



INSTITUT FÜR ARBEITSMARKT- UND
BERUFSFORSCHUNG

Die Forschungseinrichtung der Bundesagentur für Arbeit

Methodische Herausforderungen bei der Aufbereitung und Auswertung von Smartphone- Daten zur Messung sozialer Interaktion

Sonja Malich ¹, Sebastian Bähr ¹, Georg-Christoph Haas ^{1, 2},
Florian Keusch ², Frauke Kreuter ^{3, 4}, Mark Trappmann ^{1, 5}

¹ Institut für Arbeitsmarkt- und Berufsforschung

² Universität Mannheim

³ University of Maryland, USA

⁴ Ludwig-Maximilians-Universität, München

⁵ Universität Bamberg

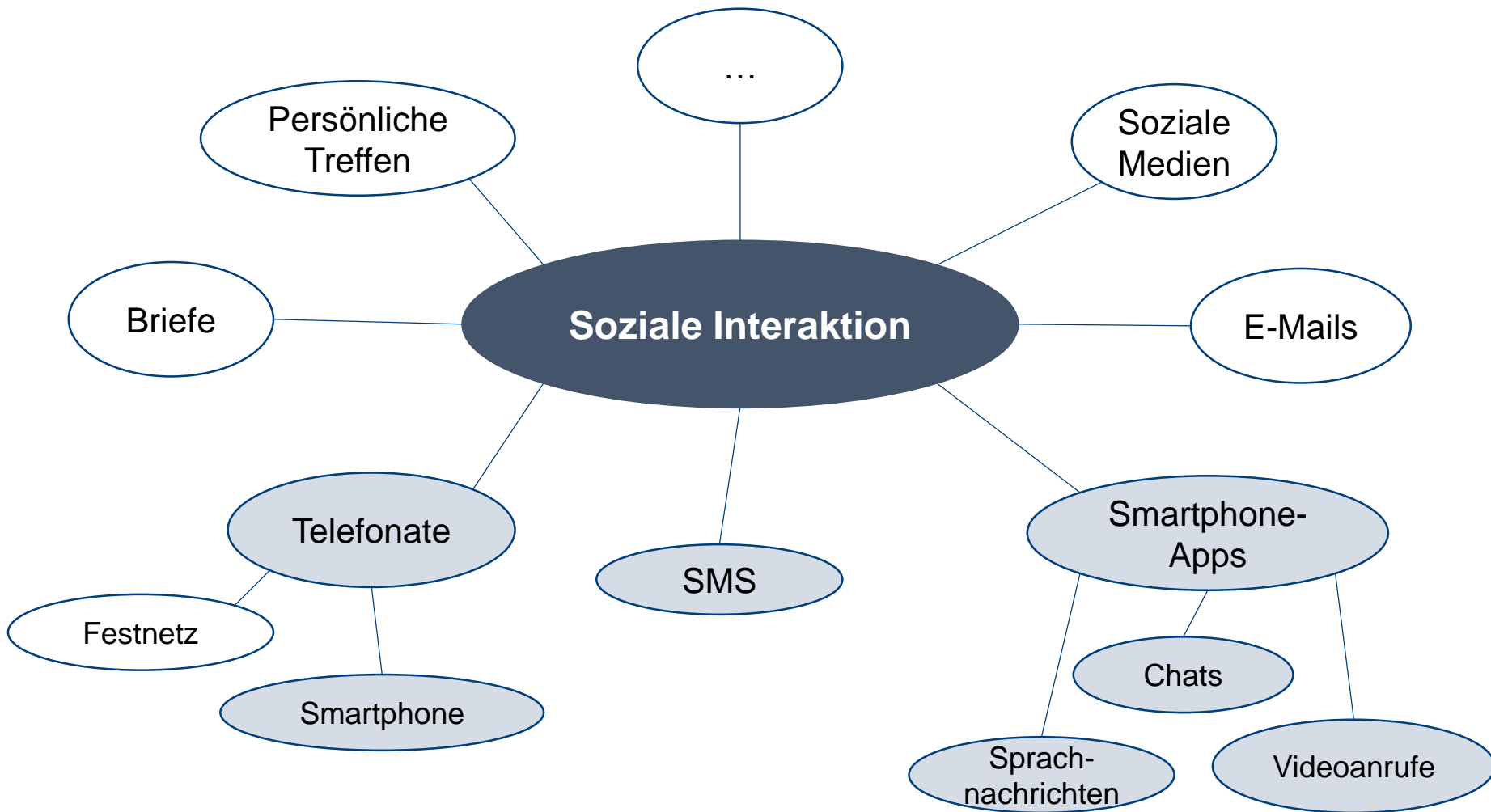
25.03.2022

AGENDA

1. Soziale Interaktion:
 - a) Formen
 - b) Messung von sozialer Interaktion
2. Daten: IAB-SMART
3. Methodische Herausforderungen im Umgang mit Smartphone-Daten
4. Lösungsvorschlag: Sequenz-Cluster-Analyse
5. Zusammenfassung und Fazit

1. SOZIALE INTERAKTION

FORMEN SOZIALER INTERAKTION



ERHEBUNGSMETHODEN SOZIALER INTERAKTION

