal of Sociology*, 84(2):427–436, 1978.
- S.-G. Kang, J. M. Kang, K.-P. Ko, S.-C. Park, S. Mariani, and J. Weng. Validity of a commercial wearable sleep tracker in adult insomnia disorder patients and good sleepers. *Journal of psychosomatic research*, 97:38–44, 2017.
- R. E. Kass and L. Wasserman. A reference bayesian test for nested hypotheses and its relationship to the schwarz criterion. *Journal of the american statistical association*, 90(431):928–934, 1995.
- R. Kent. Self-tracking health over time: From the use of instagram to perform optimal health to the protective shield of the digital detox. *Social Media+ Society*, 6(3):2056305120940694, 2020.
- Y. Kim, T. Schneider, E. Faß, and M. Lochbaum. Personal social capital and self-rated health among middle-aged and older adults: A cross-sectional study exploring the roles of leisure-time physical activity and socioeconomic status. *BMC Public Health*, 21(1):1–11, 2021.
- A. C. King. Interventions to promote physical activity by older adults. *The Journals of Gerontology Series A: Biological Sciences and Medical Sciences*, 56(suppl_2):36–46, 2001.
- A. C. King, K. Glanz, and K. Patrick. Technologies to measure and modify physical activity and eating environments. *American journal of preventive medicine*, 48(5):630–638, 2015.
- S. Kiritchenko, X. Zhu, and S. M. Mohammad. Sentiment analysis of short informal texts. *Journal of Artificial Intelligence Research*, 50:723–762, 2014.
- H. W. Kohl 3rd, C. L. Craig, E. V. Lambert, S. Inoue, J. R. Alkandari, G. Leetongin, S. Kahlmeier, L. P. A. S. W. Group, et al. The pandemic of physical inactivity: global action for public health. *The Lancet*, 380(9838):294–305, 2012.

- M. Kosinski, S. C. Matz, S. D. Gosling, V. Popov, and D. Stillwell. Facebook as a research tool for the social sciences: Opportunities, challenges, ethical considerations, and practical guidelines. *American psychologist*, 70(6):543, 2015.
- J. A. Kulik and H. I. Mahler. Social support and recovery from surgery. *Health Psychology*, 8(2):221, 1989.
- H. Kwak, C. Lee, H. Park, and S. Moon. What is twitter, a social network or a news media? In *Proceedings of the 19th international conference on World wide web*, pages 591–600. ACM, 2010.
- Y. Laird, S. Fawcner, and A. Niven. A grounded theory of how social support influences physical activity in adolescent girls. *International journal of qualitative studies on health and well-being*, 13(1):1435099, 2018.
- C. Lampe. Behavioral trace data for analyzing online communities. *The SAGE Handbook of Digital Technology Research*, page 236, 2013.
- H. Lee, C. Wildeman, E. A. Wang, N. Matusko, and J. S. Jackson. A heavy burden: the cardiovascular health consequences of having a family member incarcerated. *American journal of public health*, 104(3):421–427, 2014.
- A. Lenhart, K. Purcell, A. Smith, and K. Zickuhr. Social media & mobile internet use among teens and young adults. millennials. *Pew internet & American life project*, 2010.
- C. A. Levinson, L. Fewell, and L. C. Brosos. My fitness pal calorie tracker usage in the eating disorders. *Eating behaviors*, 27:14–16, 2017.
- K. Lewis, M. Gonzalez, and J. Kaufman. Social selection and peer influence in an online social network. *Proceedings of the National Academy of Sciences*, 109(1):68–72, 2012.
- N. Lin. Building a network theory of social capital. In *Social capital*, pages 3–28. Routledge, 2017.

- J. M. Linenger, C. V. Chesson, and D. S. Nice. Physical fitness gains following simple environmental change. *American Journal of Preventive Medicine*, 7(5):298–310, 1991.
- A. Lipsman, G. Mudd, M. Rich, and S. Bruich. The power of “like”: How brands reach (and influence) fans through social-media marketing. *Journal of Advertising research*, 52(1):40–52, 2012.
- C. Lister, J. H. West, B. Cannon, T. Sax, and D. Brodegard. Just a fad? gamification in health and fitness apps. *JMIR serious games*, 2(2), 2014.
- Y. Liu, X. Lv, and Z. Tang. The impact of mortality salience on quantified self behavior during the covid-19 pandemic. *Personality and Individual Differences*, 180:110972, 2021.
- D. Lupton. *Self-tracking, health and medicine*, 2017.
- K. S. Lyons, K. M. Winters-Stone, J. A. Bennett, and T. M. Beer. The effects of partnered exercise on physical intimacy in couples coping with prostate cancer. *Health Psychology*, 35(5):509, 2016.
- J. Mahmud, J. Chen, and J. Nichols. When will you answer this? estimating response time in twitter. In *ICWSM*, 2013.
- A. Malhotra, C. K. Malhotra, and A. See. How to create brand engagement on facebook. *MIT Sloan Management Review*, 54(2):18–20, 2013.
- P. V. Marsden and N. E. Friedkin. Network studies of social influence. *Sociological Methods & Research*, 22(1):127–151, 1993.
- E. Martínez-Cámara, M. T. Martín-Valdivia, L. A. Urena-López, and A. R. Montejo-Ráez. Sentiment analysis in twitter. *Natural Language Engineering*, 20(1):1–28, 2014.
- E. McAuley, B. Blissmer, D. X. Marquez, G. J. Jerome, A. F. Kramer, and J. Katula. Social relations, physical activity, and well-being in older adults. *Preventive medicine*, 31(5):608–617, 2000.

