Measuring Facebook Use: The Accuracy of Self-Reported Data Versus Digital Trace Data

Paulina Pankowska[1], Florian Keusch[2], Ruben Bach[2], and Alexandru Cernat[3]

[1] Utrecht University, The Netherlands

[2] University of Mannheim, Germany

[3] University of Manchester, UK

WEB DATA OPP Workshop

18-19 March 2024

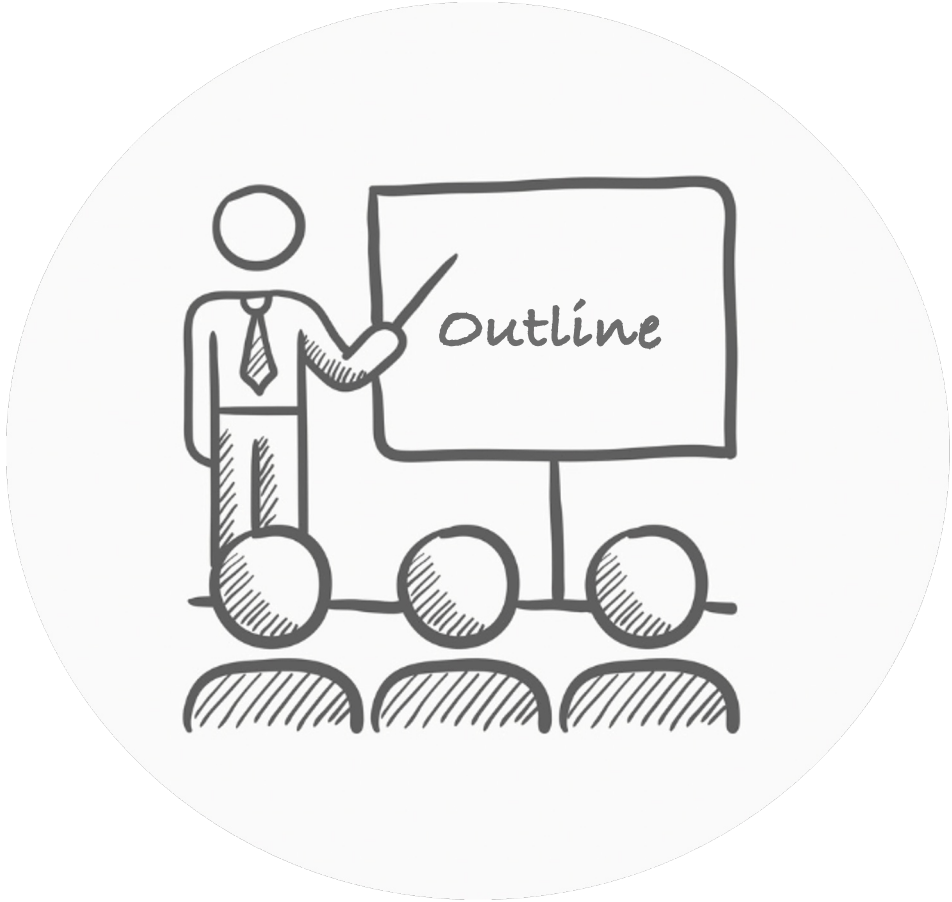
RECSM-Universitat Pompeu Fabra

Utrecht University

UNIVERSITY OF MANNHEIM

MANCHESTER 1824

The University of Manchester

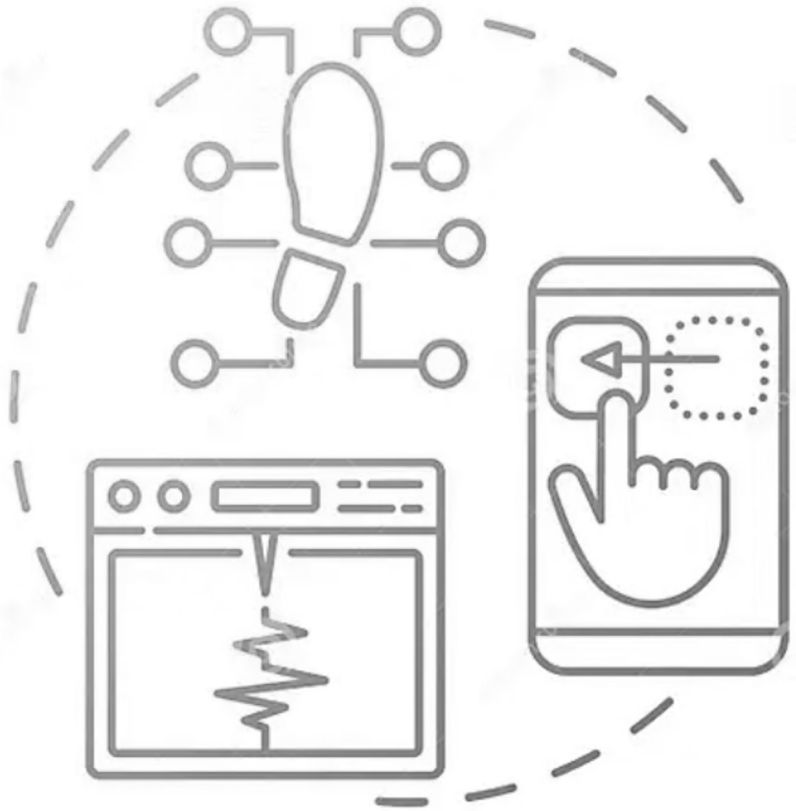


Background

- Increased use of digital trace data in social sciences

Our research

- Comparing digital trace and survey measures of FB use
