

Big Data vs. Small Data?

Erforschung digitaler Kommunikation zwischen Komplexität und Vereinfachung

Für die Analyse von digitaler Kommunikation werden in den letzten Jahren unter dem Stichwort „Big Data“ neue Ansätze diskutiert. Einige sehen hier neue Möglichkeiten, gar ein neues Paradigma für die Sozialwissenschaften (Anderson, 2008; Lazer et al., 2009; Mayer-Schönberger & Cukier, 2013). Große Datenmengen aus Onlineplattformen würden klassische sozialwissenschaftliche Verfahren überflüssig machen. Stattdessen könne in „natürlichen“ Daten nach Mustern gesucht werden, um so auch völlig unerwartete Erkenntnisse zu gewinnen. Der Beitrag diskutiert die methodische Komplexität bzw. Einfachheit typischer Big Data-Studien und macht Vorschläge, wie die Kombination von Verfahren aus Big Data- und herkömmlicher Sozialforschung kombiniert werden können, um kommunikationswissenschaftlichen Forschungsanliegen besser gerecht zu werden.

In zweierlei Hinsicht bedingt das typische Vorgehen von Big Data-Forschung, dass derartige Studien stark zu Vereinfachung neigen. Zum einen sind die ausgelesenen Kategorien i.d.R. durch die untersuchte Plattform vorgegeben; nur dann sind sie automatisiert in großer Zahl auslesbar. Was diese Kategorien inhaltlich abbilden, ist häufig unklar (boyd & Crawford, 2012; Mahrt & Scharkow, 2013). Wenn zu einem YouTube-Video z.B. eine große Zahl an Nutzerkommentaren vorliegt, ist das Video dann populär oder kontrovers? Beziehen sich die Kommentare überhaupt auf das Video selbst oder hat sich die Unterhaltung zwischen den Nutzerinnen und Nutzern längst von dessen Inhalt entkoppelt? In aggregierten Werten wie der Anzahl von Klicks auf Schaltflächen einer Plattform können sich also eine Vielzahl von Nutzerinteraktionen oder -motivationen ausdrücken.

Die Datenstruktur von Big Data führt häufig auch zu wenig komplexen Auswertungen. Typische unabhängige, intervenierende oder Kontrollvariablen sind meist nicht vorhanden, bspw. Informationen über die Personen, deren Verhalten online Spuren hinterlassen hat. So finden sich in Big Data-Studien häufig deskriptive Auswertungen oder Visualisierungen, etwa von Netzwerkstrukturen. Selten sind dagegen kausale Analysen, z.B. zu den Auswirkungen, die bestimmte Verteilungen oder Netzwerke haben.

Mehr Komplexität täte Big Data an vielen Stellen daher gut. Während komplexe Analysen und Designs in sozialen Netzwerkseiten vermutlich den Forschungsabteilungen der

entsprechenden Unternehmen vorbehalten bleiben werden (etwa Bakshy, Messing, & Adamic, 2015), stellt sich ansonsten die Frage, ob statt Big nicht eher „Small Data“ ein fruchtbarer Ansatz für die Kommunikationswissenschaft sind.

Mit Blick auf YouTube sollten nicht nur aggregierte Zahlen zur Häufigkeit von Nutzung, Kommentaren, positiver oder negativer Bewertung untersucht werden. Rotman, Golbeck und Preece (2009) zeigen bspw., wie sehr sich die quantitative, datenorientierte Netzwerkstruktur der Plattform von der qualitativen Sicht der Nutzerinnen und Nutzer unterscheidet. Ebenso können der Inhalt der angebotenen bzw. gesehenen Videos oder die Eigenschaften der Nutzerinnen und Nutzer sehr viel mehr Aufschluss über die Funktionen geben, die die Videoplattform erfüllen kann.

Diese Funktionen sollen mit einerseits Daten einer manuellen Inhaltsanalyse von YouTube-Videos und andererseits Nutzungsdaten der Plattform aufgezeigt werden. Zur Nutzung von YouTube wird ein Datensatz der Marktforschungsfirma Nielsen ausgewertet, der für die Teilnehmer ihres „Clickstream“-Panels alle im Juni 2014 aufgerufenen YouTube-URLs ausweist. Es liegen damit Informationen zu 8.147 Personen vor, die insgesamt über 430.000mal auf YouTube zugegriffen haben. In diesem großen Datensatz werden aber weniger allgemeine Verteilungsmuster, etwa der Beliebtheit, untersucht. Stattdessen geht es um die Erkundung der Publika, die sich um bestimmte Inhalte und Kanäle auf YouTube bilden, und damit um das Integrations- oder Fragmentierungspotenzial der Plattform. Abb. 1 zeigt für einen kleinen Ausschnitt, die zehn beliebtesten Videos (ohne Musikvideos), wie sich die Nutzerschaften dieser Inhalte überschneiden. Es zeigen sich vergleichsweise große Schnittmengen für Videos desselben Kanals (z.B. Blau markiert für den YouTuber „LeFloid“), während ein viraler Erfolg eines WDR-Ausschnitts aus der „Sendung mit der Maus“ (Rot markiert) zwar eine ähnlich große, aber völlig anders zusammengesetzte Nutzerschaft erreicht.

