

Sozialwissenschaften durch Big Data? Ein "systematic review" zu Methoden, Daten und Fragestellungen der Computational Social Science

Der Begriff "Computational Social Science" (CSS) beschreibt interdisziplinäre Forschungsansätze an der Schnittstelle der Sozialwissenschaften mit der Informatik, der Physik komplexer Systeme und der angewandten Mathematik (vgl. Golder und Macy, 2014, Lazer et al., 2009, Strohmaier und Wagner, 2014). In CSS-Ansätzen werden statistische und computergestützte Methoden wie Clustering, Association Rule Mining, Topic Modeling und Graphenanalyse für die Bearbeitung sozialwissenschaftlicher Fragestellungen miteinander kombiniert und anhand großer Datensätze bearbeitet, oftmals mit dem Ziel, prädiktive Modelle zu entwickeln. Aus einer anwendungsorientierten Perspektive lassen sich so beispielsweise Vorhersagen über Konsumverhalten und Wählerpräferenzen treffen, und eröffnen sich andererseits auch interessante theoretische Optionen, etwa bei der Untersuchung des Zusammenhangs von Online- und Offline-Protesten (Bastos, Mercea und Charpentier, 2015), oder mit Blick auf die Vorhersagbarkeit von Informationskaskaden (Cheng et al., 2014). Das wachsende wissenschaftliche Interesse an CSS lässt sich unter anderem an der Veröffentlichung von mehreren Special Issues in diesem und letztem Jahr zu diesem Thema erkennen (*Annals of the American Academy of Political and Social Science*, *Annual Review of Sociology*, *Journal of Communication*), aber auch an der zunehmenden Anzahl von Tagungen und Studiengängen mit Bezügen zu CSS. Die Verfügbarkeit von Daten aus sozialen Medien befeuert diese Tendenz zusätzlich. Gleichzeitig ist die Methodologie der Computational Social Science jedoch noch vergleichsweise wenig standardisiert, was sowohl Kritik an der Validität bestimmter Forschungsdesigns provoziert, als auch ethische und rechtliche Fragen aufwirft (Jungherr, 2015, Mahrt und Scharkow, 2013). Die populärsten Verfahren weisen einen explorativen Grundansatz auf, der die Replikation von Studien und die Generalisierung von Ergebnissen häufig erschwert. Auch ist die Anschlussfähigkeit an bestehende Theorien und Frameworks oftmals schwierig, da nur solche Konzepte für die CSS nutzbar sind, die sich klar operationalisieren und quantifizieren lassen. Daraus ergeben sich auch deshalb interessante Fragen für die Kommunikations- und Medienwissenschaft, weil in der CSS oftmals Kommunikation als eine Form des Verhaltens operationalisiert wird, aus der sich etwa die Merkmale von Akteuren (bspw. Einfluss, Status, Engagement) ableiten lassen. Auch ergeben sich Spannungen zwischen etablierten Methoden der Medienanalyse und computergestützten Ansätzen, die mitunter miteinander in Konkurrenz treten (vgl. Zamith und Lewis, 2015).

Kern meines Vortrags ist eine Auswertung von 60 häufig zitierten Publikationen aus dem Feld der Computational Social Science, mit dem Ziel, zentrale Methoden, Datentypen und Fragestellungen zu identifizieren. Dabei wird der Ansatz des *systematic reviews* (Petticrew und Roberts, 2006) einerseits gewählt, weil die CSS disziplinübergreifend und international operiert, und somit eine breite Auswahl wissenschaftlicher Quellen vonnöten ist, um ihrer Vielfalt von Datenquellen, Methoden und Fragestellungen gerecht zu werden, und andererseits, um auf die zum Teil erheblichen Unterschiede zwischen den Ergebnissen einzelner Studien hinzuweisen, welche den emergenten Charakter von CSS unterstreichen. Die Samplinggrundlage meiner Untersuchung bilden dabei einerseits Proceedings maßgeblichen Konferenzen aus der Informatik in diesem Bereich (*ACM Web Science*, *ACM Hypertext*, *ICSWM* und *WWW*) und andererseits die o.g. Special Issues aus etablierten sozialwissenschaftlichen Journals. Im Mittelpunkt steht die Frage, wie innerhalb des Paradigmas der Computational Social Science wissenschaftliches Wissen generiert wird, und

welche Annahmen und Grundlagen dabei eine Rolle spielen. Dabei grenze ich erstens CSS als distinktive wissenschaftliche Community mit einem eigenständigen Ansatz von "Big Data" im seiner inzwischen sehr unscharfen und allgemeinen Bedeutung ab. Zweitens gehe ich auf den emergenten Charakter von CSS ein, der derzeit auch deshalb in der Kritik steht, weil er kaum die Untersuchung kausaler Zusammenhänge erlaubt und deshalb häufig als stark interpretativ empfunden wird. Drittens wird der Operationalisierungsansatz und die Theoriebindung von CSS kritisch diskutiert. Viertens werde ich den Binnencharakter von CSS zwischen angewandter Industrieforschung und theoretisch orientierter akademischer Forschung hervorheben und davor warnen, die Bedeutung der Computational Social Science wegen ihrer Anwendungsnähe zu unterschätzen.

Quellen

- Bastos, M. T., Mercea, D., & Charpentier, A. (2015). Tents, tweets, and events: The interplay between ongoing protests and social media. *Journal of Communication*, 65(2), 320–350. doi:10.1111/jcom.12145
- Cheng, J., Adamic, L. A., Dow, P. A., Kleinberg, J., & Leskovec, J. (2014). Can cascades be predicted? In W. Lee, H.-C. Rim, & D. Schwabe (Eds.), *Proceedings of the 23rd International World Wide Web Conference (WWW '14)* (pp. 1–11). Seoul, Republic of Korea: ACM Press. doi:10.1145/2566486.2567997
- Golder, S. A., & Macy, M. W. (2014). Digital footprints: Opportunities and challenges for online social research. *Annual Review of Sociology*, 40, 129–152. doi:10.1146/annurev-soc-071913-043145
- Jungherr, A. (2015). *Analyzing Political Communication with Digital Trace Data*. Heidelberg: Springer International Publishing. doi:10.1007/978-3-319-20319-5
- Lazer, D., Pentland, A., Adamic, L. A., Aral, S., Barabasi, A.-L., Brewer, D., ... Van Alstyne, M. (2009). Computational social science. *Science*, 323(5915), 721–723. doi:10.1126/science.1167742
- Mahrt, M., & Scharkow, 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
- Petticrew, M., & Roberts, H. (2006). *Systematic Reviews in the Social Sciences: A Practical Guide*. Malden, MA: Blackwell Publishing.
- Strohmaier, M., & Wagner, C. (2014). Computational social science for the world wide web. *IEEE Intelligent Systems*, 29(5), 84–88. doi:10.1109/MIS.2014.80
- Zamith, R., & Lewis, S. C. (2015). Content Analysis and the algorithmic coder: What Computational Social Science means for traditional modes of media analysis. *Annals of the American Academy of Political and Social Science*, 659(1), 307–318. doi:10.1177/0002716215570576