

Analyse sozialer Netzwerke mit digitalen Verhaltensdaten

Lietz, Haiko; Schmitz, Andreas; Schaible, Johann

Veröffentlichungsversion / Published Version

Zeitschriftenartikel / journal article

Zur Verfügung gestellt in Kooperation mit / provided in cooperation with:

GESIS - Leibniz-Institut für Sozialwissenschaften

Empfohlene Zitierung / Suggested Citation:

Lietz, H., Schmitz, A., & Schaible, J. (2021). Analyse sozialer Netzwerke mit digitalen Verhaltensdaten. *easy_social_sciences*, 66, 90-98. <https://doi.org/10.15464/easy.2021.010>

Nutzungsbedingungen:

Dieser Text wird unter einer CC BY Lizenz (Namensnennung) zur Verfügung gestellt. Nähere Auskünfte zu den CC-Lizenzen finden Sie hier: <https://creativecommons.org/licenses/by/4.0/deed.de>

Terms of use:

This document is made available under a CC BY Licence (Attribution). For more information see: <https://creativecommons.org/licenses/by/4.0>



Analyse sozialer Netzwerke mit digitalen Verhaltensdaten

Haiko Lietz, Andreas Schmitz & Johann Schaible

Die Nutzung von digitalen Technologien, wie Social-Media-Plattformen, hinterlässt riesige Mengen an Verhaltensspuren, die für die Sozialforschung überaus interessant sind. Andere digitale Technologien, wie Mobiltelefone, ermöglichen die Erhebung von Verhaltensspuren zu Forschungszwecken. Diese digitalen Verhaltensdaten bestehen aus genuin relationalen Beziehungen, die als Netzwerke betrachtet werden können. Diese Art von Daten erfordert eine Verlagerung der Perspektive von Individuen zu Mikroereignissen (z.B. ein Post in den sozialen Medien) als Beobachtungseinheiten und stellt etablierte Techniken, wie die Analyse sozialer Netzwerke, in den Mittelpunkt. Wir argumentieren, dass die Verwendung dieses Ansatzes, die Ermittlung individueller Eigenschaften und Einstellungen sowie die Offenlegung von Mikro-Makro-Verhaltensdynamiken mittels Auswertung von Mustern potenziell nützliche Anwendungen sind. Wir diskutieren methodische Herausforderungen und gelangen zu der Schlussfolgerung, dass soziale Theorie ein Grundpfeiler für die Konsolidierung der Computational Social Science (computergestützte Sozialwissenschaft) ist.

Keywords: Digitale Verhaltensdaten, Analyse sozialer Netzwerke, Computational Social Science, Transaktionen, Eigenschaften und Einstellungen, Muster

Durch digitale Medien wie Facebook sind wir heute alle mit dem Konzept der sozialen Netzwerke vertraut. Online-Praktiken sind als „social networking“ Teil unseres Alltags geworden. Aber auch außerhalb von digitalen Plattformen stoßen wir immer wieder auf das Netzwerk-Konzept: Stellensuche, sportliche Aktivitäten, die Frage, mit wem wir zusammenarbeiten möchten (oder eher nicht) – immer mehr Dinge sind faktisch als Netzwerke organisiert. Tatsächlich interessieren sich die Sozialwissenschaften seit ihren Anfängen im 19. Jahrhundert für soziale Beziehungen und haben sich infolgedessen zunehmend explizit mit Netzwerken befasst. Die Idee besteht nicht nur darin, dass soziale Formationen als durch Kanten miteinander verbundene Knoten und die entstehenden Strukturen beschrieben werden können, sondern dass sie tatsächlich als Netzwerke funktionieren (White, 2008).

Die Bedeutung dieser theoretischen Perspektive und der zugehörigen Forschungsmethoden ist in letzter Zeit erheblich gewachsen. Wir leben einen stetig wachsenden Teil unserer Zeit in der digitalen Welt. In diesem Leben in digitalen Ökosystemen hinterlässt jede Handlung Spuren. Facebook erfasst, wer mit wem befreundet ist, Google, wer was gesucht hat, und Amazon, wer was gekauft hat. Alle diese Daten werden in Echtzeit erhoben. Aber das ist noch nicht alles: Facebook weiß, welche Nutzer potenziell miteinander befreundet sind, Google weiß, welche Suchbegriffe häufig miteinander kombiniert werden, und Amazon weiß, welche Produkte häufig zusammen gekauft werden. Plattformbetreiber werten die *Muster* aus, die sich in der Gesamtheit der zusammen ausgewählten Dinge herausbilden. Die Kenntnis von Mustern ist das neue Gold in der Digitalwirtschaft. Die Plattform-

betreiber verkaufen sie oder nutzen sie, um neue Freunde, Suchbegriffe oder Produkte zu empfehlen. Die Netzwerkperspektive hat an Bedeutung gewonnen, da Muster die Form von Netzwerken annehmen und die zugehörigen Methoden für die Datenauswertungsprozesse hilfreich sind (Nassehi, 2019). Für die Sozialwissenschaften ist es besonders interessant, wie diese Verhaltensdaten, die häufig als Nebenprodukt der Geschäftstätigkeit privater Unternehmen anfallen, so genutzt werden können, dass sie systematisch mit theoretischen Konzepten in Zusammenhang stehen.