- K. R. McLeroy, D. Bibeau, A. Steckler, and K. Glanz. An ecological perspective on health promotion programs. *Health Education & Behavior*, 15(4):351–377, 1988.
- M. McLuhan. *Understanding media: The extensions of man*. MIT press, 1994.
- M. McPherson, L. Smith-Lovin, and J. M. Cook. Birds of a feather: Homophily in social networks. *Annual review of sociology*, 27(1):415–444, 2001.
- F. Meng, X. Zhang, L. Liu, and C. Ren. Converting readers to patients? from free to paid knowledge-sharing in online health communities. *Information Processing & Management*, 58(3):102490, 2021.
- A. Mislove, M. Marcon, K. P. Gummadi, P. Druschel, and B. Bhattacharjee. Measurement and analysis of online social networks. In *Proc. of the 7th ACM SIGCOMM conference on Internet measurement*, pages 29–42. ACM, 2007.
- K. Miyata. Social support for japanese mothers online and offline. *The internet in everyday life*, pages 520–548, 2002.
- M. D. Molina and S. S. Sundar. Can mobile apps motivate fitness tracking? a study of technological affordances and workout behaviors. *Health communication*, 2018.
- S. A. Moorhead, D. E. Hazlett, L. Harrison, J. K. Carroll, A. Irwin, and C. Hoving. A new dimension of health care: systematic review of the uses, benefits, and limitations of social media for health communication. *Journal of medical Internet research*, 15(4):e85, 2013.
- M. E. Morris, S. Consolvo, S. Munson, K. Patrick, J. Tsai, and A. D. Kramer. Facebook for health: opportunities and challenges for driving behavior change. In *CHI'11 Extended Abstracts on Human Factors in Computing Systems*, pages 443–446. ACM, 2011.
- L. Muchnik, S. Aral, and S. J. Taylor. Social influence bias: A randomized experiment. *Science*, 341(6146):647–651, 2013.

- F. Mueller, F. Vetere, M. Gibbs, D. Edge, S. Agamanolis, J. Sheridan, and J. Heer. Balancing exertion experiences. In *Proceedings of the SIGCHI Conference on Human Factors in Computing Systems*, pages 1853–1862. ACM, 2012.
- S. A. Munson and S. Consolvo. Exploring goal-setting, rewards, self-monitoring, and sharing to motivate physical activity. In *2012 6th International Conference on Pervasive Computing Technologies for Healthcare (PervasiveHealth) and Workshops*, pages 25–32. IEEE, 2012.
- S. A. Munson, H. Cavusoglu, L. Frisch, and S. Fels. Sociotechnical challenges and progress in using social media for health. *Journal of Medical Internet Research*, 15:e226, 2013.
- P. Nakov, A. Ritter, S. Rosenthal, F. Sebastiani, and V. Stoyanov. Semeval-2016 task 4: Sentiment analysis in twitter. In *Proceedings of the 10th International Workshop on Semantic Evaluation (SemEval-2016)*, pages 1–18, 2016.
- A. Nazi, Z. Zhou, S. Thirumuruganathan, N. Zhang, and G. Das. Walk, not wait: Faster sampling over online social networks. *arXiv preprint arXiv:1410.7833*, 2014.
- M. Newman. *Networks*. Oxford university press, 2018.
- M. E. Newman. The structure and function of complex networks. *SIAM review*, 45(2): 167–256, 2003.
- M. W. Newman, D. Lauterbach, S. A. Munson, P. Resnick, and M. E. Morris. It’s not that i don’t have problems, i’m just not putting them on facebook: challenges and opportunities in using online social networks for health. In *Proc. of the ACM 2011 conference on Computer supported cooperative work*, pages 341–350. ACM, 2011.
- P. K. Novak, J. Smailović, B. Sluban, and I. Mozetič. Sentiment of emojis. *PloS one*, 10(12):e0144296, 2015.

