Using Digital Trace Data in the Social Sciences: Syllabus

Junprof. Dr. Andreas Jungherr

Room: D 307
Postfach 85
Universitätsstraße 10
78464 Konstanz

Phone: +49-7531-882150
Email: andreas.jungherr@gmail.com

Office hours: By appointment

Secretary:
Karin Becker
Room: D 312
Phone: +49-7531-882602
Email: office.cdm@uni-konstanz.de

Course website:
`http://ajungherr.github.io/dtd-konstanz-summer2018`

1 Course Description

In the course, students will learn fundamental techniques of data collection preparation, and analysis with digital trace data in the social sciences. In this, we will focus on working with the microblogging-service Twitter. Over the course, students are expected to become proficient in the use of two programming languages, Python and R.

Level: Create—Students are expected to independently perform theory-driven data collections on the microblogging-service Twitter and use these data in the context of a series of specified prototypical analyses.

Vst.-Nr. & ECTS-Punkte:
POL-19640: Vertiefungsseminar—6 ECTS
POL-19650: Seminar—7 ECTS
POL-19630: Doktorandenseminar—4 ECTS / 6 ECTS (GSDS)

2 Housekeeping

The goal of the course is to enable students to independently design a research question based on current theory and field-specific debates, collect data on Twitter, and run a fitting data analysis on said data. The course itself will focus heavily on the practicalities of collecting data and running selected types of analyses on said data. Information,

readings, and example scripts for the respective sessions will be posted on the course website http://ajungherr.github.io/dtd-konstanz-summer2018.

The course follows closely a tutorial written by Pascal Jürgens and me, A Tutorial for Using Twitter Data in the Social Sciences: Data Collection, Preparation, and Analysis. The tutorial is freely available on the Social Science Research Network (SSRN) at https://papers.ssrn.com/sol3/papers.cfm?abstract_id=2710146. I recommend all participants in the course to download the tutorial and the accompanying set of scripts available at https://github.com/trifle/twitterresearch. You will very likely profit from preparing the respective sections of the tutorial before and after the corresponding session.

– Pascal Jürgens and Andreas Jungherr. 2016. *A Tutorial for Using Twitter-Data in the Social Sciences: Data Collection, Preparation, and Analysis*. Social Science Research Network (SSRN). doi:10.2139/ssrn.2710146.

The articles in readings are linked to from the syllabus and should be available to you by using your Uni-Konstanz VPN-access. The books listed are not linked to from the document but should be available to you through the dedicated online services of the respective publishers. Books by O'Reilly, PACKT, and Apress are available through the Proquest Safari-Books Online-Shelf, at http://proquest.techbus.safaribooksonline.de. Access to Safari-Books Online is freely available by using your Uni-Konstanz VPN-access.

3 Requirements

3.1 Course Work

1. Regular and active participation.
2. Independent study of Python and R in order to follow the example scripts.
3. Regular independent follow-up of example scripts introduced in the respective sessions.
4. Presentation: In the last weeks of the course, you will be asked to present the research project forming the basis of your term paper. For this presentation please prepare a slide deck introducing your research question, your motivation, proposing a mechanism linking signals you hope to find on Twitter to your phenomenon of interest, your proposed approach to the data collection, and open questions. This presentation will take approximately 10 minutes and will be followed by a quick round of feedback from the other participants.

3.2 Term Paper

Your grade for the course will be based on a paper that you hand in following the course. For the paper, you will be asked to perform and report an independent data analysis based on data collected on Twitter by you on a research question chosen by you. The aim of this paper is for you to demonstrate that you are able to independently apply and adapt the techniques learned during the course in the context of a specific research question developed by you.