- Beobachtung (Audio-, Videoaufnahmen, Labor)
 - Kleine Samples
 - Geringe externe Validität
- Retrospektive Selbstberichte
 - Erinnerungsfehler
- Selbstgeführte Tagebücher bzw. Protokolle
 - Aufwand
- **Digitale Verhaltensdaten (Soziale Medien, **Smartphone-Daten**, ...)**
 - Geringer Aufwand für Teilnehmende
 - Keine Verzerrung durch Erinnerungslücken
 - Natürlicher Rahmen
 - Große Informationsvielfalt

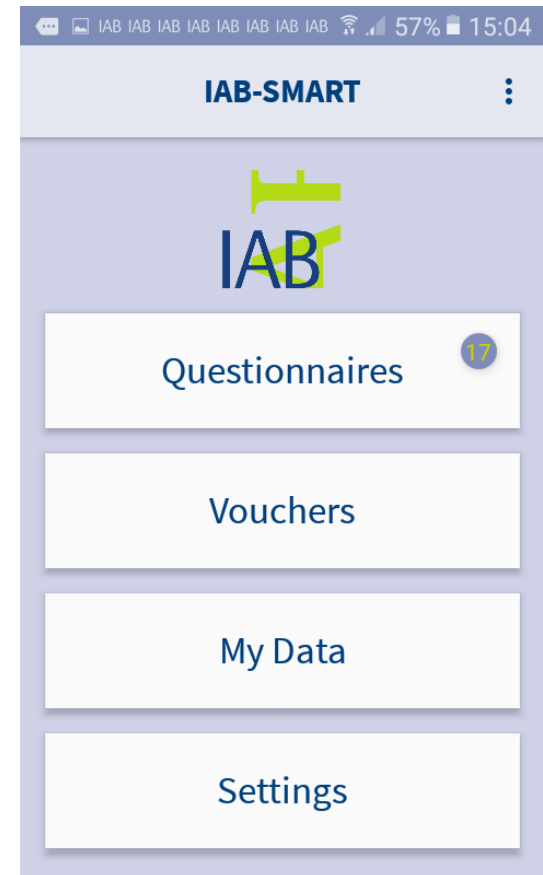
2. DATEN

IAB-SMART STUDIE (I) (KREUTER ET AL. 2020)

- **Panel Arbeitsmarkt und soziale Sicherung (PASS)** (Trappmann et al. 2019):
 - Zentrale Datenquelle für die Arbeitslosigkeits-, Armuts- & Sozialpolitikforschung
 - ~15.000 Personen in ~10.000 Haushalten jährlich seit 2007
- **IAB-SMART:** Zufallsstichprobe aus PASS-Teilnehmenden (18-64 J.)
 - Von 4.293 eingeladenen Android Smartphonebesitzer*innen haben 623 (14,5%) die IAB-SMART App heruntergeladen und Daten geliefert
 - Jan-Aug 2018