- M. J. Paul and M. Dredze. You are what you tweet: Analyzing twitter for public health. *ICWSM*, 20:265–272, 2011.
- C. Pechmann, L. Pan, K. Delucchi, C. M. Lakon, and J. J. Prochaska. Development of a twitter-based intervention for smoking cessation that encourages high-quality social media interactions via automessages. *Journal of medical Internet research*, 17(2):e50, 2015.
- K. A. Pelphrey, T. V. Mitchell, M. J. McKeown, J. Goldstein, T. Allison, and G. McCarthy. Brain activity evoked by the perception of human walking: controlling for meaningful coherent motion. *The Journal of Neuroscience*, 23(17):6819–6825, 2003.
- F. J. Penedo and J. R. Dahn. Exercise and well-being: a review of mental and physical health benefits associated with physical activity. *Current opinion in psychiatry*, 18(2):189–193, 2005.
- E. M. Piras and F. Miele. Clinical self-tracking and monitoring technologies: negotiations in the ict-mediated patient–provider relationship. *Health Sociology Review*, 26(1):38–53, 2017.
- T. G. Plante, L. Coscarelli, and M. Ford. Does exercising with another enhance the stress-reducing benefits of exercise? *International Journal of Stress Management*, 8(3):201–213, 2001.
- S. A. Prince, K. B. Adamo, M. E. Hamel, J. Hardt, S. C. Gorber, and M. Tremblay. A comparison of direct versus self-report measures for assessing physical activity in adults: a systematic review. *International Journal of Behavioral Nutrition and Physical Activity*, 5(1):56, 2008.
- S. Purpura, V. Schwanda, K. Williams, W. Stubler, and P. Sengers. Fit4life: the design of a persuasive technology promoting healthy behavior and ideal weight. In *Proceedings of the SIGCHI conference on human factors in computing systems*, pages 423–432, 2011.

- M. Raggatt, C. J. Wright, E. Carrotte, R. Jenkinson, K. Mulgrew, I. Prichard, and M. S. Lim. “i aspire to look and feel healthy like the posts convey”: engagement with fitness inspiration on social media and perceptions of its influence on health and wellbeing. *BMC public health*, 18(1):1–11, 2018.
- L. B. Ransdell, E. Eastep, A. Taylor, D. Oakland, J. Schmidt, L. Moyer-Mileur, and B. Shultz. Daughters and mothers exercising together (damet): Effects of home-and university-based interventions on physical activity behavior and family relations. *American Journal of Health Education*, 34(1):19–29, 2003.
- R. Robbins, A. Seixas, L. Walton Masters, N. Chanko, F. Diaby, D. Vieira, and G. Jean-Louis. Sleep tracking: A systematic review of the research using commercially available technology. *Current sleep medicine reports*, 5(3):156–163, 2019.
- E. M. Rogers. *Diffusion of innovations*. Simon and Schuster, 2010.
- C. E. Ross and D. Hayes. Exercise and psychologic well-being in the community. *American Journal of Epidemiology*, 127(4):762–771, 1988.
- T. Sakaki, M. Okazaki, and Y. Matsuo. Earthquake shakes twitter users: real-time event detection by social sensors. In *Proc. of the 19th international conference on World wide web*, pages 851–860. ACM, 2010.
- M. M. Saleh, S. A. Abd Rahman, A. Nazarudin, N. A. A. Kasim, and F. A. A. Razak. Eatntrack malaysia mobile application food calories tracker-a conceptual paper. *e-Academia Journal*, 10(1), 2021.
- M. Sharma. *Theoretical foundations of health education and health promotion*. Jones & Bartlett Learning, 2021.
- L. H. Shaw and L. M. Gant. In defense of the internet: The relationship between internet communication and depression, loneliness, self-esteem, and perceived social support. *Cyberpsychology & behavior*, 5(2):157–171, 2002.

- H. A. Simon. On a class of skew distribution functions. *Biometrika*, 42(3/4):425–440, 1955.
- M. Skauge and Ø. Seippel. Where do they all come from? youth, fitness gyms, sport clubs and social inequality. *Sport in Society*, pages 1–21, 2020.
- E. S. Spiro and Z. W. Almquist. Mechanisms of tie formation in activity-based online social networks, 2016, 2016.
- E. S. Spiro, Z. W. Almquist, and C. T. Butts. The persistence of division: Geography, institutions, and online friendship ties. *Socius*, 2:2378023116634340, 2016.
- C. Steinfield, N. B. Ellison, and C. Lampe. Social capital, self-esteem, and use of online social network sites: A longitudinal analysis. *Journal of Applied Developmental Psychology*, 29(6):434–445, 2008.
- M. A. Stinson, C. R. Bhat, et al. A comparison of the route preferences of experienced and inexperienced bicycle commuters. In *TRB 84th Annual Meeting Compendium of Papers*, number 05-1434, 2005.
- J. Stragier, M. V. Abeele, P. Mechant, and L. De Marez. Understanding persistence in the use of online fitness communities: comparing novice and experienced users. *Computers in Human Behavior*, 64:34–42, 2016.
- Strava. Strava just hit a huge activity milestone. http://womensrunning.competitor.com/2017/05/news/strava-one-billion-activities_75359, 2017.
- K. H. Sung and M. J. Lee. Do online comments influence the public’s attitudes toward an organization? effects of online comments based on individuals’ prior attitudes. *The Journal of psychology*, 149(4):325–338, 2015.
- R. Teodoro and M. Naaman. Fitter with twitter: Understanding personal health and fitness activity in social media. In *ICWSM*, 2013.