- Style requirements: Font—Times New Roman, 12pt; Line-separation—1.5; Page borders—2.5 cm left and right, 2cm above and below; Page set—Block; The first line of each paragraph is indented.
- Citation Style: Please follow the citation convention of the American Political Science Review (APSR) available at `http://www.apsanet.org/APSR-Submission-Guidelines-August-2016`.
- Cover page: University, department, course title, paper title, name, Matriknr., semester count, study program, and e-mail-address.
- Length: ca. 4000 words +/-10%
- Deadline: Please return the paper on the date specified by the department (BA: 15. September; MA: 15. September) electronically at andreas.jungherr@gmail.com and by hardcopy with Karin Becker (Room D 312). The date is mandatory and can only be extended in case of officially certified illness.

3.3 Reading Papers

Although the course will not feature the dedicated discussion of research papers, you are advised to read at least a selection of the papers listed for each session. This will allow you to familiarize with some of the most prominent research papers using digital trace data. This familiarity will allow you to formulate better research questions and facilitate the work on your term paper. Make sure to use the time during the semester to familiarize yourself with the papers listed in the syllabus.

In reading the paper keep the following questions in mind:

1. What is the research question?
2. What is the phenomenon under examination?
3. How explicitly does the author link their research question to concepts in the social sciences?
4. What signals in digital trace data does the author take as indicators of the phenomenon of interest?
5. What is the mechanism that leads the author to expect signals found in digital trace data to be linked to the phenomenon of interest?
6. To which literature does the article contribute?
7. What is the method of analysis? How are data collected? How appropriate are the method and data?
8. What are the empirical findings? How convincingly are they linked to the phenomenon ostensively of interest? How transparent is the author in discussing limitations arising from the mediated nature of digital trace data?

Further, in assessing causal claims put forward in papers using digital trace data make sure to ask the following questions:[1]

1. Is there a plausible mechanism for the effect?
2. Does evidence come from peer-reviewed sources?
3. Are all relevant studies considered?
4. Are results of specific studies misrepresented?
5. Are causal claims based on experiment, correlation or analogy?
6. Is technical, scientific terminology used to obfuscate rather than clarify?

You may find it useful to keep notes on the papers read by you.

[1] Questions slightly shortened from David Robert Grimes and Dorothy V. M. Bishop. 2018. "Distinguishing Polemic From Commentary in Science: Some Guidelines Illustrated With the Case of Sage and Burgio (2017)". *Child Development* 89 (1): 141–147. doi:10.1111/cdev.13013

4 Course Outline

Class will meet at the following times and locations:

Thursday 15:15-16:45 (C421)

4.1 Week 1: Introduction and Conceptual Issues in the Use of Digital Trace Data in Social Science (April 19)

4.2 Week 2: Set Up and Introduction to Collecting Data on Twitter (April 26)

4.3 Week 3: Introduction to Python (May 3)

4.4 Week 4: Christi Himmelfahrt (May 10)—no meeting

4.5 Week 5: Collecting Data Through Twitter's API (May 17)

4.6 Week 6: How to Find A Research Question? & Data Lab (May 24)

4.7 Week 7: Fronleichnam (May 31)—no meeting

4.8 Week 8: Loading Twitter Data Into a Database (June 7)

4.9 Week 9: Sample Analyses: Counts & Time Series (June 14)

4.10 Week 10: Sample Analyses: Networks (June 21)

4.11 Week 11: Independent Study and Preparation of Presentations (June 28)—no meeting

4.12 Week 12: Student Presentations I. (July 5)

4.13 Week 13: Student Presentations II. (July 12)

4.14 Week 14: Student Presentations III. & Where to take it from here? Discussion of Open Questions and Paper (July 19)

4.1 Week 1: Introduction and Conceptual Issues in the Use of Digital Trace Data in Social Science (April 19)

Required Readings:

– Andreas Jungherr. 2018. "Normalizing Digital Trace Data". In *Digital Discussions: How Big Data Informs Political Communication*, ed. by Natalie Jomini Stroud and Shannon C. McGregor. New York, NY: Routledge. http://andreasjungherr.net/wp-content/uploads/2017/05/Jungherr-2017-Normalizing-Digital-Trace-Data-Preprint.pdf.