IAB-SMART STUDIE (II) (KREUTER ET AL. 2020)

- Passive Datenerfassung (Standortdaten, Smartphone-/Appnutzung, Schrittzähler, Interaktionshistorie, Kontaktinformation, ...)
 - Gemäß DS-GVO
 - Hoch frequente Messung
 - Vielzahl an Informationen/Indikatoren
- Kurze In-App-Befragungen
- Datenverknüpfung mit PASS and Registerdaten (Bundesagentur für Arbeit)
- Methodische Analysen zu Coverage, Nichtteilnahme, Messfehlern (z.B. Keusch et al. 2020, Keusch et al. 2022, Bähr et al. 2020)



3. METHODISCHE HERAUSFORDERUNGEN IM UMGANG MIT SMARTPHONE-DATEN

3. METHODISCHE HERAUSFORDERUNGEN IM UMGANG MIT SMARTPHONE-DATEN

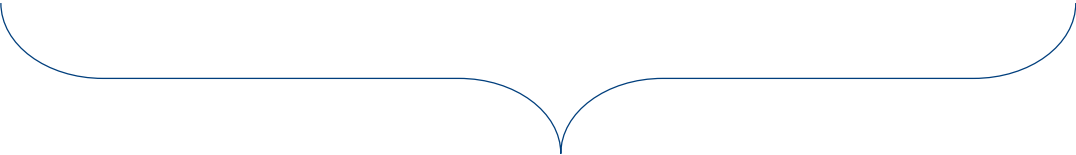
- 1) Umgang mit unbekanntem Werten
- 2) Vordergrund Modus
- 3) App-Kategorisierungen
- 4) Zählzeiten: Schiefe Verteilungen mit extremen Ausreißern

1) UMGANG MIT UNBEKANNTEN WERTEN

id	type	time_stamp_start	time_stamp_end	call_type
1	<i>call</i>	2018-01-09 16:01:23	2018-01-09 16:03:54	<i>incoming</i>
1	<i>call</i>	2018-01-09 18:10:40	2018-01-09 20:19:05	<i>outgoing</i>
1	<i>call</i>	2018-01-09 23:40:29	2018-01-09 23:40:29	<i>missed</i>
1	<i>call</i>	2018-01-11 17:00:45	2018-01-11 18:27:12	<i>incoming</i>

1) UMGANG MIT UNBEKANNTEN WERTEN

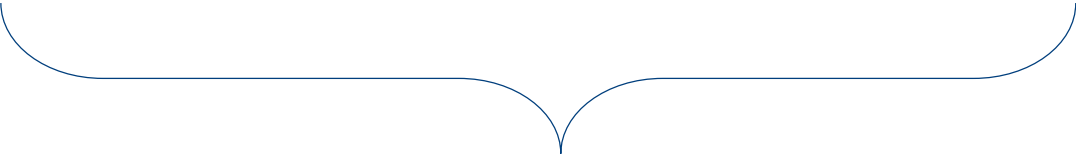
id	type	time_stamp_start	time_stamp_end	call_type
1	call	2018-01-09 16:01:23	2018-01-09 16:03:54	incoming
1	call	2018-01-09 18:10:40	2018-01-09 20:19:05	outgoing
1	call	2018-01-09 23:40:29	2018-01-09 23:40:29	missed
1	call	2018-01-11 17:00:45	2018-01-11 18:27:12	incoming