» Ein „Teleskop“ der Sozialwissenschaften: die gleichzeitige Verfügbarkeit riesiger Mengen von Verhaltensdaten und der technologischen Fähigkeiten zu ihrer Analyse. «

Watts (2013, S. 239) vergleicht die Sozialwissenschaften zu Beginn des 21. Jahrhunderts mit den Anfängen der modernen Astronomie zu Beginn des 17. Jahrhunderts und proklamiert, dass „[wir] endlich unser Teleskop entdeckt [haben].“ Dieses „Teleskop“ der Sozialwissenschaften wird in der gleichzeitigen Verfügbarkeit riesiger Mengen von Verhaltensdaten und der technologischen Fähigkeiten zu ihrer Analyse gesehen. Ein Teil dieser technologischen Innovation besteht darin, dass das „Teleskop“ massenhaft mikroskopische Ereignisse oder Transaktionen (z.B. Posts in sozialen Medien) für ein komplettes digitales Ökosystem verarbeiten kann und das Potenzial zur Schaffung einer völlig neuen Form von Sozialwissenschaft hat. Die Computational Social Science (CSS), ein aufstrebendes Feld an der Schnittstelle zwischen Sozialwissenschaft und Informatik, stellt sich den damit verbundenen Herausforderungen (Lazer et al., 2020).

In diesem Artikel schlagen wir eine Definition für digitale Verhaltensdaten vor und erörtern ihre Eigenschaften sowie beispielhafte Anwendungen. Wir stellen die These auf, dass sie erhebliches Potenzial bei der Ermittlung von individuellen Eigenschaften und Einstellungen sowie als Makroskop zur Offenlegung von Mikro-Makro-Verhaltensdynamiken aufweisen. Zum Abschluss folgt eine Diskussion fünf wichtiger Einschränkungen und Herausforderungen in Zusammenhang mit der Nutzung des „neuen Teleskops“ der Sozialwissenschaft.

Digitale Verhaltensdaten im Kontext

Wir definieren digitale Verhaltensdaten (DVD) als durch Nutzung digitaler Technologie hinterlassene oder durch digitale Technologie erhobene Verhaltensspuren. Um unser Verständnis von Verhalten zu entwickeln, charakterisieren wir diese Spuren als Aufzeichnungen von *Transaktionen*, also beobachtete Phänomene, die nicht als atomistische Entitäten, sondern vielmehr als genuin relationale Emanationen definiert sind. Diese Objekte haben vier Dimensionen (Abbildung 1): Erstens umfassen sie soziale Beziehungen zwischen Akteur*innen, üblicherweise zwischen Sender*in und Empfänger*in oder mehreren Empfänger*innen. Zweitens enthalten sie Kommunikationsinhalte und -bezüge.

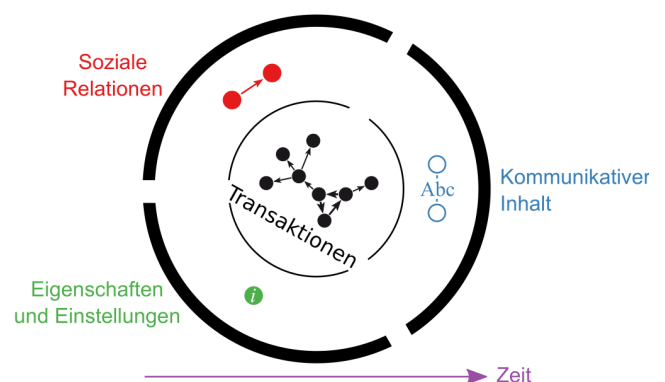


Abbildung 1 Vier Dimensionen von Transaktionen als Mikroereignisse

Drittens haben Entitäten Attribute (wie die Einstellungen der Akteur*innen), die ebenfalls relationalen Ursprungs sein können. Viertens beinhalten Transaktionen eine zeitliche Dimension, d.h. sie werden als Mikroereignisse zu einem bestimmten Zeitpunkt initiiert (Emirbayer, 1997). Auf Twitter werden Transaktionen beispielsweise „Tweets“ genannt. Sie stellen soziale Beziehungen dar, insofern als sie Aktionen in einem Netzwerk von Nutzenden sind und einen Bedeutungsinhalt (bis 280 Zeichen) haben. Darüber hinaus können Nutzer-Attribute den Profilen entnommen werden, und Tweets sind mit einem Zeitstempel versehen.