– Pages 7-14 from Pascal Jürgens and Andreas Jungherr. 2016. *A Tutorial for Using Twitter-Data in the Social Sciences: Data Collection, Preparation, and Analysis*. Social Science Research Network (SSRN). doi:10.2139/ssrn.2710146.

Background Readings:

– David Donoho. 2015. "50 Years of Data Science". In *Paper Presented at the Tukey Centennial Workshop*. Princeton, NJ. http://courses.csail.mit.edu/18.337/2015/docs/50YearsDataScience.pdf.

– Bradley Efron and Trevor Hastie. 2016. *Computer Age Statistical Inference: Algorithms, Evidence, and Data Science*. Cambridge, UK: Cambridge University Press.

– Scott A. Golder and Michael W. Macy. 2014. "Digital Footprints: Opportunities and Challenges for Online Social Research". *Annual Review of Sociology* 40:129–152. doi:10.1146/annurev-soc-071913-043145.

– James Howison, Andrea Wiggins, and Kevin Crowston. 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.

– Andreas Jungherr. 2015. *Analyzing Political Communication with Digital Trace Data: The Role of Twitter Messages in Social Science Research*. Cham, CH: Springer. doi:10.1007/978-3-319-20319-5.

– Andreas Jungherr and Pascal Jürgens. 2013. "Forecasting the pulse: How deviations from regular patterns in online data can identify offline phenomena". *Internet Research* 23 (5): 589–607. doi:10.1108/IntR-06-2012-0115.

– Andreas Jungherr, Harald Schoen, and Pascal Jürgens. 2016. "The Mediation of Politics Through Twitter: An Analysis of Messages Posted During the Campaign for the German Federal Election 2013". *Journal of Computer-Mediated Communication* 21 (1): 50–68. doi:10.1111/jcc4.12143.

– Andreas Jungherr, Harald Schoen, Oliver Posegga, and Pascal Jürgens. 2017. "Digital Trace Data in the Study of Public Opinion: An Indicator of Attention Toward Politics Rather Than Political Support". *Social Science Computer Review* 35 (3): 336–356. doi:10.1177/0894439316631043.

– David Lazer, Alex Pentland, Lada Adamic, Sinan Aral, Albert-László Barabási, Devon Brewer, Nicholas Christakis, Noshir Contractor, James Fowler, Myron Gutmann, Tony Jebara, Gary King, Michael W. Macy, Deb Roy, and Marshall Van Alstyne. 2009. "Computational Social Science". *Science* 323 (5915): 721–723. doi:10.1126/science.1167742.

– David Lazer, Ryan Kennedy, Gary King, and Alessandro Vespignani. 2014. "The Parable of Google Flu: Traps in Big Data Analysis". *Science* 343 (6176): 1203–1205. doi:10.1126/science.1248506.

– David Lazer and Jason Radford. 2017. "Data ex Machina: Introduction to Big Data". *Annual Review of Sociology* 43:19–39. doi:10.1146/annurev-soc-060116-053457.

– Viktor Mayer-Schönberger and Kenneth Cukier. 2013. *Big Data: A Revolution that Will Transform How We Live, Work, and Think*. New York, NY: Houghton Mifflin.

– Cornelius Puschmann and Jean Burgess. 2013. "The politics of Twitter data". In *Twitter and Society*, ed. by Katrin Weller, Axel Bruns, Jean Burgess, Marja Mahrt, and Cornelius Puschmann, 43–54. New York, NY: Peter Lang Publishing.

– Richard Rogers. 2013b. *Digital Methods*. Cambridge, MA: The MIT Press.

– Richard Rogers. 2013a. "Debanalizing Twitter: The transformation of an object of study". In *WebSci 2013: Proceedings of the 5th Annual ACM Web Science Conference*, ed. by Hugh Davis, Harry Halpin, Alex Pentland, Mark Bernstein, and Lada Adamic, 356–365. New York, NY: ACM. doi:10.1145/2464464.2464511.

