Online Actions with Offline Impact: How Online Social Networks Influence Online and Offline User Behavior

Althoff, T., Jindal, P., & Leskovec, J. (2016)

Master-Seminar Internet of People: Connectivity, Mobility and Privacy (IN2107, IN4962)

Garching, 13. June 2019

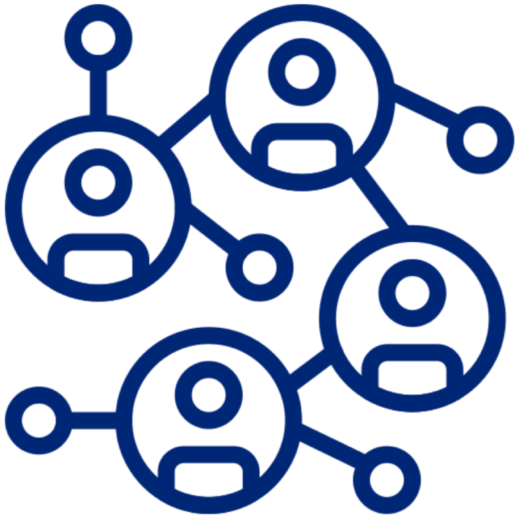
THE INFLUENCE OF SOCIAL NETWORKS ON USER BEHAVIOR REMAINS ELUSIVE

Social networks are everywhere.

In contrast to previous research this study avoids the **bias of self-reporting**.

Estimating effects on user behavior is difficult due to many unobserved factors and **selection effects** often occur.

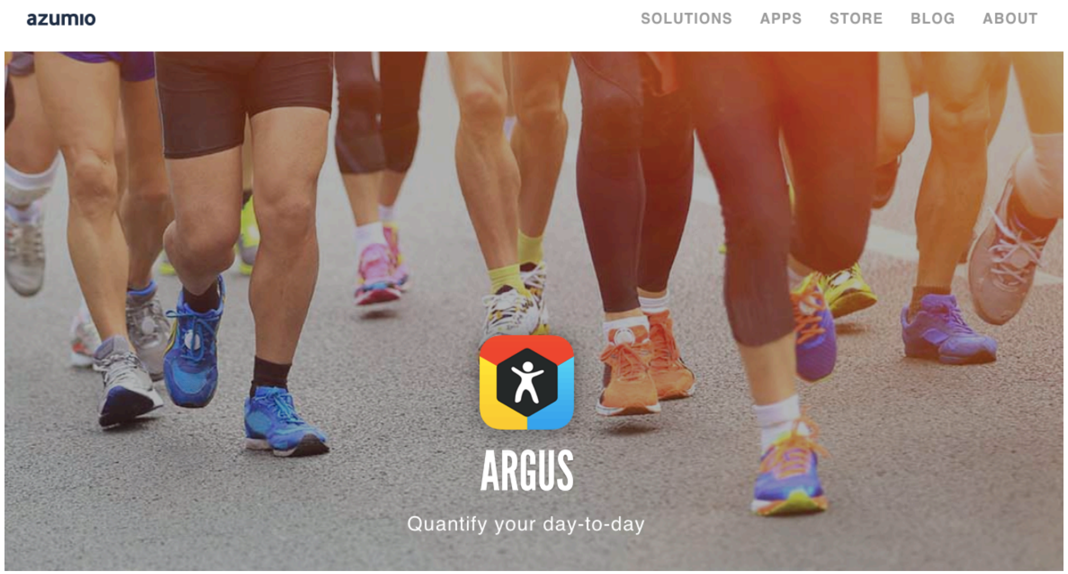
This paper studies **user behavior** in smartphone physical activity tracking application, observing in-app online engagement and offline real-world physical activity through the smartphone accelerometer.



THE LARGEST* DATASET ON HUMAN ACTIVITY TRACKING AND SOCIAL NETWORK INTERACTIONS TO DATE

Data from the Azumio Argus smartphone app

- Tracks exercise and physical activity of **6 million users** from over 100 different countries.
- Over a time period of **5 years** (January 2011 and January 2016)
- **631 million self-reported activity posts** (including running, walking, sleep, heart rate, yoga, cycling, weight, etc.)
- **160 million days of steps tracking** (objectively measured through the smartphone accelerometers)



*ten thousand times more users and a million times more activity tracking than comparable studies