- Using longitudinal survey, tracking apps, and FB donated data



- Recent increased interest in the social sciences
- Considered an attractive alternative to surveys
  - Superior quality
  - Allow to overcome limitations
- Have limitations that are often not addressed
  - Can lead to biased results and incorrect findings

- High quality data key in this context
  - integral part of modern society
- Digital traces particularly attractive
  - Directly observing online behaviors
  - Via tracking apps or donated data
- Limitations are getting acknowledged
  - In practice not much is done
  - Problematic

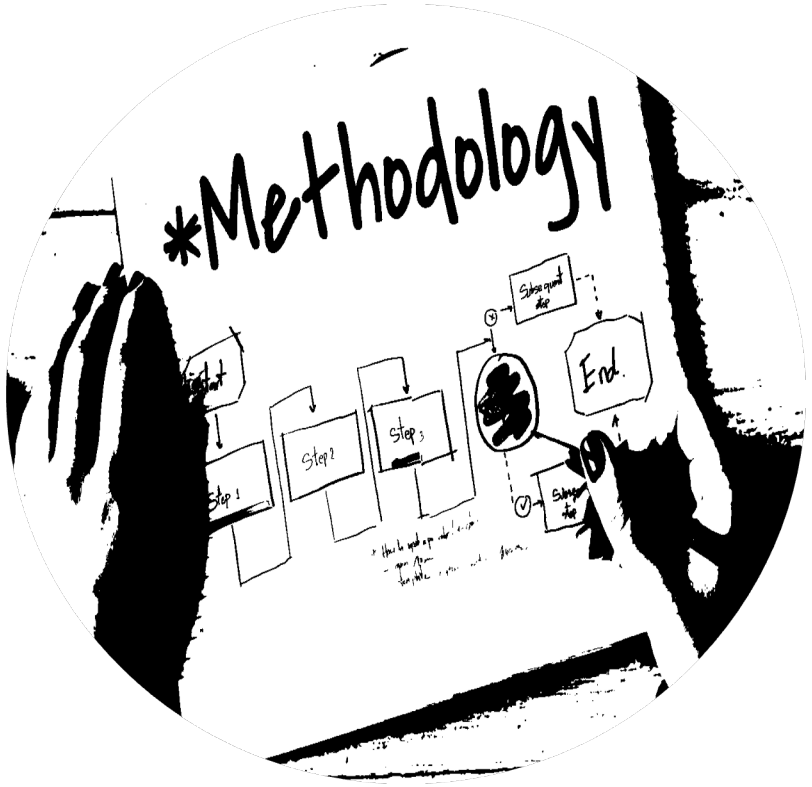
Compare digital trace and survey measures of FB use