– Derek Ruths and Jürgen Pfeffer. 2014. "Social media for large studies of behavior". *Science* 346 (6213): 1063–1064. doi:10.1126/science.346.6213.1063.

– Matthew J. Salganik. 2017. *Bit By Bit: Social Research in the Digital Age*. Princeton, NJ: Princeton University Press.

– Markus Strohmaier and Claudia Wagner. 2014. "Computational Social Science for the World Wide Web". *IEEE Intelligent Systems* 29 (5): 84–88. doi:10.1109/MIS.2014.80.

4.2 Week 2: Set Up and Introduction to Collecting Data on Twitter (April 26)

Required Readings:
– Pages 15-20 from Pascal Jürgens and Andreas Jungherr. 2016. *A Tutorial for Using Twitter-Data in the Social Sciences: Data Collection, Preparation, and Analysis*. Social Science Research Network (SSRN). doi:10.2139/ssrn.2710146.

Background Readings:
– Peter Bell and Brent Beer. 2018. *Introducing GitHub: A Non-Technical Guide*. 2nd ed. Sebastopol, CA: O'Reilly Media.

– Nick Eubank. 2015. *Data Analysis in Python*. http://www.data-analysis-in-python.org.

– Matthew A. Russell. 2018. *Mining the Social Web*. 3rd ed. Sebastopol, CA: O'Reilly Media.

4.3 Week 3: Introduction to Python (May 3)

Required Readings:
– Nick Eubank. 2015. *Data Analysis in Python*. http://www.data-analysis-in-python.org.

Background Readings:
– Naomi R. Ceder. 2018. *The Quick Python Book*. 3rd ed. Greenwich, CT: Manning

Publications Co.

– Swaroop Chitlur. 2016. *A Byte of Python*. https://www.gitbook.com/book/swaroopch/byte-of-python/details.

– Mark Lutz. 2013. *Learning Python*. 5th ed. Sebastopol, CA: O'Reilly Media.

– Wes McKinney. 2017. *Python for Data Analysis: Data Wrangling with Pandas, NumPy, and IPython*. 2nd ed. Sebastopol, CA: O'Reilly Media.

– Sebastian Raschka and Vahid Mirjalili. 2017. *Python Machine Learning: Machine Learning and Deep Learning with Python, scikit-learn, and TensorFlow*. 2nd ed. Birmingham, UK: PACKT Publishing.

– Albert Sweigart. 2015. *Automate the Boring Stuff with Python: Practical Programming for Total Beginners*. San Francisco, CA: No Starch Press. https://automatetheboringstuff.com.

4.4 Week 4: Christi Himmelfahrt (May 10)—no meeting

This is an official holiday, so no course meeting on this day. Maybe use the time to further familiarize yourself with Python in independent study.

4.5 Week 5: Collecting Data Through Twitter's API (May 17)

Required Readings:

– Pages 21-28 from Pascal Jürgens and Andreas Jungherr. 2016. *A Tutorial for Using Twitter-Data in the Social Sciences: Data Collection, Preparation, and Analysis*. Social Science Research Network (SSRN). doi:10.2139/ssrn.2710146.

Background Readings:

– Ryan Mitchell. 2018. *Web Scraping with Python: Collecting More Data from the Modern Web*. 2nd ed. Sebastopol, CA: O'Reilly Media.

– Matthew A. Russell. 2018. *Mining the Social Web*. 3rd ed. Sebastopol, CA: O'Reilly Media.

4.6 Week 6: How to Find A Research Question? & Data Lab (May 24)

Required Readings:

– Andreas Jungherr. 2016. "Twitter use in election campaigns: A systematic literature review". *Journal of Information Technology & Politics* 13 (1): 72–91. doi:10.1080/19331681.2015.1132401.

Background Readings:

How to Find a Research Question:

– Howard S. Becker. 1998. *Tricks of the Trade: How to Think About Your Research While You're Doing It*. Chicago, IL: The University of Chicago Press.