In unserer Definition von DVD unterscheiden wir zwischen zwei Arten von Spuren, die in der Praxis auftreten: organische und entworfene DVD. Durch die Nutzung von digitaler Technologie hinterlassene Spuren entsprechen *organischen* DVD (Howison et al., 2011). Dies subsumiert Daten wie Logs von digitalen Plattformen, wo die Spur gleichzeitig die Aktion und ihre technisch durchgeführte Beobachtung ist. Das idealtypische Beispiel ist die Nutzung von sozialen Medien und Kurznachrichtendiensten, wie Facebook, WhatsApp oder YouTube. Praktiken wie posten, kommentieren, taggen oder liken bilden „digitales Leben“ ab (Lazer & Radford, 2017). Onlinekäufe, Surfen im Internet und Suchen im Web sind weitere Beispiele. Diese spezifischen Spuren bestehen mindestens aus Daten zu sozialen Beziehungen und Kommunikationsinhalten. E-Mails, Telefonanrufe und SMS-Kurznachrichten hinterlassen ebenfalls Spuren durch die Nutzung der zugrunde liegenden technischen Infrastruktur, aber sie sind keine Formen von digitalem Leben, da die Spuren nur Aufzeichnungen der Aktionen sind aber nicht die Aktionen an sich (Lazer & Radford, 2017). Bei E-Mails werden soziale Beziehungen und Kommunikationsinhalte auf E-Mail-Servern protokolliert. Einzelverbindungenachweise, die heutzutage von Telefonanbietern verkauft werden, enthalten hingegen keine Kommunikationsinhalte.

Durch digitale Technologie erhobene Ver-

haltensspuren entsprechen *entworfenen* DVD, das heißt, die Daten sind kein Nebenprodukt des Betriebs einer digitalen Plattform, sondern werden explizit für einen Forschungszweck generiert. Ein Beispiel sind Interaktionen von Angesicht zu Angesicht. Diese Spuren können von tragbaren Sensoren aufgezeichnet werden, die messen, ob zwei Träger*innen eines Sensors nahe beieinander stehen und sich zugewandt sind oder nicht. Diese Formen von Mikroverhalten sind reine Aufzeichnungen sozialer Beziehungen mit Zeitstempel (Schaible et al., 2022). Andere digitale Spuren verfügen nur über die Transaktionseigenschaften von Attributen mit Zeitstempel, sogenannte Metadaten zu Mikroereignissen. Beispiele sind der durchgehend erfasste GPS-Standort von Mobiltelefonnutzenden (physisches Attribut), von Fitnesstrackern aufgezeichnete Schlafphasen und Herzfrequenzen (biologische Attribute), von einem Bluetooth-Sensor erfasste andere Mobiltelefone in der Nähe (soziales Attribut) oder der Geräuschpegel am Aufenthaltsort von Mobiltelefonnutzenden (ein Umgebungsattribut). Es ist Aufgabe des Forschers, diese Spuren Forschungskonzepten zuzuordnen (z.B. Sensornähe zu Zusammenarbeit).

Eigenschaften digitaler Verhaltensdaten

DVD werden häufig als „Big Data“ bezeichnet, definiert durch die drei V-Eigenschaften: großes „Volume“ (übersteigen die Kapazität konventioneller Hardware), große „Variety“ (liegen in anderen Formen als rechteckigen Tabellen vor) und/oder große „Velocity“ (sind longitudinal und variabel). Tatsächlich sind DVD recht vielfältig, und jede Datenquelle erfordert eine eigene Verarbeitungsroutine. In der Praxis können sie jedoch zum Teil auf einem Laptop- oder Desktop-Rechner verarbeitet werden. Ein grundlegenderer Aspekt von DVD ist ihre relationale Logik. Abbildung 1 zeigt diesbezüglich die soziale Beziehung zwischen Sender*in und Empfänger*in. Aber

Tabelle 1 Sozialdaten im Vergleich

	Befragungsdaten	Digitale Verhaltensdaten
Beobachtungseinheit	Individuum	Transaktion
Struktur	Querschnitt	Longitudinal
Umfang	Repräsentativ	Umfassend
Verzerrungsquelle	Wahrnehmung	Rückkopplung

es können beliebige Entitäten sein, die miteinander in Beziehung stehen oder auf verschiedene Weise miteinander in Beziehung gebracht werden können (z.B. Transaktionen, die auf frühere Transaktionen verweisen, zusammen verwendete Wörter oder gemeinsam auftretende Attribute).

Die Bedeutung von DVD wird durch einen Vergleich mit Befragungsdaten deutlicher (Tabelle 1). In der Tradition der Befragungsmethode, die vor einem Jahrhundert aufkam, ist die Analyseeinheit, über die Informationen gewonnen werden sollen, die Gesellschaft als Ganzes oder eine ihrer Teilpopulationen. Jedoch wurde aus methodologischen Gründen das Individuum zur paradigmatischen Beobachtungseinheit. Im Rahmen einer Befragung gibt ein Individuum seine Eigenschaften und Einstellungen zu einem bestimmten Thema selbst an. Da Vollerhebungen kaum durchgeführt werden können, repräsentative Stichproben jedoch geeignete Ergebnisse liefern, wurde die Befragungsmethodik im engen Zusammenhang mit Verfahren der Stichprobenerhebung entwickelt. Methodische Einschränkungen sind noch gravierender, wenn Forschenden longitudinale Beobachtungen anstreben.