- Allow errors in both sources
- Do not treat any source as ’gold standard’

- Non-probability online panel in Germany
  - N=2,100
  - Survey with three waves around September 2021
  - Tracking app installed on computer and/or mobile device
  - FB donated data (for a sub-sample)

- Survey – 5-point Likert scale asking about frequency of use

- Tracking app – harmonizing to fit the survey question
  - Three 10-day periods corresponding to survey waves
  - Summing number of days in which FB was used on mobile & PC in each period
  - Categorizing the usage variable in line with survey question



Hidden Markov models (HMMs)

- Latent class models used to correct for measurement error in categorical, longitudinal data
- Do not require any data source to be error-free
- Use repeated measures of indicator(s) to extract information about the error from the data
- Multiple-indicator HMMs correct for error in all sources simultaneously

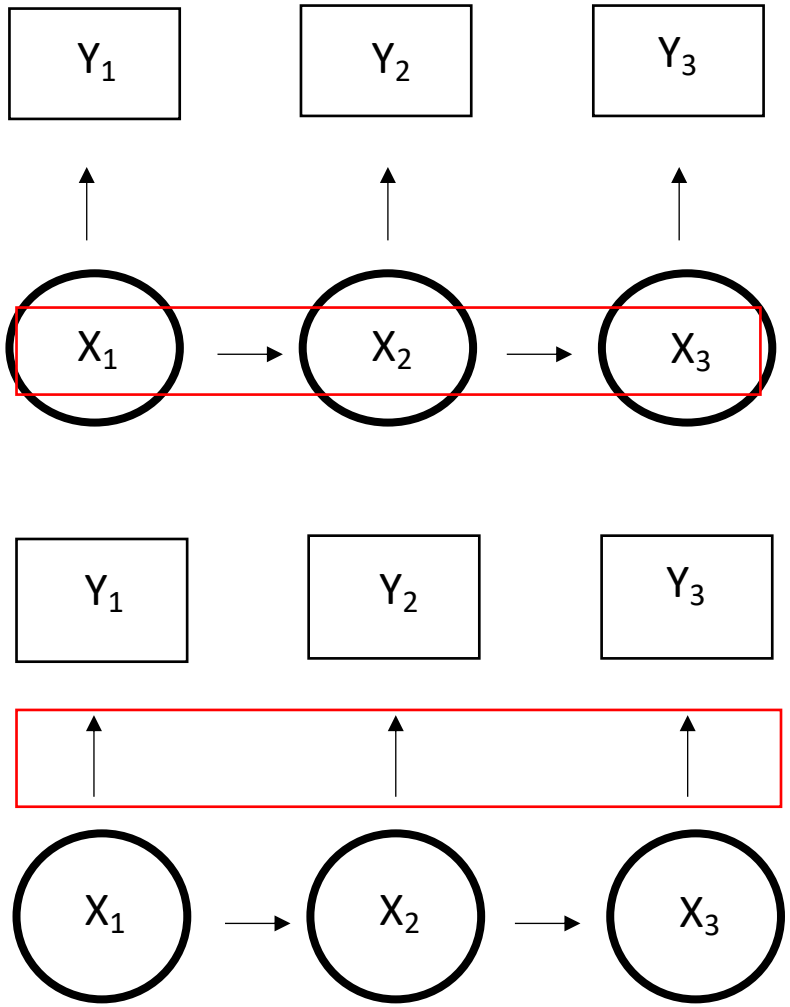
The standard HMM

Markov assumption

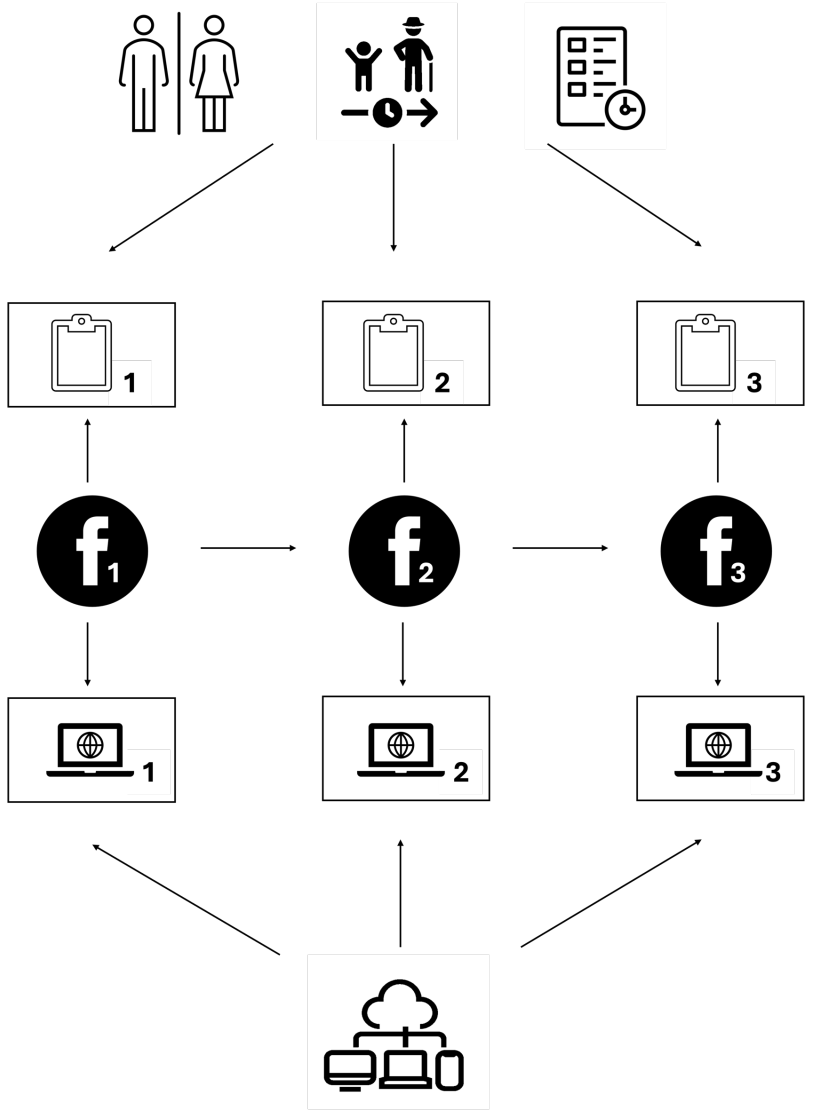
true state at time *t* only depends on true state at time *t-1*

