

MAPPING A DARK SPACE: CHALLENGES IN SAMPLING AND CLASSIFYING NON- INSTITUTIONALIZED ACTORS ON TELEGRAM

Social Media Observatory (SMO) Starter Workshop Telegram

23. Januar 2024 | Leibniz-Institut für Medienforschung | Hans-Bredow-Institut

Pablo Jost (Johannes Gutenberg-Universität Mainz)

Annett Heft (Weizenbaum Institut für die vernetzte Gesellschaft und Freie Universität Berlin)

Kilian Bühling (Weizenbaum Institut für die vernetzte Gesellschaft und Freie Universität Berlin)

Maximilian Zehring (Technische Universität Ilmenau)

Heidi Schulze (Ludwig-Maximilians-Universität München)

Hendrik Bitzmann (BAG «Gegen Hass im Netz»)

Emese Domahidi (Technische Universität Ilmenau)

Potenziale & Herausforderungen bei der Forschung auf Telegram

- Telegram als aufstrebende Plattform mit Fokus auf offener (nicht zensierter) Kommunikation (Telegram, 2022)
- Bedeutung als sicherer Raum für Aktivismus (in repressiven politischen Settings) (Herasimenka, 2022; Urman & Katz, 2022)
- Telegram als **wertvolle Datenquelle** für die Erforschung nicht institutionalisierter Aktivitäten
- *Herausforderungen:* **Pluralität sozialer Bewegungen** und **Vielfalt der Kommunikationsfunktionen** auf Telegram (Diani, 1992; Urmann & Katz, 2022b)

Fokus des Papiers: Methodologische Herausforderungen bei der Stichprobenziehung und Akteursklassifizierung



Agenda

- **Identifikation von Herausforderungen** für nicht-institutionalisierte Akteure auf Basis der Forschung zu sozialen Bewegungen
- Diskussion ausgewählter Ansätze zur **Definition und Ziehung von Stichproben** (in mehrstufigen Verfahren)
- Strategien zur **Klassifizierung von Akteuren**
- Ausblick: Herausforderungen und potenzielle Strategien für **zukünftige Forschung**

Struktur sozialer Bewegungen

- Sozialer Bewegungen als informelle Netzwerke mit gemeinsamen kollektiven Identitäten
- **kaum formale Merkmale** zur Definition von Mitgliedschaft
- Komplexität bei der Identifikation relevanter Akteure: Abhängig von Organisationsstruktur, **Hierarchie und Formalisierung der Bewegung** (Kriesi, 1996; Willems & Jegers, 2012)
 - Formell und hierarchisch organisierte Bewegungen haben stärkere Verbindungen und klare Ziele
 - Weniger formelle Bewegungen ohne Hierarchien neigen zu mehr Ideologenvielfalt und abstrakten Zielen
 - Instabilität und wechselnde Koalitionen in weniger formalen Bewegungen (Willems & Jegers, 2012)
 - Interne Struktur der Akteure beeinflusst durch Ressourcenfluss: Formalisierung, Professionalisierung, interne Differenzierung und Integration (Kriesi, 1996)
- Bedeutung von geografischem und zeitlichem Rahmen für die Analyse von Bewegungen und deren Einfluss auf die Identifikation relevanter Akteure

ERFASSUNG DER POPULATION / STICHPROBENBILDUNG

Ansätze zur Stichprobenbildung

- Zentrale Entscheidungen zur Definition, Klassifizierung und Auswahl davon ab, inwieweit diese Population bekannt und a priori spezifizierbar ist oder ob die Population a priori nicht definiert werden kann (eine "unbekannte" Population).
- zuvor **definierte / bekannte** Population (a priori)
- (mehrstufige) **inhaltsbasierte Strategien** (unbekannte Population)
- (mehrstufiger) **akteurbasiertes Netzwerksampling** (unbekannte Population)

Definierte / bekannte Population (a priori)

- Definition, Auswahl und Klassifikation von Akteuren als Ein-Schritt-Prozess.
- Ziele: Vollständige Sammlung von wenigen Fällen oder Auswahl typischer, extremer, prominenter Akteure.
- Häufig bei Studien mit bewusster Auswahl einzelner Akteure (zB. ISIS) und Expertenwissen.
- Zwei Strategien für Auswahl/Klassifikation:
 1. A-priori-Festlegung bekannter Kanäle als (a) vollständige Sammlung oder (b) feste Stichprobe.
 2. Verwendung vorhandener Akteurslisten (z.B. von Strafverfolgungsbhörden).

Einschränkungen:

- Ergebnisse nicht verallgemeinerbar; Geltung nur für die untersuchten Kanäle und Gruppen
- Expertenentscheidungen erfordern theoretische Grundlagen und Rechtfertigung, da sie kaum empirisch validiert werden können.

Inhaltsbasierte Strategien zur Identifikation unbekannter Populationen

- Multistep-Ansatz: Nachrichten oder Kanalbeschreibungen nach Stichworten durchsuchen
- Inhaltsbasierte Strategie zur Datensammlung, entweder
 - (a) themenspezifische Schlüsselwörter für bestimmte Akteure (z.B. “Pepe the frog”, “QAnon”, “Querdenken”) oder
 - (b) breite Schlüsselwortbasierte Verfahren für allgemeine Population (z.B. Liste aktueller politischer Themen).
- Inhaltsbasierte Stichprobenziehung für weniger bekannte Akteursgruppen geeignet.