DVD sind hingegen umfassend. Grundsätzlich kann Jemand mit den erforderlichen Zugangsrechten und Fähigkeiten alle Transaktionen auf einer digitalen Plattform analysieren. Dies wäre eine Vollerhebung, wobei die Beobachtungseinheit nicht Individuen, sondern Transaktionen sind. DVD sind zumeist stark schiefenverzerrt, da wenige Nutzer*innen eine unverhältnismäßige Zahl an Beobachtungen beitragen. Aber dies ist eine

Eigenschaft des Systems, das beobachtet wird, und nicht ein Problem. Als konstanter Datenstrom (das Velocity-Argument von Big Data) setzen DVD keine Grenzen für die Analyse dieser kollektiven Dynamiken (Diaz et al., 2016).

» **Grundsätzlich kann Jemand mit den erforderlichen Zugangsrechten und Fähigkeiten alle Transaktionen auf einer digitalen Plattform analysieren.** «

Die Tatsache, dass Transaktionen die Beobachtungseinheit sind, bietet vielfältige Möglichkeiten für die Analyseeinheit. Informationen aus den Transaktionen (z.B. Akteur*innen, Wörter, sogar das Ereignis an sich) können als Netzwerkknoten oder -kanten oder beides verwendet werden. Die Analyseeinheiten hängen dann von der Analyseebene ab und können alles sein, von den individuellen Knoten und, im Falle von Akteur*innen, ihrem Mikroverhalten (z.B. ihre Positionsdynamiken) bis hin zum gesamten Netzwerk und seinem Makroverhalten (White, 2008). In der Einleitung haben wir Makroverhalten als Muster bezeichnet. Auch Muster können vielfältige Formen annehmen, von der Existenz von Akteursgruppen bis hin zu einem strukturierten Diskurs.

Eine wichtige Eigenschaft von DVD betrifft Rückkopplungen (Feedback), einen Mechanismus, bei dem das individuelle Verhalten vom kollektiven Verhalten oder individuellen

Verhalten in der Vergangenheit beeinflusst wird. Rückkopplungen sind entscheidend für die Entstehung von sozialen Strukturen (Keuschnigg et al., 2018). Eine besondere Eigenschaft von organischen DVD (digitales Leben) besteht darin, dass sie auch Rückkopplungsschleifen ausgesetzt sind, die durch die Plattformen in Kraft gesetzt werden. Dies geschieht in Form von Mustern, die aus Mikroverhalten hervorgehen, dieses später aber auch beeinflussen (z.B. durch Empfehlungen oder gefilterte Ereignisströme). Rückkopplung ist eine wesentliche Verzerrungsquelle von organischen DVD und steht im Gegensatz zur Wahrnehmung als Hauptverzerrungsquelle bei Befragungsdaten (z.B. Erwünschtheitsverzerrung, Erinnerungsverlust).

Die Möglichkeit zur Untersuchung realer Situationen, zur Rekonstruktion von manifester und symbolischen Beziehungen, zur Offenlegung von Verhaltenslogiken auf Mikro- und Makroebenen mit hoher zeitlicher Auflösung und in skalierender Größenordnung sind die Hauptgründe dafür, dass DVD das aufstrebende Feld der CSS vorantreiben. Im Folgenden werden zwei Anwendungsszenarien basierend auf relationalen Analysen erörtert. Für andere Ansätze, wie maschinelles Lernen, soziale Simulationen oder Experimente, verweisen wir auf andere Quellen (Lazer & Radford, 2017; Lazer et al., 2020).

Anwendungen digitaler Verhaltensdaten

Szenario 1: Ermittlung individueller Eigenschaften und Einstellungen

Ziel der Befragungsmethode ist die Ermittlung der Eigenschaften und Einstellungen von Individuen. Viele davon lassen sich mit einem bezifferbaren Fehler auch aus DVD ableiten. Kosinski et al. (2013) haben durch Befragung und psychometrische Tests von beinahe 60.000 Facebook-Nutzenden detaillierte psycho-

demografische Profile erhoben und versucht, die resultierenden Variablen mithilfe von sozialwissenschaftlichen Standardmethoden nur anhand ihres Liking-Verhaltens auf Facebook vorherzusagen. Attribute wie Alter, Geschlecht, politische und religiöse Ansichten sowie sexuelle Orientierung ließen sich alle mit einer Genauigkeit von mindestens 75 % vorhersagen. Sozialpsychologische Merkmale ließen sich wesentlich ungenauer vorhersagen.