Local independence assumption

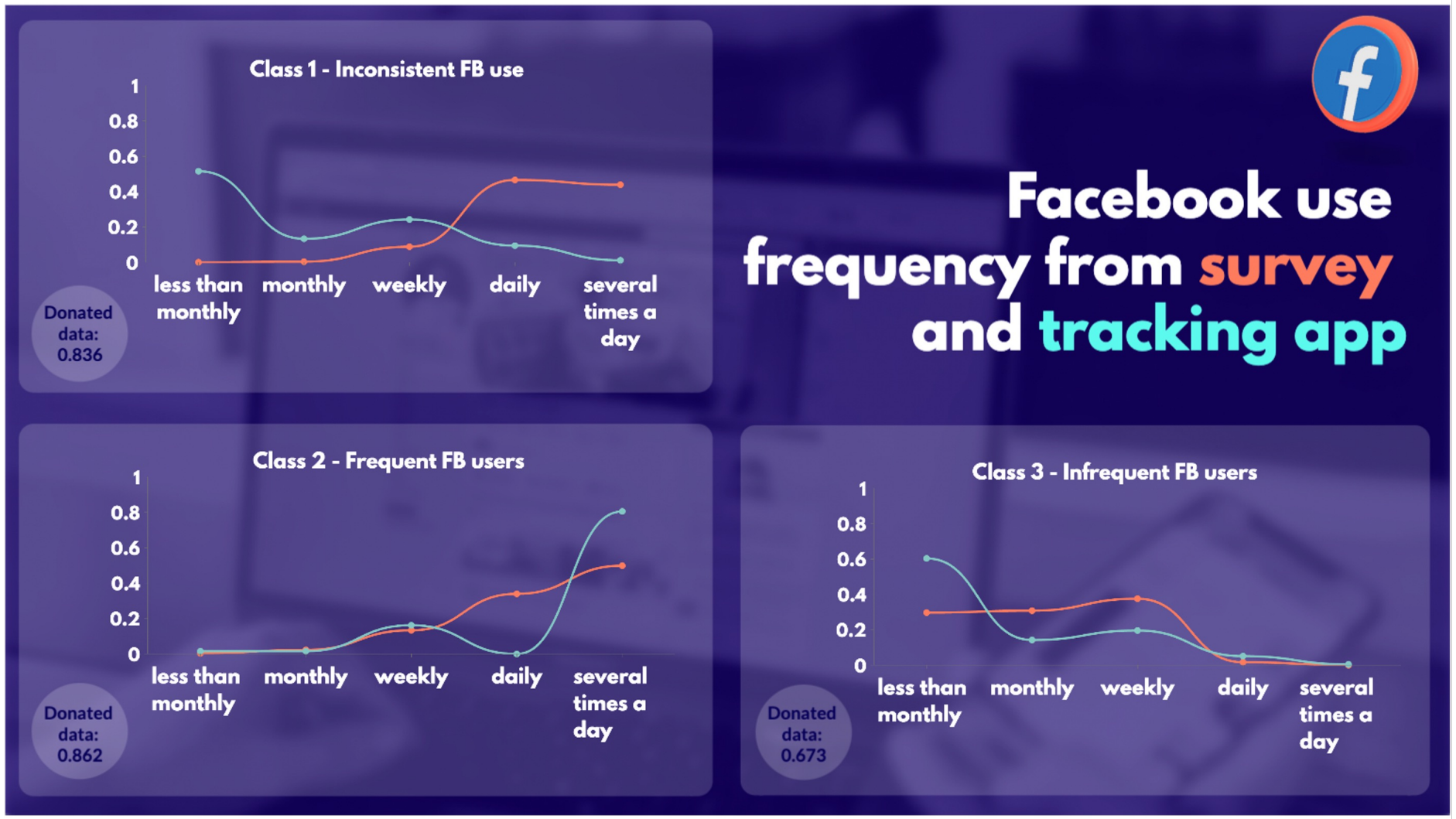
observed state at time *t* only depends on true state at time *t*