id	type	date	total_calls_per_day
1	call	2018-01-09	3
1	call	2018-01-11	1

1) UMGANG MIT UNBEKANNTEN WERTEN

id	type	time_stamp_start	time_stamp_end	call_type
1	call	2018-01-09 16:01:23	2018-01-09 16:03:54	incoming
1	call	2018-01-09 18:10:40	2018-01-09 20:19:05	outgoing
1	call	2018-01-09 23:40:29	2018-01-09 23:40:29	missed
1	call	2018-01-11 17:00:45	2018-01-11 18:27:12	incoming



id	type	date	total_calls_per_day
1	call	2018-01-09	3
1	call	2018-01-10	0
1	call	2018-01-11	1



fehlende Tage mit
Info befüllen

2) APPNUTZUNG IM VORDERGRUND

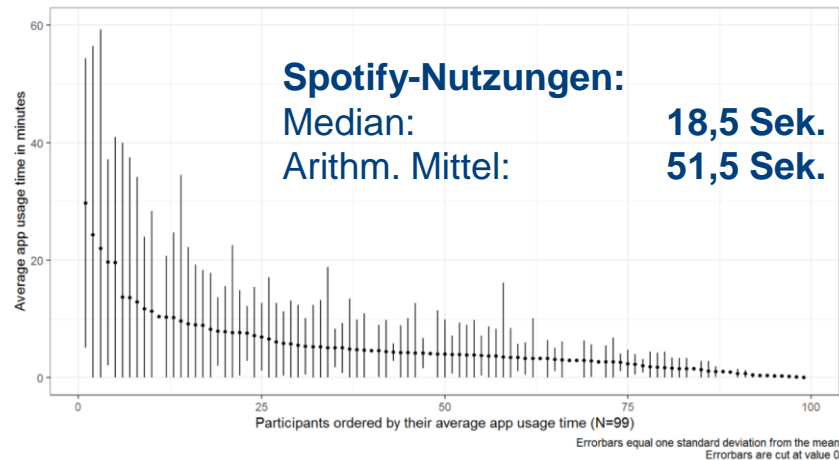
id	time_start	times_end	app
1	2018-03-01 16:01:23	2018-03-01 16:01:34	<i>com.whatsapp</i>
1	2018-03-01 18:10:40	2018-03-01 18:15:05	<i>com.google.android.calendar</i>
1	2018-03-01 20:49:29	2018-03-01 20:53:29	<i>com.instagram.android</i>
1	2018-03-01 21:00:24	2018-03-01 21:00:40	<i>com.spotify.music</i>

2) APPNUTZUNG IM VORDERGRUND

id	time_start	times_end	app
1	2018-03-01 16:01:23	2018-03-01 16:01:34	<i>com.whatsapp</i>
1	2018-03-01 18:10:40	2018-03-01 18:15:05	<i>com.google.android.calendar</i>
1	2018-03-01 20:49:29	2018-03-01 20:53:29	<i>com.instagram.android</i>
1	2018-03-01 21:00:24	2018-03-01 21:00:40	<i>com.spotify.music</i>

2) APPNUTZUNG IM VORDERGRUND

id	time_start	times_end	app
1	2018-03-01 16:01:23	2018-03-01 16:01:34	com.whatsapp
1	2018-03-01 18:10:40	2018-03-01 18:15:05	com.google.android.calendar
1	2018-03-01 20:49:29	2018-03-01 20:53:29	com.instagram.android
1	2018-03-01 21:00:24	2018-03-01 21:00:40	com.spotify.music



3) KATEGORISIERUNG VON APPS

IAB-SMART: ~ 8.000 verschiedene Apps

App	Google Play Store Category
Whatsapp	Communication
Microsoft Edge	Communication
Web.de	Communication
Outlook	Efficiency
Twitter	Social network
Instagram	Social network
Telegram	Social network

3) KATEGORISIERUNG VON APPS

IAB-SMART: ~ 8.000 verschiedene Apps

App	Google Play Store Category
Whatsapp	Communication
Microsoft Edge	Communication
Web.de	Communication
Outlook	Efficiency
Twitter	Social network
Instagram	Social network
Telegram	Social network