Zahlreiche Teilbereiche der CSS befassen sich mit der Entwicklung von Methoden zur Erkennung von Meinungen, Begründungen, Haltungen, Gerüchten, Emotionen, Sarkasmus oder Ironie aus dem Textteil von DVD. Derartige Inferenzen leiten sich aus Methoden der Verarbeitung natürlicher Sprache, d.h. automatisierten Verfahren zur Beschaffung, Verarbeitung und Analyse von Kommunikation, ab. Diese Ansätze haben unlängst besondere Bekanntheit erlangt, da DVD die Erzeugung, Verarbeitung und Verwendung von Bedeutung empirisch zugänglich machen (Bail, 2014).

Während Eigenschaften und Einstellungen aus organischen DVD abgeleitet werden müssen, können Attribute von den Forschenden mithilfe digitaler Technologie auch direkt erhoben werden. Die Copenhagen Networks Study ist ein Paradebeispiel für die Schlagkraft einer Mischung aus organischen DVD, entworfenen DVD und Befragungsdaten sowie methodenintegrativen Verfahren. 2013 verteilten Forschenden 1.000 Mobiltelefone an Studierende der Technischen Universität Dänemarks und erfassten deren physischen Standort (über den GPS-Sensor des Telefons), wer sich in wessen Nähe aufhielt (über den Bluetooth-Sensor), wer wen anrief und wer wem eine SMS-Kurznachricht schickte. Außerdem wurden ihre Transaktionen auf Facebook gesammelt sowie ihre demografischen und psychologischen Merkmale erfragt. In Abbildung 2 sind vier Arten von sozialen Beziehungen aus dem öffentlich verfügbaren Teil dieses Datensatzes dargestellt.

Die Forschenden konnten allein durch Untersuchung der relationalen Schicht der

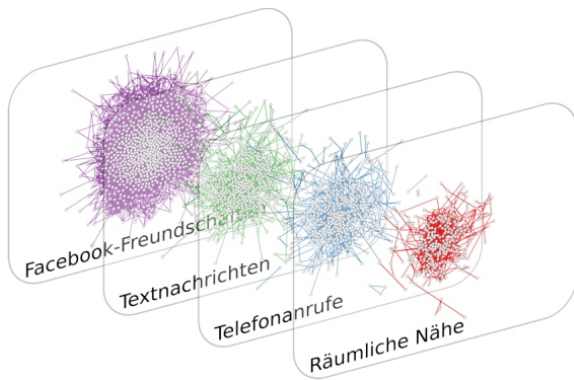


Abbildung 2 Soziale Beziehungen in der Copenhagen Networks Study

physischen Nähe die Mikroverhaltensursprünge für Gruppenbildung offenlegen (Sekara et al., 2016). Im Rahmen eines weiteren Projekts haben sie versucht, die „Big Five“-Persönlichkeitsmerkmale (Offenheit für Erfahrungen, Gewissenhaftigkeit, Extraversion, Verträglichkeit und Neurotizismus) der Studierenden anhand ihres erfassten Verhaltens vorherzusagen. Neben den Variablen, die sich aus ihrem Anruf- und SMS-Verhalten sowie der Zahl ihrer Facebook-Freundschaften ableiten ließen, wurden auch Prädiktoren durch Messung ihres Mobilitätsverhaltens (mittels GPS-Verfolgung) und ihrer geosozialen Einbettung (mittels Bluetooth-Nähe) hergeleitet. Von den fünf Merkmalen konnte nur die Extraversion besser als durch Zufall vorhergesagt werden, was erneut die Grenzen einer Vorhersage der Persönlichkeit anhand des Verhaltens aufzeigt (Mønsted et al., 2018).

Szenario 2: Offenlegung von Mikro-Makro- Verhaltensdynamiken

Die zweite auf DVD beruhende Anwendungsart betrifft die Auswertung von Makroverhaltensmustern, deren Erkundung die CSS während ihrer datengetriebenen Anfänge umfangreich verfolgt hat. Dies ist zwar sinnvoll, jedoch liefert die reine Beschreibung der Makroebene keine Erklärung für ihre Genese. Das Versprechen von DVD besteht darin, dass sie als Makroskop zur Offenlegung von Mikro-

Makro-Verhaltensdynamiken dienen können. Makroverhaltensmuster sind als Ursachen von Mikroverhalten, Aggregate von Mikroverhalten oder Teil eines Ursachen und Aggregate umfassenden Mechanismus von Bedeutung. Die oben beschriebene Rückkopplungsdynamik ist ein solcher Mechanismus (Keuschnigg et al., 2018). Die Frage, ob die Mikroebene Vorrang vor der Makroebene hat oder umgekehrt, zählt zu den ältesten Fragestellungen der Sozialwissenschaften, und DVD liefern neue Antworten. Der entscheidende Schritt dabei ist die Ausnutzung der longitudinalen Eigenschaft von DVD.