Our model: two-indicator HMM



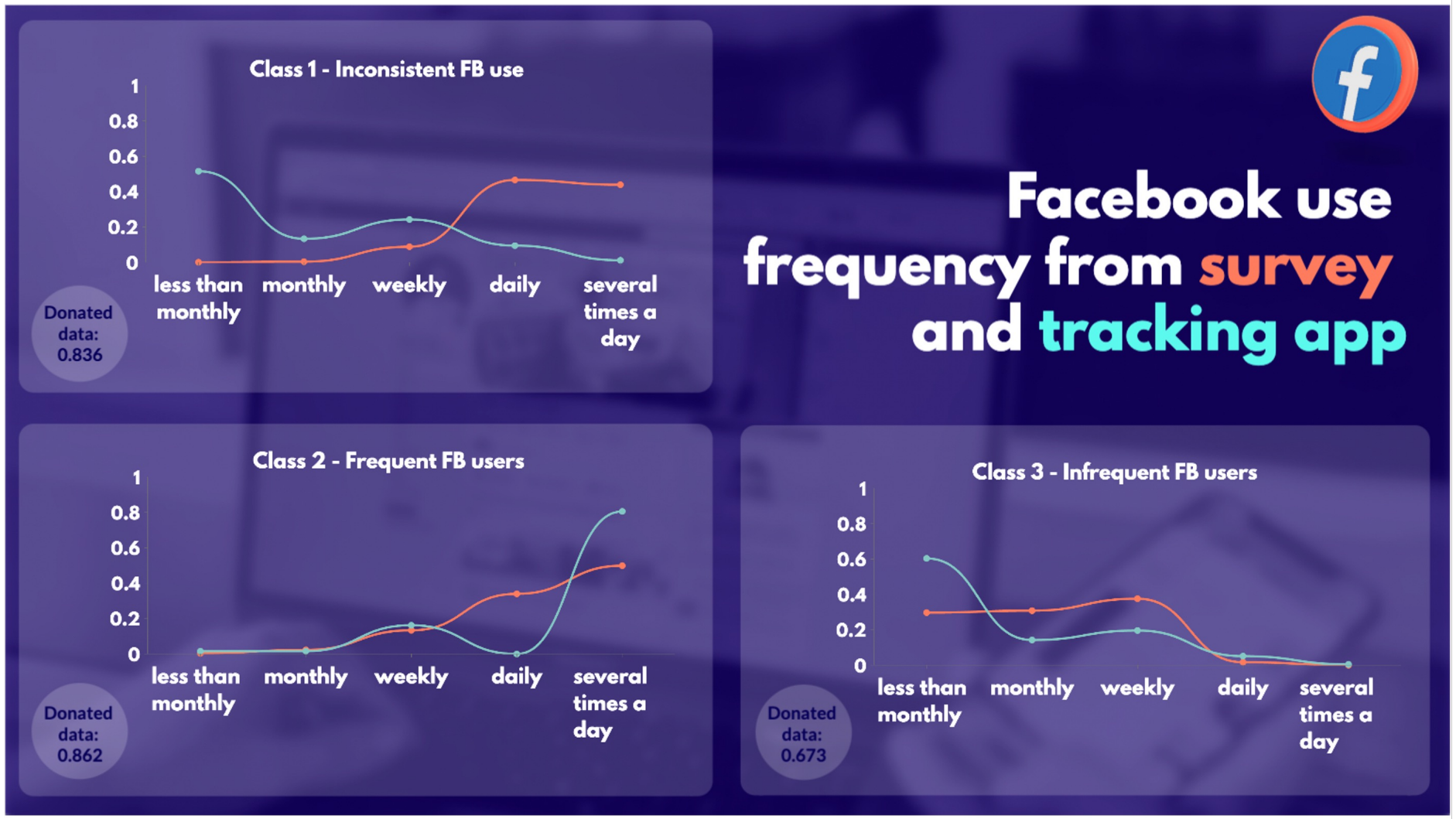
- Model with 3 FB use classes fits data well
- Both sources measure frequent (32.8%) and infrequent users relatively accurately (29.5%)
- One class (37.7%) is characterized by large inconsistencies between the two sources