– Andreas Jungherr and Yannis Theocharis. 2017. "The Empiricist's Challenge: Asking

Meaningful Questions in Political Science in the Age of Big Data". *Journal of Information Technology & Politics* 14 (1): 97–109. doi:10.1080/19331681.2017.1312187.

Conceptual Issues in Working with Digital Trace Data:

– Christopher Bail. 2014. "The cultural environment: measuring culture with big data". *Theory and Society* 43 (3-4): 465–482. doi:10.1007/s11186-014-9216-5.

– Fernando Diaz, Michael Gamon, Jake M. Hofman, Emre Kıcıman, and David Rothschild. 2016. "Online and Social Media Data As an Imperfect Continuous Panel Survey". *PLoS One* 11 (1): e0145406. doi:10.1371/journal.pone.0145406.

– James Howison, Andrea Wiggins, and Kevin Crowston. 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.

– Lilli Japec, Frauke Kreuter, Marcus Berg, Paul Biemer, Paul Decker, Cliff Lampe, Julia Lane, Cathy O'Neil, and Abe Usher. 2015. "Big Data in Survey Research: AAPOR Task Force Report". *Public Opinion Quarterly* 79 (4): 839–880. doi:10.1093/poq/nfv039.

– Andreas Jungherr, Harald Schoen, and Pascal Jürgens. 2016. "The Mediation of Politics Through Twitter: An Analysis of Messages Posted During the Campaign for the German Federal Election 2013". *Journal of Computer-Mediated Communication* 21 (1): 50–68. doi:10.1111/jcc4.12143.

– Andreas Jungherr, Harald Schoen, Oliver Posegga, and Pascal Jürgens. 2017. "Digital Trace Data in the Study of Public Opinion: An Indicator of Attention Toward Politics Rather Than Political Support". *Social Science Computer Review* 35 (3): 336–356. doi:10.1177/0894439316631043.

– David Lazer, Ryan Kennedy, Gary King, and Alessandro Vespignani. 2014. "The Parable of Google Flu: Traps in Big Data Analysis". *Science* 343 (6176): 1203–1205. doi:10.1126/science.1248506.

– Filipe N. Ribeiro, Matheus Araújo, Pollyanna Gonçalves, Marcos André Gonçalves, and Fabricio Benevenuto. 2016. "SentiBench–a benchmark comparison of state-of-the-practice sentiment analysis methods". *EPJ Data Science* 5 (23): 1–29. doi:10.1140/epjds/s13688-016-0085-1.

– Matthew J. Salganik. 2017. *Bit By Bit: Social Research in the Digital Age*. Princeton, NJ: Princeton University Press.

– Michael F. Schober, Josh Pasek, Lauren Guggenheim, Cliff Lampe, and Frederick G. Conrad. 2016. "Social Media Analyses for Social Measurement". *Public Opinion Quarterly* 80 (1): 180–211. doi:10.1093/poq/nfv048.

Case Studies Illustrating Different Approaches to the Use of Twitter data:

– R. Michael Alvarez, ed. 2016. *Computational Social Science: Discovery and Prediction*. New York, NY: Cambridge University Press. doi:10.1017/CBO9781316257340.

– Christopher Bail, Lisa Argyle, Taylor Brown, John Bumpus, Haohan Chen, M.B. Hunzaker, Jaemin Lee, Marcus Mann, Friedolin Merhout, and Alexander Volfovsky. 2018. "Exposure to Opposing Views can Increase Political Polarization: Evidence from a Large-Scale Field Experiment on Social Media". *SocArXiv*. doi:10.17605/OSF.IO/4YGUX.

– Pablo Barberá. 2015. "Birds of the same feather tweet together: Bayesian ideal point estimation using Twitter data". *Political Analysis* 23 (1): 76–91. doi:10.1093/pan/mpu011.

– Marco T. Bastos, Dan Mercea, and Arthur Charpentier. 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.

– Marco Bastos, Carlo Piccardi, Michael Levy, Neil McRoberts, and Mark Lubell. 2018. "Core-periphery or decentralized? Topological shifts of specialized information on Twitter". *Social Networks* 52 (1): 282–293. doi:10.1016/j.socnet.2017.09.006.