Schon dieser Ausschnitt zeigt, dass „Erfolg“ auf YouTube auf ganz unterschiedliche Weise zustandekommen kann. Weitere Analysen der Nielsen-Daten dauern an, insbesondere zur Zusammensetzung der jeweiligen Publika von Kanälen und Inhalten. Dies soll zeigen, dass es jenseits von Big Data andere fruchtbare Ansätze zur Erforschung von Onlineplattformen gibt, während man große Datenmengen nicht zwingend in Big Data-Manier auswerten muss.

Literatur

- Anderson, C. (2008, June 23). The end of theory: The data deluge makes the scientific method obsolete. *Wired*. Abgerufen am 3. Juni 2015 von http://archive.wired.com/science/discoveries/magazine/16-07/pb_theory/
- Bakshy, E., Messing, S., & Adamic, L. (2015). Exposure to ideologically diverse news and opinion on Facebook. *Science*. doi: 10.1126/science.aaa1160
- boyd, d., & Crawford, K. (2012). Critical questions for Big Data. Provocations for a cultural, technological, and scholarly phenomenon. *Information, Communication & Society*, 15(5), 662-679. doi: 10.1080/1369118x.2012.678878
- Lazer, D., Pentland, A., Adamic, L., Aral, S., Barabási, A.-L., Brewer, D., . . . Van Alstyne, M. (2009). Computational social science. *Science*, 323(5915), 721-723. doi: 10.1126/science.1167742
- Mahrt, M., & Scharrow, M. (2013). The value of big data in digital media research. *Journal of Broadcasting & Electronic Media*, 57(1), 20-33. doi: 10.1080/08838151.2012.761700
- Mayer-Schönberger, V., & Cukier, K. (2013). *Big data. A revolution that will transform how we live, work, and think*. New York, NY: Houghton Mifflin Harcourt.
- Rotman, D., Golbeck, J., & Preece, J. (2009). *The community is where the rapport is – On sense and structure in the YouTube community*. Proceedings of the fourth international conference on Communities and technologies, University Park, PA. doi: 10.1145/1556460.1556467

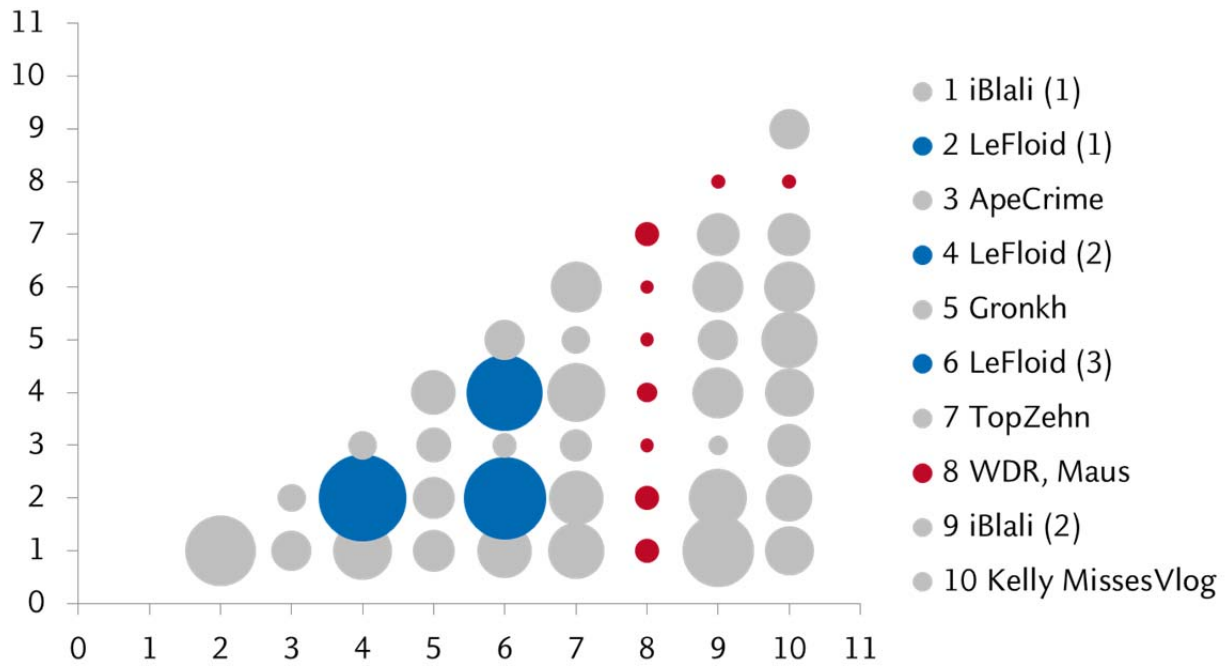


Abb. 1: Paarweise Überlappung der Nutzerschaften von zehn beliebten YouTube-Videos