THE SOCIAL CONNECTIONS INCLUDE NOTIFICATIONS, ACTIVITY FEED AND COMMENTS

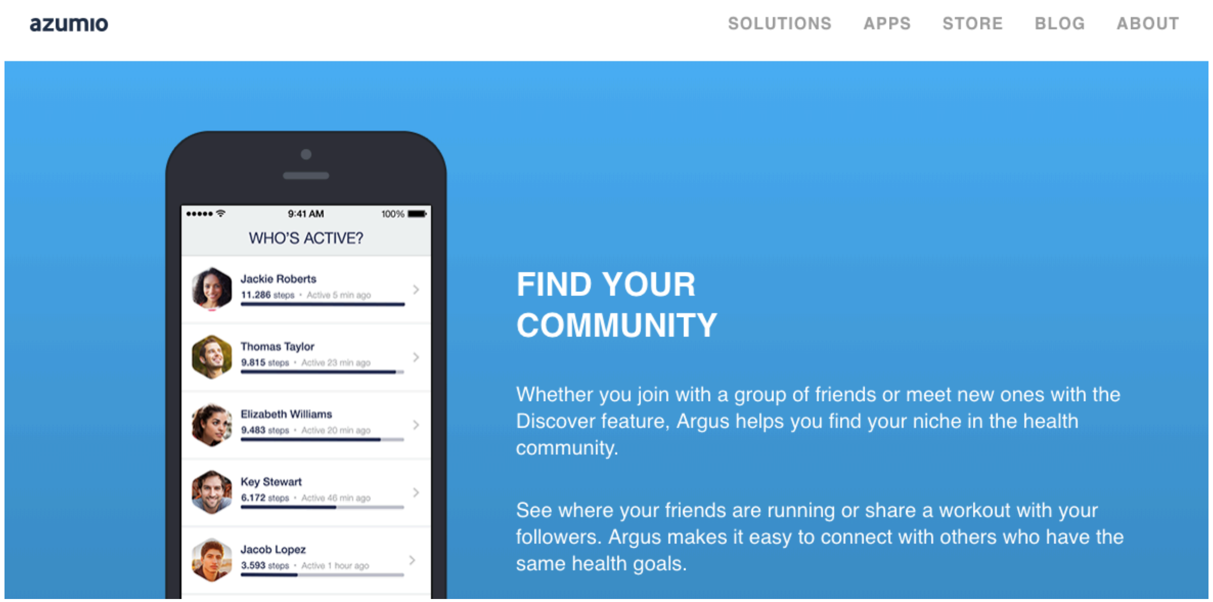
The social network in this study has **two types of connections**.

- **bi-directional friend connections** (after approval of a friend request by the receiver)
- **uni-directional follower connections** (without need for approval).

All edges in the network was created **organically** without friend recommendation algorithms.

Physical activity (offline behavior) is defined as number of accelerometer-defined steps

In-app activity (online behavior) is defined as number of posts the user creates within the app each corresponding to a self-reported action such as running, cycling or sleeping

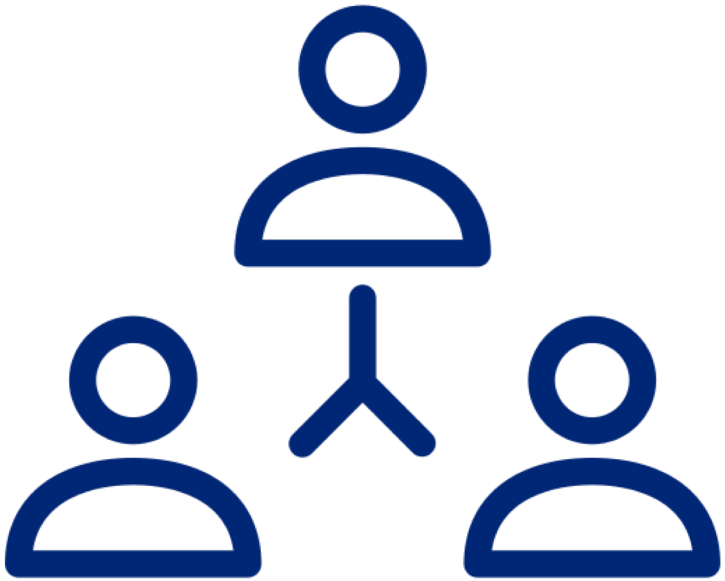


THE INTRODUCTION OF A SOCIAL NETWORK ALLOWS FOR QUANTIFYING THE CAUSAL EFFECT

After 3 years (Nov. 2013), the app introduces an internal **social network**.

The data allows for **quantify the causal effect of the social network** on user behavior by using it as a natural experiment on delayed social network edge formation.

Distinguish the **causal effect of social influence** from the simultaneous increase in motivation of the user to use the app (i.e., a **selection effect**).



USER ACTIVITY SIGNIFICANTLY INCREASES AFTER EACH EDGE CREATION

We see that the **activity level increases** after a new friend connection.

Just an **unobservable motivation** boost?

Or is it due to the **social influence** of a new online friend?

To estimate the effect of the social network, it is crucial to disentangle the **selection effect** of intrinsically motivated users who send friend requests and the **social network effect** of the new connection.