4) ZÄHL-/VERHALTENS DATEN

- Unsere Daten zählen Verhalten, d.h.:
 - Zähldaten
 - Nullenüberschuss
 - Linksschiefe Verteilung
 - Long tail (viele Extremwerte am rechten Ende der Verteilung)
 - Stark heterogenes Verhalten unterschiedlichster Individuen erfasst auf feingliedriger Ebene

3. METHODISCHE HERAUSFORDERUNGEN IM UMGANG MIT SMARTPHONE-DATEN

- 1) Umgang mit unbekanntem Werten
- 2) Vordergrund Modus
- 3) App-Kategorisierungen
- 4) Zählzeiten: Schiefe Verteilungen mit extremen Ausreißern

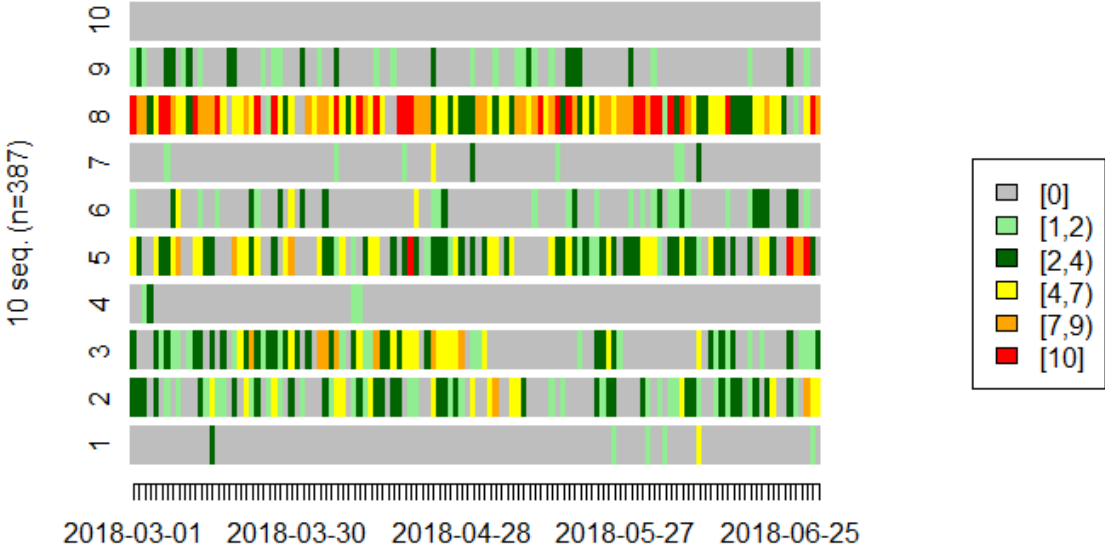
4. LÖSUNGSVORSCHLAG: SEQUENZ- CLUSTER-ANALYSE (SCA)

UMGANG MIT DEN HERAUSFORDERUNGEN ANHAND DER SCA

- 1) Umgang mit unbekanntem Werten
 - „0“ hat nicht dieselbe Bedeutung wie bei kontinuierlichen Variablen => Die Sequenzanalyse ordnet nicht: jeder „Status“ (jede Kategorie) ist gleichwertig. Damit haben Tage mit „0“ nicht mehr so einen großen Einfluss.
- 2) Vordergrund Modus
 - Können wir nicht lösen => Annahmen treffen.
- 3) App-Kategorisierungen
 - Können wir nicht lösen => händisch kategorisieren.
- 4) Zähldaten: Schiefe Verteilungen mit extremen Ausreißern
 - Schiefe Verteilung & long tail werden durch Kategorisierung reduziert ... (s.o.)