Beispielsweise besteht ein großes Forschungsinteresse an Homophilie, dem häufig beobachteten Muster, dass soziale Beziehungen vorzugsweise zwischen Identitäten bestehen, die in einem oder mehreren Attributen übereinstimmen. Kossinets and Watts (2009) untersuchten 30.000 Personen an einer US-Hochschule und die 7 Millionen E-Mails, die diese in einem Zeitraum von 270 Tagen austauschten. Sie fanden heraus, dass sich zwei Personen umso näher im Kommunikationsnetzwerk stehen, je ähnlicher sie in Bezug auf ihre soziodemografischen Merkmale sind. Aber was sind die Ursprünge dieser Homophilie? Entsteht das Muster aus Mikroverhalten, weil Gleich und Gleich sich gern gesellt? Oder schränkt das Muster Mikroverhalten ein, so dass Gleich keine andere Wahl hat, als sich mit Gleich zu gesellen? Die Antwort der Autoren lautet: beides. Nach Untersuchung der Netzwerkdynamiken kamen die Autoren zu dem Ergebnis, dass Mikroverhalten und Makroverhaltensstrukturen ko-konstitutiv sind. Wenn voneinander entfernte, aber ähnliche Personen sich einander annähern, ermöglicht durch gemeinsame Aktivitäten wie Unterricht, konvergieren Nähe im Netzwerk und Attributähnlichkeit.

Ein weiterer Ansatz zur Modellierung der wechselseitigen Abhängigkeit von Mikro- und Makroebene sind generative Modelle. Stochastische Bayes'sche Blockmodelle von sozialen Netzwerken sind ein anschauliches Beispiel. Das Konzept von stochastischen

Blockmodellen besteht darin, dass Knoten in einem Netzwerk in einer Gruppe zusammengeführt werden, wenn sie das gleiche Muster von Verbindungen zu Knoten in anderen Gruppen aufweisen. Ein einfaches Beispiel sind Knoten in der Netzwerkperipherie, die mit Knoten im Netzwerkkern, aber nicht mit anderen peripheren Knoten verbunden sind. Der Bayes-Algorithmus lernt ein Blockmodell für ein Netzwerk unter der Annahme, dass das Blockmodell das Netzwerk erzeugt hat. Peixoto (2015) untersuchte ein Netzwerk persönlicher Interaktionen von High-School-Schüler*innen, die mithilfe von sozialen Sensoren mit hoher zeitlicher Auflösung erfasst wurden. Er zeigte, dass ein Blockmodell, in dem die Schüler*innen im Zeitverlauf denselben Gruppen zugeordnet werden, das dynamische Netzwerk am besten beschreibt. Dies impliziert, dass das Schüler*innen-Netzwerk seinen Makroverhaltenszustand fortlaufend über einen Rückkopplungsmechanismus reproduziert.

Diese Beispiele nutzen die soziale Dimension von Transaktionen. Aber die Analyse der natürlichen Sprache – der inhaltlichen Dimension von Transaktionen – ist eine direkte Möglichkeit zur Untersuchung der Ko-Konstitution von Mikro- und Makroverhalten (Bail, 2014). Wie im vorherigen Unterabschnitt treffen sich hier die Verarbeitung natürlicher Sprache und das maschinelle Lernen. Sogenannte „Topic Models“ sind generative Modelle zur automatisierten Textanalyse. Sie sind in der Lage, latente Makroverhaltensmuster offenzulegen, aus denen Transaktionen mutmaßlich entstehen. Es handelt sich um relationale Methoden, da Themen (Topics) aus zusammen verwendeten Wörtern bestehen. Beispielsweise gingen Stier et al. (2018) von den in repräsentativen Umfragen genannten wichtigsten Themen aus und fanden heraus, dass Politiker in sozialen Medien auch andere Themen ansprechen. Die digitalen Diskursmuster geben Auskunft über Mikroverhaltenspraktiken: Facebook wird stark für Wahlkämpfe genutzt, während Twitter als Kanal für politische Diskussionen dient.

Herausforderungen bei der Nutzung des „neuen Teleskops“

Das „neue Teleskop“ der Sozialwissenschaften ist eine Metapher für die Verfügbarkeit von riesigen DVD-Mengen und den entsprechenden Analysetools. DVD sind genuin relational, und etablierte Techniken, wie Netzwerk- und Textanalyse, nutzen diese Relationalität und zählen zu den Hauptpfeilern von CSS. Mit der zunehmenden Bedeutung dieses Felds müssen sich Sozialwissenschaftler*innen mit dem Potenzial, den Einschränkungen und den Herausforderungen der methodischen Innovationen von DVD vertraut machen. Unserer Ansicht nach bestehen mindestens fünf Hauptherausforderungen für die aktuelle und künftige Forschung.