- M. Thelwall, D. Wilkinson, and S. Uppal. Data mining emotion in social network communication: Gender differences in myspace. *Journal of the Association for Information Science and Technology*, 61(1):190–199, 2010.
- T. N. Y. Times. What is your city’s twin? <https://www.nytimes.com/interactive/2018/04/03/upshot/what-is-your-citys-twin.html>, 2018.
- F. A. Treiber, T. Baranowski, D. S. Braden, W. B. Strong, M. Levy, and W. Knox. Social support for exercise: relationship to physical activity in young adults. *Preventive medicine*, 20(6):737–750, 1991.
- A.-L. B. Ulseth and Ø. Seippel. Fitness, class and culture: Social inequality in fitness. In *The Nordic Sport Science Forum*, volume 11, pages 1–17. Citeseer, 2011.
- J. M. Vaterlaus, E. V. Patten, C. Roche, and J. A. Young. # gettinghealthy: The perceived influence of social media on young adult health behaviors. *Computers in Human Behavior*, 45:151–157, 2015.
- R. Vilhjalmsson and G. Kristjansdottir. Gender differences in physical activity in older children and adolescents: the central role of organized sport. *Social science & medicine*, 56(2):363–374, 2003.
- S. N. Walker, K. R. Sechrist, and N. J. Pender. The health-promoting lifestyle profile: development and psychometric characteristics. *Nursing research*, 36(2):76–81, 1987.
- J. B. Wang, J. K. Cataldo, G. X. Ayala, L. Natarajan, L. A. Cadmus-Bertram, M. M. White, H. Madanat, J. F. Nichols, and J. P. Pierce. Mobile and wearable device features that matter in promoting physical activity. *Journal of mobile technology in medicine*, 5(2):2, 2016.
- J. H. West, P. C. Hall, C. L. Hanson, M. D. Barnes, C. Giraud-Carrier, and J. Barrett. There’s an app for that: content analysis of paid health and fitness apps. *Journal of medical Internet research*, 14(3), 2012.

- WHO. *Obesity: preventing and managing the global epidemic*. Number 894. World Health Organization, 2000.
- WHO et al. Fair play: building a strong physical activity system for more active people. In *Fair play: building a strong physical activity system for more active people*. 2021.
- R. R. Wing and R. W. Jeffery. Benefits of recruiting participants with friends and increasing social support for weight loss and maintenance. *Journal of consulting and clinical psychology*, 67(1):132, 1999.
- K. M. Winters-Stone, K. S. Lyons, L. M. Nail, and T. M. Beer. The exercising together project: Design and recruitment for a randomized, controlled trial to determine the benefits of partnered strength training for couples coping with prostate cancer. *Contemporary clinical trials*, 33(2):342–350, 2012.
- K. Wright. Perceptions of on-line support providers: An examination of perceived homophily, source credibility, communication and social support within on-line support groups. *Communication Quarterly*, 48(1):44–59, 2000.
- D. R. Young, W. L. Haskell, C. B. Taylor, and S. P. Fortmann. Effect of community health education on physical activity knowledge, attitudes, and behavior: The stanford five-city project. *American journal of epidemiology*, 144(3):264–274, 1996.
- L. Zeng, Z. W. Almquist, and E. S. Spiro. Let’s workout! exploring social exercise in an online fitness community. *iConference 2017 Proc. Vol. 2*, 2017.
- L. Zeng, Z. W. Almquist, and E. S. Spiro. Stay connected and keep motivated: Modeling activity level of exercise in an online fitness community. In *International Conference on Social Computing and Social Media*, pages 137–147. Springer, 2018.
- J. Zhang, D. Brackbill, S. Yang, and D. Centola. Efficacy and causal mechanism of an online social media intervention to increase physical activity: results of a randomized controlled trial. *Preventive medicine reports*, 2:651–657, 2015.

S. Zhao. Toward a taxonomy of copresence. *Presence: Teleoperators & Virtual Environments*, 12(5):445–455, 2003.