– Marco Bastos and Dan Mercea. 2018. "Parametrizing Brexit: Mapping Twitter political space to parliamentary constituencies". *Information, Communication & Society*. doi:10.1080/1369118X.2018.1433224.

– Michael D. Conover, Jacob Ratkiewicz, Matthew Francisco, Bruno Goncalves, Alessandro Flammini, and Filippo Menczer. 2011. "Political Polarization on Twitter". In *ICWSM 2011: Proceedings of the 5th International AAAI Conference on Weblogs and Social Media*, ed. by Nicolas Nicolov, James G. Shanahan, Lada Adamic, Ricardo Baeza-Yates, and Scott Counts, 89–96. Menlo Park, CA: Association for the Advancement of Artificial Intelligence (AAAI).

– Peter S. Dodds, Kameron Decker Harris, Isabel M. Klouman, Catherine A. Bliss, and Christopher M. Danforth. 2011. "Temporal Patterns of Happiness and Information in a Global-Scale Social Network: Hedonometrics and Twitter". *PLoS One* 6 (12): 1–26. doi:10.1371/journal.pone.0026752.

– Elizabeth Dubois and Devin Gaffney. 2014. "The Multiple Facets of Influence: Identifying Political Influentials and Opinion Leaders on Twitter". *American Behavioral Scientist* 58 (10): 1260–1277. doi:10.1177/0002764214527088.

– Sharad Goel, Ashton Anderson, Jake Hofman, and Duncan J. Watts. 2016. "The Structural Virality of Online Diffusion". *Management Science* 62 (1): 180–196. doi:10.1287/mnsc.2015.2158.

– Todd Graham, Marcel Broersma, Karin Hazelhoff, and Guido van 't Haar. 2013. "Between broadcasting political messages and interacting with voters: The use of Twitter during the 2010 UK general election campaign". *Information, Communication & Society* 16 (5): 692–716. doi:10.1080/1369118X.2013.785581.

– Andreas Jungherr. 2014. "The Logic of Political Coverage on Twitter: Temporal Dynamics and Content". *Journal of Communication* 64 (2): 239–259. doi:10.1111/jcom.12087.

– Andreas Jungherr. 2015. *Analyzing Political Communication with Digital Trace Data: The Role of Twitter Messages in Social Science Research*. Cham, CH: Springer. doi:10.1007/978-3-319-20319-5.

– Andreas Jungherr and Pascal Jürgens. 2014. "Through a glass, darkly: Tactical support and symbolic association in Twitter messages commenting on Stuttgart 21". *Social Science Computer Review* 32 (1): 74–89. doi:10.1177/0894439313500022.

– Pascal Jürgens, Andreas Jungherr, and Harald Schoen. 2011. "Small worlds with a difference: New gatekeepers and the filtering of political information on Twitter". In *WebSci 2011: Proceedings of the 3rd International Web Science Conference*, ed. by David De Roure and Scott Poole. 21. New York: ACM. doi:10.1145/2527031.2527034.

– Daniel Kreiss. 2016. "Seizing the Moment: The Presidential Campaigns' Use of Twitter During the 2012 Electoral Cycle". *New Media & Society* 18 (8): 1473–1490. doi:10.1177/1461444814562445.

– Yu-Ru Lin, Brian Keegan, Drew Margolin, and David Lazer. 2014. "Rising tides or rising stars? Dynamics of shared attention on Twitter during media events". *PLoS One* 9 (5): e94093. doi:10.1371/journal.pone.0094093.

– Panagiotis Takis Metaxas, Eni Mustafaraj, and Daniel Gayo-Avello. 2011. "How (not) to predict elections". In *SocialCom 2011: The 3rd IEEE International Conference on Social Computing*, ed. by Alessandro Vinciarelli, Maja Pantic, Elisa Bertino, and Justin Zhan, 165–171. Washington, DC: IEEE. doi:10.1109/PASSAT/SocialCom.2011.98.