- (1) **Datenmanagement.** Die Schlagkraft von Erklärungen in der CSS ergibt sich aus dem Volume, der Variety und Velocity von DVD. Jedoch macht gerade diese Komplexität es notwendig, das Management dieser Daten zu erlernen (z.B. stark relationale Informationen zu verknüpfen, aggregieren und analysieren). Sozialwissenschaftler*innen von heute müssen nicht nur (wie in den vergangenen Jahrzehnten) Statistikmethoden, sondern auch fortgeschrittene Techniken zum Umgang mit DVD beherrschen.
- (2) **Datenqualität.** Eine grundlegende Einschränkung von DVD besteht darin, dass die von den Nutzenden subjektiv beabsichtigten Bedeutung anhand von beobachtetem Verhalten nicht rekonstruiert werden kann. Noch grundlegender ist, dass einige Akteur*innen gegebenenfalls nicht einmal Menschen, sondern Bots sind. Somit gehen die Vorteile von DVD mit beträchtlichen Einschränkungen hinsichtlich der Validität von Konstrukten einher. Außerdem sind im Forschungszyklus verschiedene Mess- und Repräsentationsfehler möglich. Daher wird empfohlen, dass diese identifiziert, ausdrücklich aufgezeigt und Maßnahmen zu ihrer Abmilderung ergriffen werden (z.B.

DVD durch Befragungsdaten komplementieren; Sen et al., 2021; Schmitz & Riebling, im Erscheinen).

- (3) **Reproduzierbarkeit.** In letzter Zeit ist die Reproduzierbarkeit von Studien in der empirischen Sozialwissenschaft ein zunehmend wichtiges Thema geworden. Aber genau dies ist in Zusammenhang mit DVD häufig schwierig. Bei DVD gibt es einen Zielkonflikt zwischen dem Reichtum von Informationen zu sozialen Beziehungen und Kommunikationsinhalten und deren freien Verfügbarkeit. Es laufen einige Initiativen zur Öffnung der „Closed Shops“ privater Plattformbetreiber. Forschenden können aber bereits ihren Computercode teilen, um Reproduzierbarkeit wenigstens dadurch zu verbessern, dass sie es anderen ermöglichen, ihre Analysen erneut durchzuführen. Beispielsweise ermöglicht der GESIS Notebook Service unter notebooks.gesis.org die Ausführung von Computercode in einem Browserfenster, ohne dass eine Programmiersprache installiert werden muss (Cloud Computing).
- (4) **Reflexivität.** Kategorische Einstellungen gegenüber CSS – grundsätzliche Ablehnung oder unkritische Übernahme – sind gleichermaßen problematisch: Zwar müssen diese Daten und Methoden mobilisiert werden, aber dabei müssen die Bedingungen, unter denen sie erzeugt werden, sowie ihre analytischen Einschränkungen offengelegt werden. Um das „neue Teleskop“ auf angemessene Weise einsetzen zu können, müssen Wissenschaftler*innen die zugrunde liegenden Architekturen und die Modi der Datengeneration, einschließlich der grundlegenden Rolle von künstlicher Intelligenz und Maschinenverhalten bei der Beeinflussung und Erzeugung von sozialen Phänomenen, besser verstehen und reflektieren (Wagner et al., 2021). Nicht zuletzt hat die Arbeit mit DVD dabei auch eine ethische Komponente.
- (5) **Theorie.** Sofern organische DVD ein Nebenprodukt des Betriebs digitaler Plattformen sind, ist die gesamte Forschung mit diesen Daten unweigerlich in gewissem Maße datengetrieben. Nichtsdestotrotz gewährleistet alleine die Verwendung von Theorie,

dass Wissen auf sinnvolle und kumulative Weise produziert wird. In Anbetracht des Potenzials von CSS zur Lösung der drängenden Probleme auf globaler Ebene (z.B. Nachhaltigkeit) beizutragen, besteht eine Möglichkeit in der Entwicklung und Anwendung von lösungsorientierten Theorien mittlerer Reichweite (Watts, 2017). Allerdings stellen DVD auch ein vielversprechendes Forschungsfeld für die Vielzahl der sozialwissenschaftlichen Paradigmen dar: Neben der Analyse sozialer Netzwerke bieten die Sozialwissenschaften eine Fülle an etablierten und ausgearbeiteten Perspektiven auf das Soziale, wie Praktiken, Mechanismen, Diskurse, Systeme, Felder und Funktionen. DVD stellen einen vielversprechenden strategischen Forschungskontext für die Anwendung, Entwicklung und Anpassung dieser verschiedenen Theorien dar. Diese verschiedenen Ansätze können sich bei der Transzendierung des Individuums als Beobachtungseinheit als hilfreich erweisen. Sie können zu lebhaften Kommunikationswegen zwischen Sozialwissenschaft und Informatik und, auf diese Weise, zu einem Grundpfeiler für die Konsolidierung des Felds der CSS werden.