Einschränkungen:

- Voraussetzung: Konsistenz der Begrifflichkeiten über Akteure hinweg (nicht gegeben bei ideologisch heterogenen Gruppen).
- Probleme mit Telegram-spezifischem Slang (spezifisches Fachwissen möglich).
- Bei allgemeiner Population bestimmt Auswahl der Schlüsselwörter Umfang und Verallgemeinerbarkeit der Ergebnisse.

Akteursbasierte Strategien zur Identifikation unbekannter Populationen

- Telegram als soziales Netzwerk: Kanäle und Chats als Knoten
- Netzwerkstruktur nicht extern beobachtbar, daher Einsatz von Sampling-Techniken für unbekannte Graphen zur Identifizierung von Akteuren.
- Beispiel: **Snowball-Sampling** (Goodman, 1961) beginnt mit **vorher ausgewählten Seed-Knoten** (z.B. Telegram-Kanäle), sammelt deren Nachrichten und identifiziert Bezüge zu anderen Kanälen oder Gruppen.
- Gefundene Kanäle dienen als neue Seed-Stichprobe, Prozess iteriert bis Erreichen eines **Abbruchkriteriums** (z.B. Semenzin & Bainotti, 2020; Gursky et al., 2022).

Einschränkungen

- Identifikation der Population abhängig von Designentscheidungen, besonders bei **Knotenkriterien**, **Saatauswahl** und **Stop-Kriterien**.

Exkurs: Spezifische Entscheidungen im Schneeball-Sampling und ihre Auswirkungen

- Kantenbasierten vs knotenbasierten vs netzwerkbasierten Kriterien.
 1. **Kantenbasierte Kriterien:** Beeinflussen Netzwerkstruktur durch Berücksichtigung verschiedener Verweisarten.
 2. **Knotenbasierte Kriterien:** Filtern relevante Kanäle basierend auf Eigenschaften, begrenzen Wachstum der Stichprobe.
 3. **Netzwerkbasierter Kriterien:** Priorisieren Knoten basierend auf ihrer Position im entdeckten Netzwerk.
- Auswahl der Seedkanäle beeinflusst den Gesamtsamplingprozess stark, Vielfalt in der Saatliste mindert Verzerrungen.
- Stop-Bedingung: Meist nicht definiert
 - Forscher stoppen bei „angemessenem“ Umfang von Kanälen/Daten im Studiendesign.
 - Begrenzung durch Iterationen mit Expert*innen oder Sprache der Kanäle

AKTEURSKLASSISFIKATION

Klassifikation von Akteuren (in mehrstufigen Samplings)

- **Manuelle Akteursklassifikationen** umfassen Inhaltsanalysen, bei denen (a) Akteurbeschreibungen, (b) Beiträge oder (c) eine Kombination für die Klassifikation verwendet werden.
- **Computergestützten Akteursklassifikationen** auf Basis von Inhalten (bspw. wörterbuchbasierte Ansätze wie Hatebase-Lexikon).
- **Netzwerkbasierende Ansätze der Akteursklassifikation**
 - Community Detection-Algorithmen.
 - Gültigkeit von Label-Propagation-Ansätzen hängt von den Merkmalen ab. Studien verwenden Schwellenwerte, um falsche Klassifikationen zu reduzieren.

ZUSAMMENFASSUNG

TL;DR - Implikationen für zukünftige Forschung

1. A-priori Akteursidentifikation (bspw. durch Expert*innen) für strukturierte, Snowball für weniger organisierte Bewegungen.
2. Große, **diverse Seed-Samples** für bessere Ergebnisse bei größeren, weniger strukturierten Bewegungen.
3. **Trade-offs bei Snowball-Sampling:** Unbeschränkt reduziert Verzerrungen, erfordert aber mehr Ressourcen. Strikte Kriterien führen zu Verzerrungen (Risiko des Ausschlusses relevanter Kanäle).
4. **Stopp-Kriterien:** Niedrige Priorität in der Forschung. Netzwerksättigung oder Referenzdatenvergleiche sind mögliche Ansätze (z.B. Münch et al. 2021).
5. Content-Analyse für Akteursklassifikation: Schwierigkeiten bei heterogenen Bewegungen, extremistischen Gruppen und Telegrams Struktur (eingeschränkt empfehlenswert).
6. **Ethik in der Forschung:** Datenschutz und Forscherversicherheit beachten, besonders in repressiven Umgebungen.

VIELEN DANK!
FRAGEN?

Jost, P., Heft, A., Buehling, K., Zehring, M., Schulze, H., Bitzmann, H., & Domahidi, E. (2023). Mapping a Dark Space: Challenges in Sampling and Classifying Non-Institutionalized Actors on Telegram. *Medien & Kommunikationswissenschaft*, 71(3–4), 212–229.
<https://doi.org/10.5771/1615-634X-2023-3-4-212>