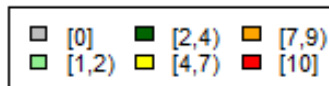
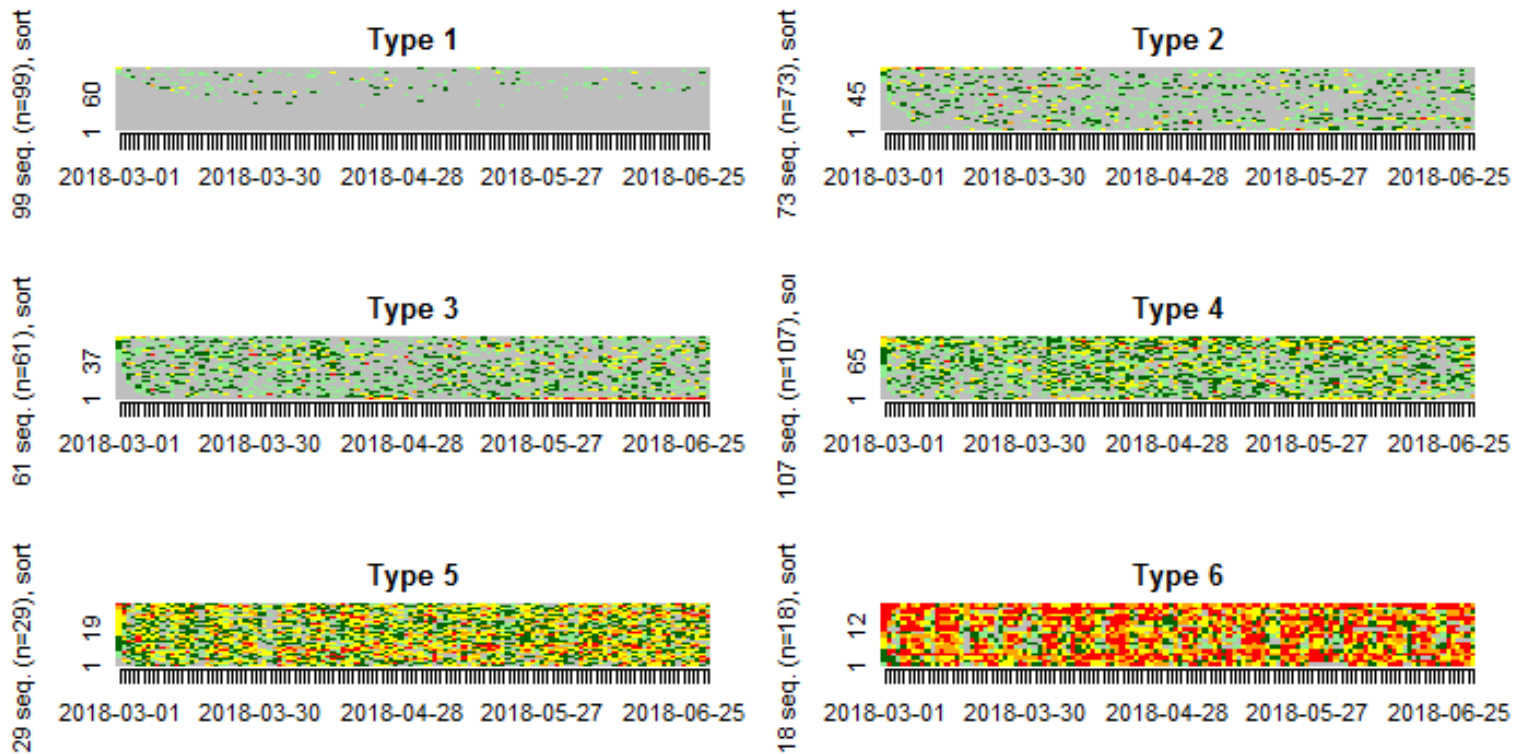
SEQUENZANALYSE

Die ersten 10 Personen (Sequenzen)



Anzahl eingehender & ausgehender Anrufe pro Tag pro Person (in Kategorien)