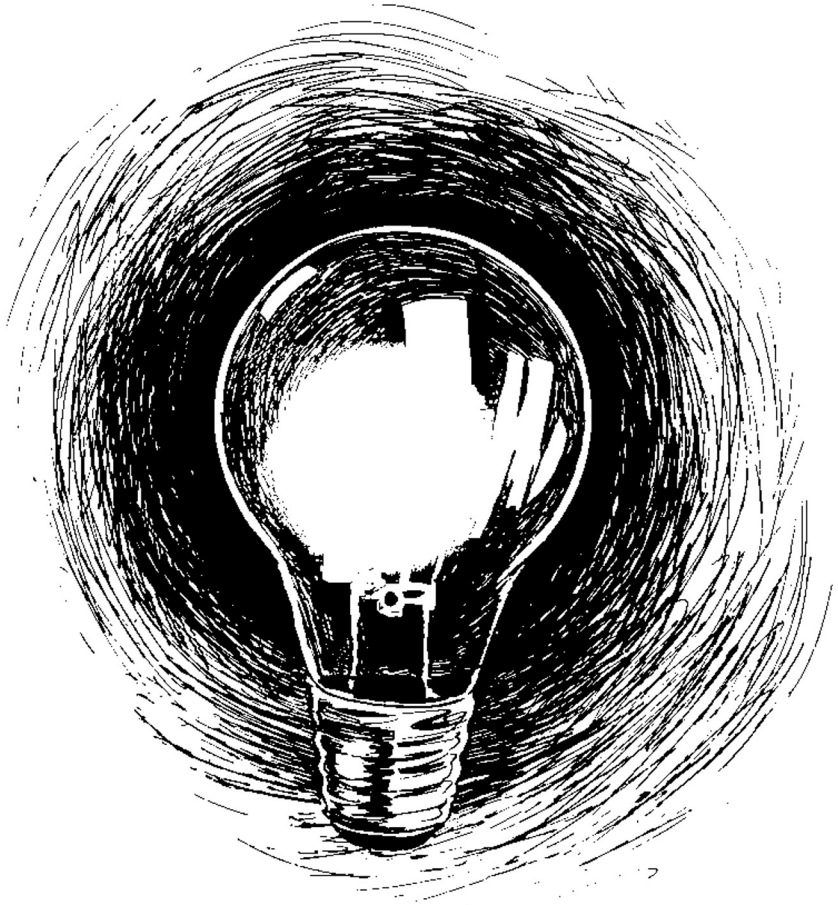

Class 1 - Inconsistent FB use
1
0.8
0.6
0.4
0.2
0
less than monthly
monthly
weekly
daily
several times a day
Donated data: 0.836
Facebook use frequency from survey and tracking app
Class 2 - Frequent FB users
1
0.8
0.6
0.4
0.2
0
less than monthly
monthly
weekly
daily
several times a day
Donated data: 0.862
Class 3 - Infrequent FB users
1
0.8
0.6
0.4
0.2
0
less than monthly
monthly
weekly
daily
several times a day
Donated data: 0.673

- Less likely to belong if both phone and PC are tracked ($\beta = -0.13$, $p < 0.05$)
- More likely to belong if own more FB compatible devices ($\beta = 0.14$, $p < 0.05$)
- Immigration background, education level, and political interest affect probability of belonging

Results confirmed by FB donated data

- Available for ±15% of sample (~ balanced distribution across classes)
- Downloaded by respondents from own FB accounts
- Used to estimate proportion of days in which FB was used around same time period


Class 1 - Inconsistent FB use
1
0.8
0.6
0.4
0.2
0
less than monthly
monthly
weekly
daily
several times a day
Donated data: 0.836
Facebook use frequency from survey and tracking app
Class 2 - Frequent FB users
1
0.8
0.6
0.4
0.2
0
less than monthly
monthly
weekly
daily
several times a day
Donated data: 0.862
Class 3 - Infrequent FB users
1
0.8
0.6
0.4
0.2
0
less than monthly
monthly
weekly
daily
several times a day
Donated data: 0.673



Conclusions

- Despite all their benefits, digital trace data are not perfect
  - E.g., they can underestimate behaviors, such as FB use, if not all devices are tracked
- We should take advantage of the good aspects, but also acknowledge the bad ones
- When using digital trace, we should account for measurement error

Funding for this study come from the *Baden-Württemberg Foundation* through the grant “Filter Bubbles, Alternative News and Political Polarization” (PI: R. Bach) and the *German Science Foundation (DFG)* through the grant “Evaluating Data Sources for Research into Political Reforms: (Non)probability Online Surveys and Big Data” (PIs: A. Blom, F. Keusch, & F. Kreuter).

We thank Ella Häcker and Leonard Bek for research assistance

Illustrations taken from freepik.com and 123rf.com

Dr Paulina Pankowska | Assistant Professor | Utrecht University | Sociology
p.k.pankowska@uu.nl | https://www.uu.nl/medewerkers/PKPankowska

Utrecht University