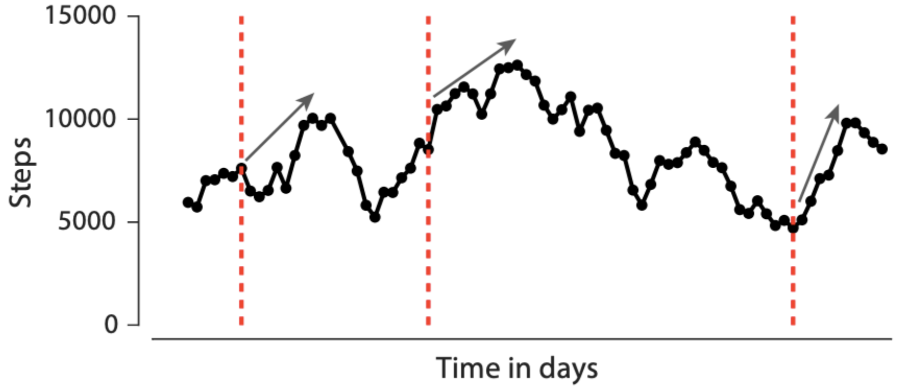


Figure 1: Time series of daily steps for an example user. Dashed vertical lines correspond to edge creations. We observe significant increase in activity after each created edge (arrows).

A NATURAL EXPERIMENT OF DELAYED ACCEPTANCE OF EDGE CREATIONS

Method: Only look at the **sender** of the friend requests and **activity 7 days before and after** sending the request.

A person that sends out a friend request is motivated M.

If the the friend request is **accepted directly** (within a day) the change in behavior (increase in steps) is due to both the **intrinsic motivation M and the social influence I**.

$$B_{\text{after}}^{\text{direct}} - B_{\text{before}}^{\text{direct}} = M + I$$

The **delayed acceptance** of friend requests does not have the social influence within the 7 days and the behavior change can therefore only be attributed to the **motivation M**.

$$B_{\text{after}}^{\text{delayed}} - B_{\text{before}}^{\text{delayed}} = M$$

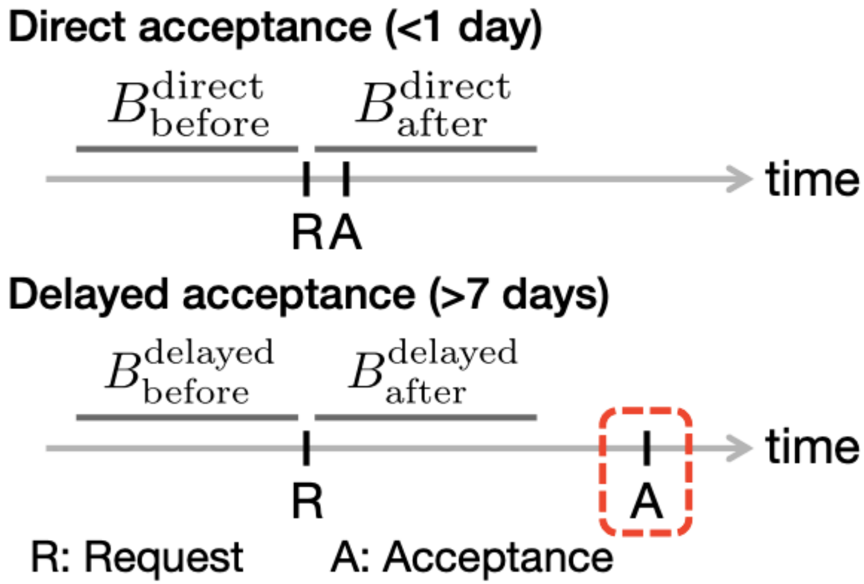


Figure 2: Conceptual framework for distinguishing intrinsic motivation from social influence in edge creations.

SOCIAL INFLUENCE ESTIMATED BY DIFFERENCE-IN-DIFFERENCE ANALYSIS

Average total step **increase of 328 daily steps** for directly accepted requests.

Delayed accepted requests lead to 148 additional daily steps. **Motivation explains 45%** of the observed effect for directly accepted requests.

The remaining **55%** or 180 daily steps can be attributed to **social influence**.

To make sure that the acceptance (delayed or direct) of the friend request is random, a balance check for the two groups is done (i.e. standardized mean difference is low).

$$(B^{\text{direct}}_{\text{after}} - B^{\text{direct}}_{\text{before}}) - (B^{\text{delayed}}_{\text{after}} - B^{\text{delayed}}_{\text{before}}) = (M + I) - (M) = I$$

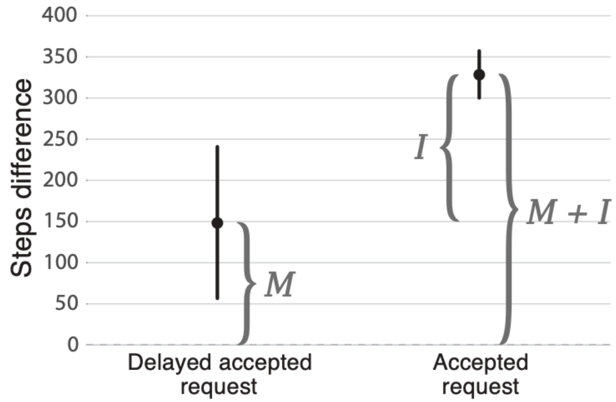


Figure 3: Steps difference after the time of friendship request for delayed accepted and directly accepted friendship requests.