Literatur

- Bail, C.A. (2014). The cultural environment: Measuring culture with Big Data. *Theory and Society*, 43(3–4), 465–482.
- Diaz, F., Gamon, M., Hofman, J.M., Kıcıman, E., & Rothschild, D. (2016). Online and social media data as an imperfect continuous panel survey. *PLoS ONE*, 11(1), e0145406.
- Emirbayer, M. (1997). Manifesto for a relational sociology. *American Journal of Sociology*, 103(2), 281–317.
- Howison, J., Wiggins, A., & Crowston, K. (2011). Validity issues in the use of Social Network Analysis with Digital Trace Data. *Journal of the Association for Information Systems*, 12(12), 767–797.
- Keuschnigg, M., Lovsjö, N., & Hedström, P. (2018). Analytical Sociology and Computational Social Science. *Journal of Computational Social Science*, 1(1), 3–14.
- Kosinski, M., Stillwell, D., & Graepel, T. (2013). Private traits and attributes are predictable from digital records of human behavior. *Proceedings of the National Academy of Sciences*, 110(15), 5802–5805.

- Kossinets, G. & Watts, D.J. (2009). Origins of homophily in an evolving social network. *American Journal of Sociology*, 115(2), 405–450.
- Lazer, D. & Radford, J. (2017). Data ex machina: Introduction to Big Data. *Annual Review of Sociology*, 43(1), 19–39.
- Lazer, D., Pentland, A., Watts, D.J., Aral, S., Athey, S., Contractor, N., Freelon, D., Gonzalez-Bailon, S., King, G., Margetts, H., Nelson, A., Salganik, M.J., Strohmaier, M., Vespignani, A., & Wagner, C. (2020). Computational Social Science: Obstacles and opportunities. *Science*, 369(6507), 1060–1062.
- 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.
- Nassehi, A. (2019). *Muster: Theorie der digitalen Gesellschaft*. München: C.H. Beck.
- Peixoto, T.P. (2015). Inferring the mesoscale structure of layered, edge-valued, and time-varying networks. *Physical Review E*, 92(4), 042807.
- Schaible, J., Oliveira, M., Zens, M., & Génois, M. (2022). Sensing close-range proximity for studying face-to-face interaction. In U. Engel, Quan-Haase, A., Liu, X., & Lyberg, L. (Hrsg.). *Handbook of Computational Social Science* (Bd. 1, Kap. 14). London: Routledge.
- Schmitz, A. & Riebling, J. (im Erscheinen). Data quality of digital process data. A generalized framework and simulation/post-hoc-identification strategy. *Kölner Zeitschrift für Soziologie und Sozialpsychologie*.
- Sekara, V., Stopczynski, A., & Lehmann, S. (2016). Fundamental structures of dynamic social networks. *Proceedings of the National Academy of Sciences*, 113(36), 9977–9982.
- Sen, I., Flöck, F., Weller, K., Weiß, B., & Wagner, C. (2021). A total error framework for digital traces of human behavior on online platforms. *Public Opinion Quarterly*, 85(S1), 399–422.
- Stier, S., Bleier, A., Lietz, H., & Strohmaier, M. (2018). Election campaigning on social media: Politicians, audiences, and the mediation of political communication on Facebook and Twitter. *Political Communication*, 35(1), 50–74.
- Wagner, C., Strohmaier, M., Olteanu, A., Kıcıman, E., Contractor, N., & Eliassi-Rad, T. (2021). Measuring algorithmically infused societies. *Nature*, 595, 197–204.
- Watts, D.J. (2013). *Alles ist offensichtlich*: *sobald man die Antwort kennt. Wie uns der gesunde Menschenverstand täuscht*. Bern: Huber.
- Watts, D.J. (2017). Should social science be more solution-oriented? *Nature Human Behaviour*, 1, 0015.
- White, H.C. (2008). *Identity and Control: How Social Formations Emerge*. Princeton, NJ: Princeton University Press.

Haiko Lietz

GESIS – Leibniz-Institut für Sozialwissenschaften

E-Mail Haiko.Lietz@gesis.org

Haiko Lietz ist Postdoc bei GESIS, Köln. Er interessiert sich für die Anwendung und Weiterentwicklung der Netzwerktheorie und -methode durch die Integration von Soziologie, Komplexitätstheorie und computergestützter Ansätze.

Andreas Schmitz

GESIS – Leibniz-Institut für Sozialwissenschaften

E-Mail Andreas.Schmitz@gesis.org

Andreas Schmitz ist Forscher bei GESIS, Köln. Seine Hauptinteressen in der Forschung sind relationale Gesellschaftstheorie, relationale Methodologie, das Zusammenspiel von Computational Social Science und Sozialtheorie, angewandte Statistik und generalisierte Feldtheorie.

Johann Schaible

EU|FH - Europäische Fachhochschule Rhein / Erft GmbH

E-Mail j.schaible@eufh.de

Johann Schaible ist Professor für anwendungsorientierte Informatik an der Europäischen Fachhochschule (EU|FH), Brühl. Sein Forschungsschwerpunkt liegt im Bereich Smart Cities mit dem Fokus auf Mobilitätsforschung und allgemein auf der Analyse spatio-temporaler Daten.