CLUSTERANALYSE DER SEQUENZEN



ZUSAMMENFASSUNG UND FAZIT

- Wir erhalten ein interpretierbares, aussagekräftiges Maß einer Dimension sozialer Interaktion
 - Reduziert Komplexität ohne den Zeitverlauf außer Acht zu lassen
 - Nutzt langen Beobachtungszeitraum
- Intuitive Herangehensweise

VIELEN DANK

Sonja Malich

Institut für Arbeitsmarkt- und Berufsforschung (IAB)

Forschungsbereich Panel Arbeitsmarkt und soziale Sicherung

sonja.malich@iab.de

REFERENCES

- Bähr, S., Haas, G.-C., Keusch, F., Kreuter, F., & Trappmann, M. (2020). Missing Data and Other Measurement Quality Issues in Mobile Geolocation Sensor Data. *Social Science Computer Review*. <https://doi.org/10.1177/0894439320944118>
- Chittaranjan, G., Blom, J., & Gatica-Perez, D. (2011). Who's Who with Big-Five: Analyzing and Classifying Personality Traits with Smartphones. 15th Annual International Symposium on Wearable Computers, 2011, 29-36. <https://doi.org/10.1109/ISWC.2011.29>
- de Montjoye, YA., Quoidbach, J., Robic, F., & Pentland A. (2013). Predicting Personality Using Novel Mobile Phone-Based Metrics. In A.M. Greenberg, W.G. Kennedy, N.D. Bos (eds), *Social Computing, Behavioral-Cultural Modeling and Prediction*. SBP 2013. Lecture Notes in Computer Science, 7812. Springer, Berlin, Heidelberg. https://doi.org/10.1007/978-3-642-37210-0_6
- Keusch, F., Bähr, S., Haas, G.-C., Kreuter, F., & Trappmann, M. (2020). Coverage Error in Data Collection Combining Mobile Surveys With Passive Measurement Using Apps: Data From a German National Survey. *Sociological Methods & Research*. <https://doi.org/10.1177/0049124120914924>
- Keusch, Florian, Sebastian Bähr, Georg-Christoph Haas, Frauke Kreuter, Mark Trappmann, and Stephanie Eckman. 2022. "Nonparticipation in Smartphone Data Collection Using Research Apps." *Journal of the Royal Statistical Society. Series A*.
- Kreuter, F., Haas, G.-C., Keusch, F., Bähr, S., & Trappmann, M. (2020). Collecting Survey and Smartphone Sensor Data With an App: Opportunities and Challenges Around Privacy and Informed Consent. *Social Science Computer Review*, 38(5), 533–549. <https://doi.org/10.1177/0894439318816389>
- Mønsted, B., Mollgaard, A., & Mathiesen, J. (2018). Phone-based metric as a predictor for basic personality traits. *Journal of Research in Personality*, 74, 16–22. <https://doi.org/10.1016/j.jrp.2017.12.004>
- Montag, C., Błaszczewicz, K., Lachmann, B., Andone, I., Sariyska, R., Trendafilov, B., Reuter, M., & Markowetz, A. (2014). Correlating personality and actual phone usage: Evidence from psychoinformatics. *Journal of Individual Differences*, 35(3), 158–165. <https://doi.org/10.1027/1614-0001/a000139>
- Stachl, C., Au, Q., Schoedel, R., Gosling, S.D., Harari, G.M., Buschek, D. et al. (2020): Predicting personality from patterns of behavior collected with smartphones. In *Proceedings of the National Academy of Sciences of the United States of America*, 117 (30), 17680–17687. <https://doi.org/10.1073/pnas.1920484117>
- Trappmann, M., Bähr, S., Beste, J., Eberl, A., Frodermann, C., Gundert, S., Schwarz, S., Teichler, N., Unger, S., & Wenzig, C. (2019). Data Resource Profile: Panel Study Labour Market and Social Security (PASS). *International Journal of Epidemiology*, 48(5), 1411–1411g. <https://doi.org/10.1093/ije/dyz041>