ESTIMATING THREE EFFECTS OF JOINING A SOCIAL NETWORK THROUGH MATCHED USERS

The previous section estimated the **effect of an average edge** in the network.

This section focuses on the **first edge** i.e. joining the social network.

Compare the treatment group (users who join the social network) to a **matched control group**.

The control user is selected by **critical constraints** on time of sign-up, activity before and same activity on the day of joining.

Physical activity (number of steps)

User engagement (number of posts in the app)

User retainment (likelihood of continuing using the app)

CREATING A FIRST EDGE SIGNIFICANTLY BOOST ACTIVITY UP TO THREE MONTHS

We observe a **significant boost in activity** of 406 additional daily steps in treatment users that **diminishes over 20 weeks** but no difference in control users.

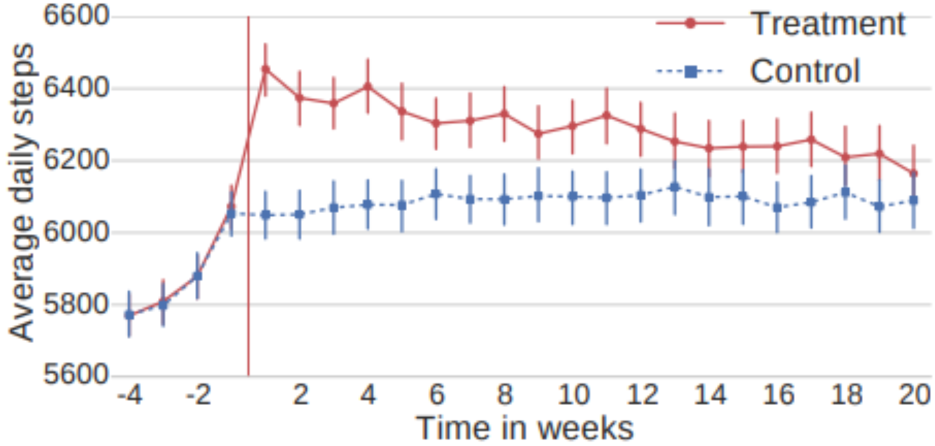


Figure 4: Average daily steps for users that do join the social network at time zero (treatment; red) and matched users that do not (control; blue).

SOCIAL NETWORK USERS ARE MORE LIKELY TO KEEP USING THE APP AND CREATE POSTS

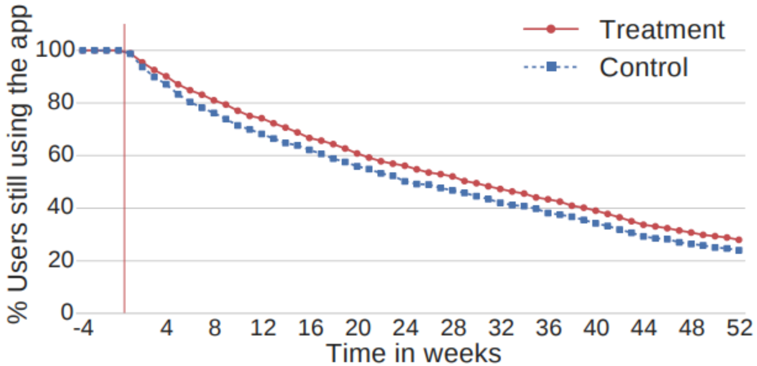


Figure 5: **Retention** of users that do join the social network at time zero (treatment; red) and matched users that do not (control; blue).

- **Increased likelihood** to keep using the activity tracking app during any of the following **52 weeks**.

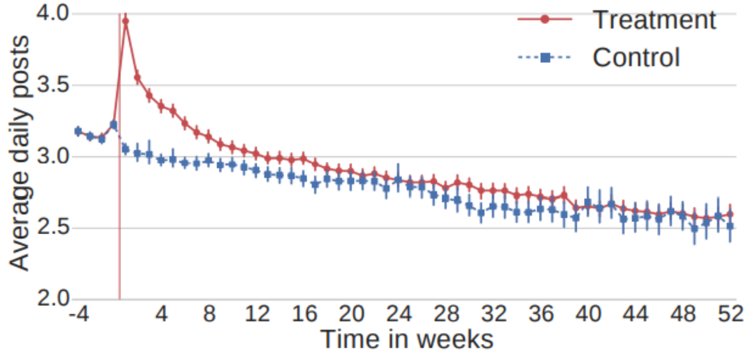


Figure 6: **App usage** of users that do join the social network at time zero (treatment; red) and matched users that do not (control; blue) among users still using the app in each week.

- **More posts created** than control users for a period of about **20 weeks** after joining the social network.

ADDITIONAL EDGES INCREASES PHYSICAL ACTIVITY, BUT THE EFFECT DECREASES

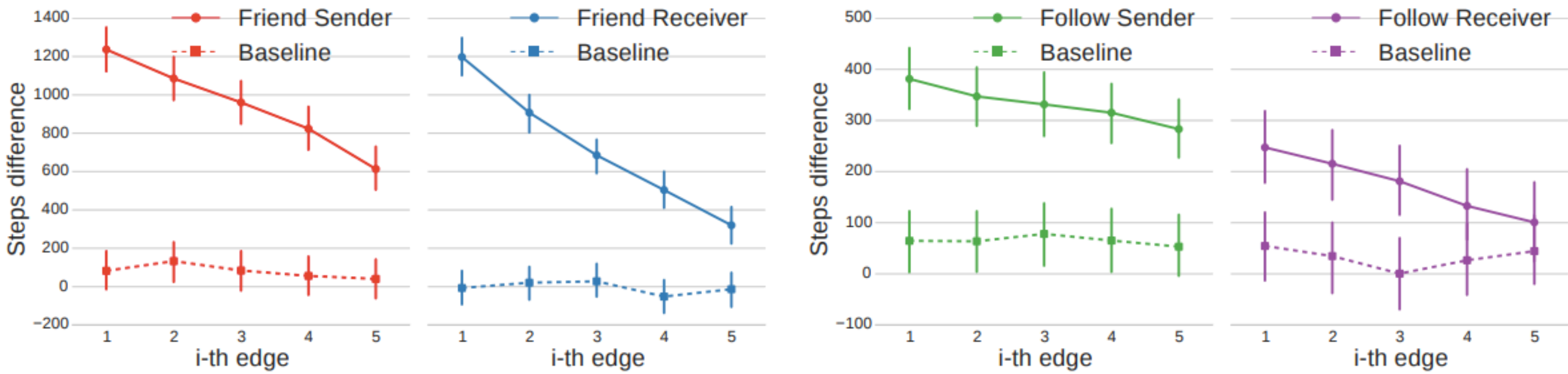


Figure 7: The average daily difference in steps 7 days before and 7 days after edge. Dashed lines show corresponding baselines.

- **Physical activity increases** after edges get created, but **decreasing effect sizes** with each additional edge
- **Larger effect for senders** compared to receivers, and **larger for friends** compared to followers.

ACTIVITY INCREASES BEFORE EDGE CREATION BUT VARIES LESS AFTER

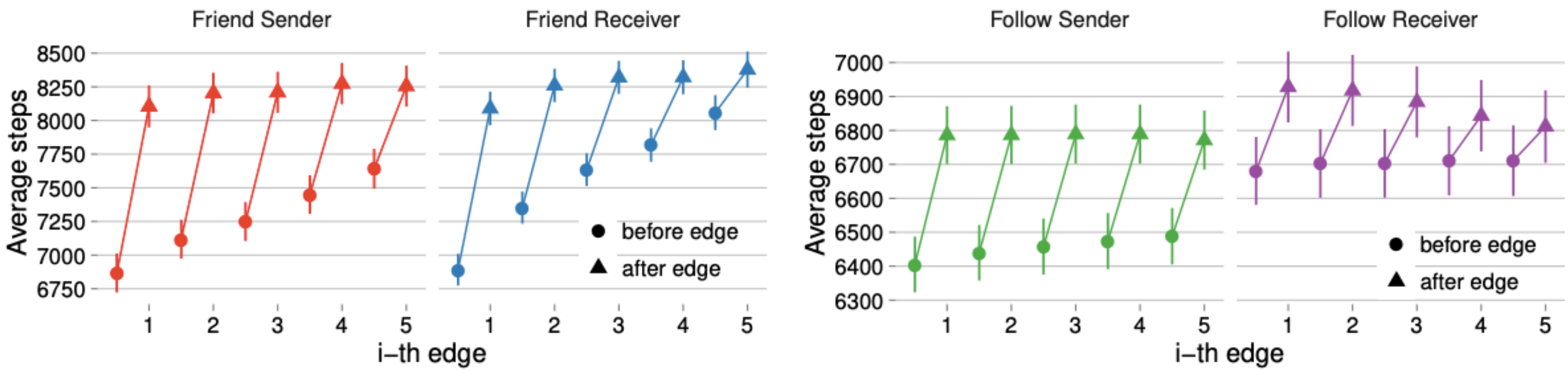


Figure 8: Steps before (circles) and after (triangles) edge creation.

- **Increasing activity levels before** edge creation (circles) while the **activity after edge creation varies less** (triangles)
- **Decreasing effect size** (smaller steps differences) at each edge.

USERS THAT ARE OLDER, HAVE HIGHER BMI, AND TAKE MORE STEPS HAVE LARGER CHANGES IN BEHAVIOR

1. **Older** (30-60)

2. **Higher BMI** for the receiver and opposite for the sender

3. Takes **more steps** in the week before

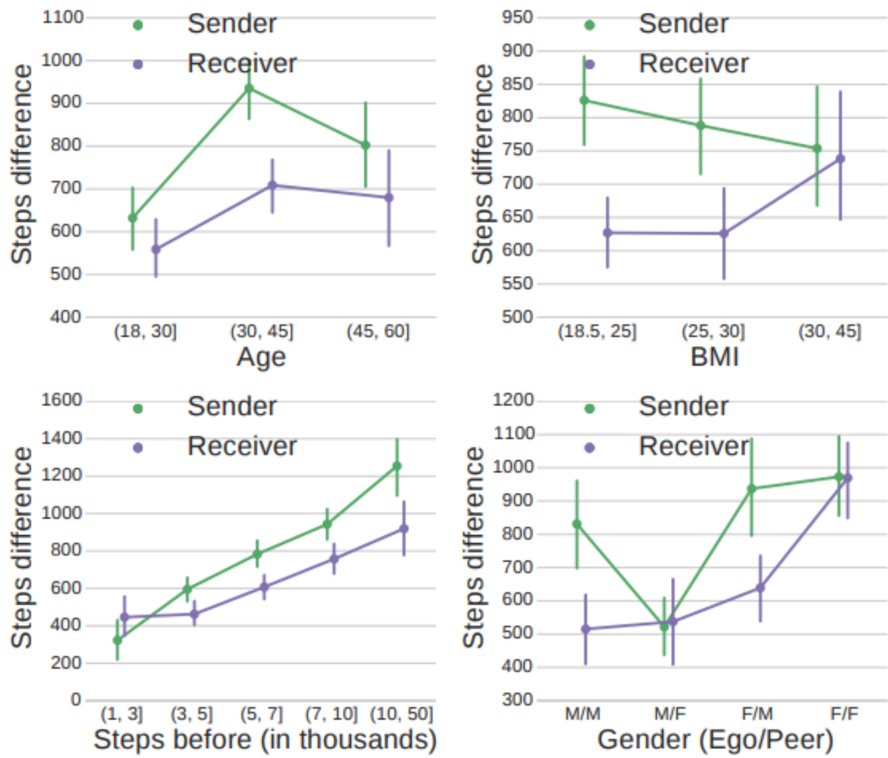


Figure 9: The difference of steps after edge creation

A DECISION TREE PREDICTS ACTIVITY INCREASE WITH HIGH ACCURACY BY COMBINING FEATURES

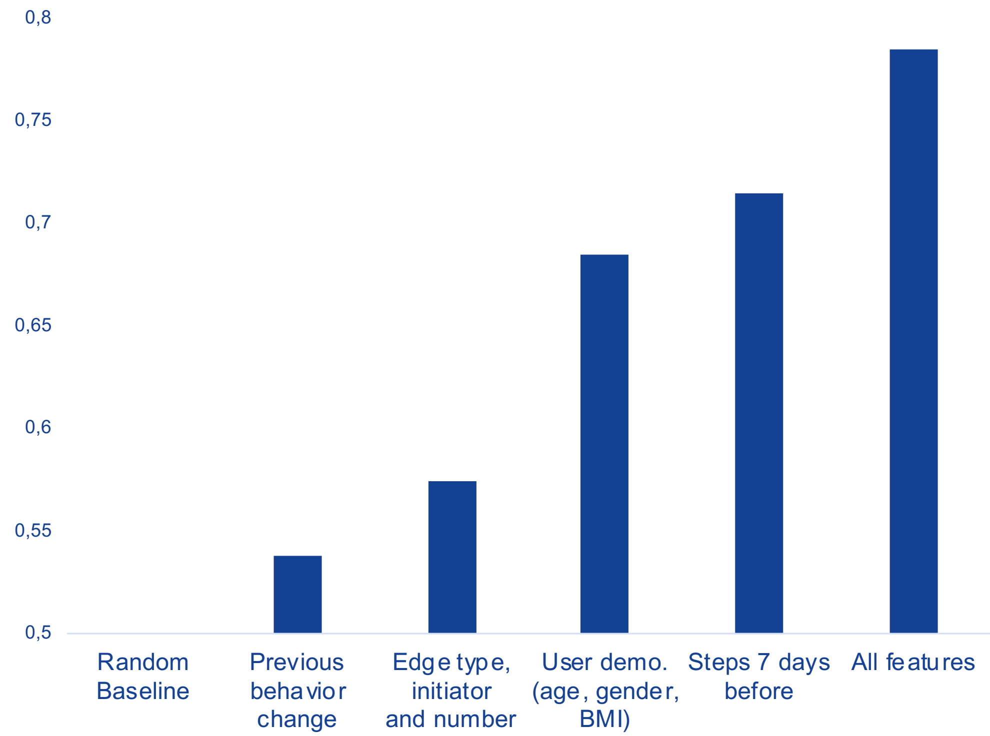
Gradient Boosted Tree models for different features and combined:
- Behavior change in steps after the previous edge creation
- Edge type, sender/receiver, number of edges
- User demographics (age, BMI, gender)
- Number of steps 7 days before

Good prediction for **user demographic** (0.685) and **activity level** the week before the edge formation (0.715)

Highest accuracy for **combining all features** (0.785).

All models were trained with 80% training data and number of threes, three depth and learning rate was optimized through cross-validation.

Accuracy of prediction model on full dataset



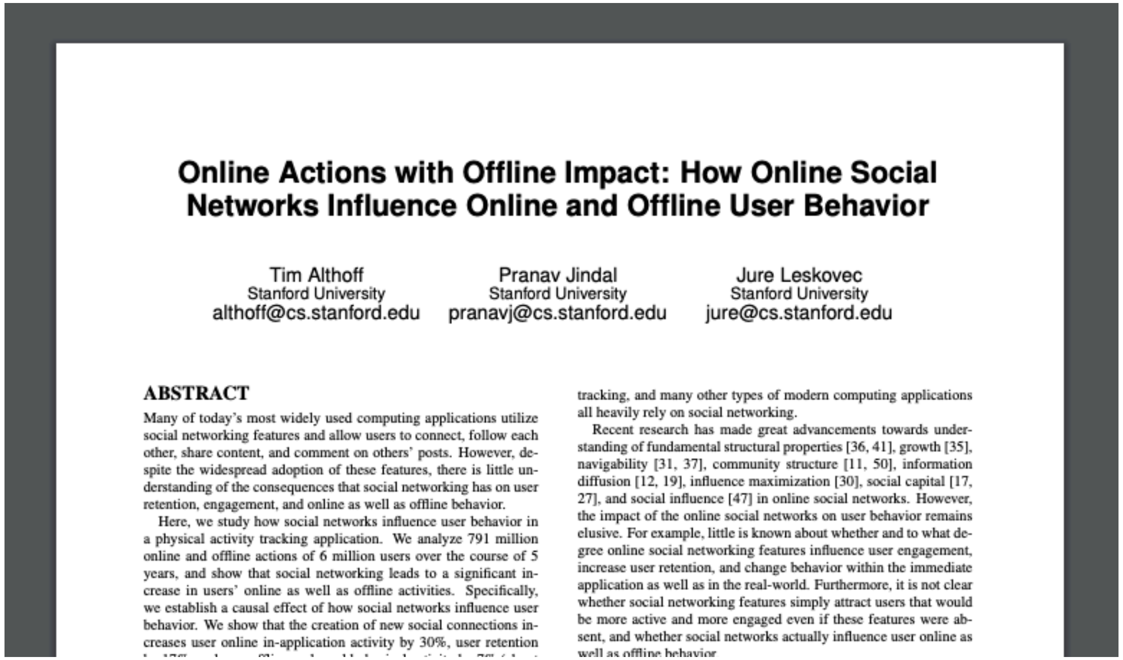
RESULTS

- **Social influence** explain **55%** of the observed average effect, while **45%** is due to **increased motivation**.
- Joining the social network has **significant positive effect** on online and offline user behavior that **diminish over time**.
- Social network users are **30% more engaged** in the app, **17% less likely to drop out** of the app within one year, and **7% more physically active** (~400steps/day) compared to a matched control group. These effects last over long periods of several months.
- Offline physical activity **temporarily increases** and the effect **diminish with each additional connection** and are **larger for friend connections** than follower connections.
- The average increases are **larger for the sender** than its receiver, and the **effect varies** with age, gender, weight, and prior physical activity level.
- Prediction models with the discovered insights can **predict with high accuracy** which users will be most influences by the creation of new social connections.

THE PAPER ADDRESSES SEVERAL ASPECTS OF SOCIAL NETWORKS IN A ROBUST WAY

- Studies on the **largest activity tracking dataset** to date
- Shows how **online social networks shape users behavior** such as user engagement, retention and real-world physical activity
- Employs **natural experi**ments, difference-in-difference models and matching-based observational studies to disentangle selection effects from causal social network effects

Robustness of the study, e.g., constraints on users to study or the standardized mean difference of the edge request groups



Online Actions with Offline Impact: How Online Social Networks Influence Online and Offline User Behavior

Tim Althoff
Stanford University
althoff@cs.stanford.edu

Pranav Jindal
Stanford University
pranavj@cs.stanford.edu

Jure Leskovec
Stanford University
jure@cs.stanford.edu

ABSTRACT

Many of today's most widely used computing applications utilize social networking features and allow users to connect, follow each other, share content, and comment on others' posts. However, despite the widespread adoption of these features, there is little understanding of the consequences that social networking has on user retention, engagement, and online as well as offline behavior.

Here, we study how social networks influence user behavior in a physical activity tracking application. We analyze 791 million online and offline actions of 6 million users over the course of 5 years, and show that social networking leads to a significant increase in users' online as well as offline activities. Specifically, we establish a causal effect of how social networks influence user behavior. We show that the creation of new social connections increases user online in-application activity by 30%, user retention

tracking, and many other types of modern computing applications all heavily rely on social networking.

Recent research has made great advancements towards understanding of fundamental structural properties [36, 41], growth [35], navigability [31, 37], community structure [11, 50], information diffusion [12, 19], influence maximization [30], social capital [17, 27], and social influence [47] in online social networks. However, the impact of the online social networks on user behavior remains elusive. For example, little is known about whether and to what degree online social networking features influence user engagement, increase user retention, and change behavior within the immediate application as well as in the real-world. Furthermore, it is not clear whether social networking features simply attract users that would be more active and more engaged even if these features were absent, and whether social networks actually influence user online as well as offline behavior

BACK-UP

THE DATASET IS LARGE AND ALLOWS FOR STUDYING OF HETEROGENOUS EFFECTS

Observation period	Jan. 2011 – Jan. 2016
Introduction of the social network	November 2013
# total users	6.0 million
# total online and offline activities	791 million
# activity posts (online engagement)	631 million
# users tracking steps	2.0 million
# days of steps tracking (offline activity)	160 million
# total steps tracked	824 billion
# users in the social network	211,383
# edges in the social network	563,007
Median age	33 years
% users female	46.1%
% underweight (BMI < 18.5)	4.7%
% normal weight ($18.5 \leq$ BMI < 25)	44.2%
% overweight ($25 \leq$ BMI < 30)	30.2%
% obese ($30 \leq$ BMI)	20.9%

Table 1: Dataset statistics.

BMI refers to body mass index.

THE VARIABLES FOR THE DIRECT AND DELAYED USER GROUPS ARE BALANCED

Group	Variable	SMD
Sender	Age	0.092
	Age NA	0.071
	Gender	0.115
	Gender NA	0.105
	BMI	0.052
	BMI NA	-0.013
	Steps 7 days before	0.034
	Days tracked 7 days before	-0.005
Receiver	Age	-0.004
	Age NA	-0.061
	Gender	-0.026
	Gender NA	-0.028
	BMI	-0.034
	BMI NA	-0.065
	Steps 7 days before	0.074
	Steps 7 days before NA	-0.002
	Days tracked 7 days before	-0.014
Relationship	#Mutual friends at request	0.080
Timing	Edge number for sender	0.049
	Edge number for receiver	0.098
	#Days on social network for sender	0.109
	#Days on social network for receiver	0.179
	Median Absolute SMD	0.057
	Maximum Absolute SMD	0.179

Table 2: Balancing statistics on relevant covariates for the natural experiment.

- Covariates with absolute SMD lower than 0.25 are considered balanced.
- NA refers to missingness indicator.

- #Days on social network refers to the number of days between the first created edge and the friendship request.

A DECISION TREE PREDICTS ACTIVITY INCREASE WITH HIGH ACCURACY BY COMBINING FEATURES

1. Random Baseline: Included for comparison.
2. Previous behavior change: Activity increase or decrease in steps after the most recent edge creation of the same type and initiator (note that this is not available for anyone's first edge). We only use previous edges that were created at least 7 days prior to the current edge because otherwise this feature could give away the true label for the current edge.
3. Edge type (friend vs follow), edge initiator (sender vs receiver) and edge number
4. User demographics: Age, gender, and BMI.
5. Steps before: Average number of steps in the 7 day window before edge creation.
6. All features: Combination of models 2-5

	Model	**All**	**Fr/S**	**Fr/R**	**Fo/S**	**Fo/R**
1	None	0.500	0.500	0.500	0.500	0.500
2	Previous behavior change	0.538	0.543	0.551	0.546	0.526
3	Edge type, initiator and number	0.574	0.515	0.518	0.506	0.510
4	User demographics (age, gender, BMI)	0.685	0.644	0.583	0.777	0.773
5	Steps 7 days before	0.715	0.665	0.614	0.808	0.781
6	All features (2-5)	**0.785**	**0.721**	**0.672**	**0.847**	**0.830**

Table 3: Performance of several models predicting activity increase or decrease after edge creation.

- The table reports predictive performance on all data (all), and split by edge type and initiator: Friend (Fr), Follow (Fo), Sender (S), Receiver (R).