– Kevin Munger. 2017. "Tweetment Effects on the Tweeted: Experimentally Reducing Racist Harassment". *Political Behavior* 39 (3): 629–649. doi:10.1007/s11109-016-9373-5.

– W. Russell Neuman, Lauren Guggenheim, S. Mo Jang, and Soo Young Bae. 2014. "The Dynamics of Public Attention: Agenda-Setting Theory Meets Big Data". *Journal of Communication* 64 (2): 193–214. doi:10.1111/jcom.12088.

– Richard Rogers. 2013a. "Debanalizing Twitter: The transformation of an object of study". In *WebSci 2013: Proceedings of the 5th Annual ACM Web Science Conference*, ed. by Hugh Davis, Harry Halpin, Alex Pentland, Mark Bernstein, and Lada Adamic, 356–365. New York, NY: ACM. doi:10.1145/2464464.2464511.

– David A. Shamma, Lyndon Kennedy, and Elizabeth F. Churchill. 2011. "Peaks and persistence: Modeling the shape of microblog conversations". In *CSCW 2011: Proceedings of the ACM 2011 Conference on Computer Supported Cooperative Work*, ed. by Pamela Hinds, John C. Tang, Jian Wang, Jakob Bardram, and Nicolas Ducheneaut, 355–358. New York, NY: ACM. doi:10.1145/1958824.195887810.1145/1958824.1958878.

– Sebastian Stier, Arnim Bleier, Haiko Lietz, and Markus Strohmaier. 2018b. "Election Campaigning on Social Media: Politicians, Audiences and the Mediation of Political Communication on Facebook and Twitter". *Political Communication* 35 (1): 50–74. doi:10.1080/10584609.2017.1334728.

– Yannis Theocharis, Pablo Barberá, Zoltán Fazekas, Sebastian Adrian Popa, and Olivier Parnet. 2016. "A Bad Workman Blames His Tweets: The Consequences of Citizens' Uncivil Twitter Use When Interacting With Party Candidates". *Journal of Communication* 66 (6): 1007–1031. doi:10.1111/jcom.12259.

– Yannis Theocharis, Silia Vitoratou, and Javier Sajuria. 2017. "Civil Society in Times of Crisis: Understanding Collective Action Dynamics in Digitally-Enabled Volunteer Networks". *Journal of Computer-Mediated Communication* 22 (5): 248–265. doi:10.1111/jcc4.12194.

– Damian Trilling. 2015. "Two Different Debates? Investigating the Relationship Between a Political Debate on TV and Simultaneous Comments on Twitter". *Social Science Computer Review* 33 (3): 259–276. doi:10.1177/0894439314537886.

Public Datasets:

– Lars Kaczmirek, Philipp Mayr, Ravi Vatrapu, Arnim Bleier, Manuela Blumenberg, Tobias Gummer, Abid Hussain, Katharina Kinder-Kurlanda, Kaveh Manshaei, Mark Thamm, Katrin Weller, Alexander Wenz, and Christof Wolf. 2013. "Social Media Monitoring of the Campaigns for the 2013 German Bundestag Elections on Facebook and Twitter". *arXiv*, no. 1312.4476v2. https://arxiv.org/abs/1312.4476.

– Sebastian Stier, Arnim Bleier, Malte Bonart, Fabian Mörsheim, Mahdi Bohlouli, Margarita Nizhegorodov, Lisa Posch, Jürgen Maier, Tobias Rothmund, and Steffen Staab.

2018a. *Systematically Monitoring Social Media: The Case of the German Federal Election 2017*. 4. Köln, DE: GESIS Papers. http://nbn-resolving.de/urn:nbn:de:0168-ssoar-56149-4.

4.7 Week 7: Fronleichnam (May 31)—no meeting

This is an official holiday, so no course meeting on this day.

4.8 Week 8: Loading Twitter Data Into a Database (June 7)

Required Readings:
– Pages 29-41 from Pascal Jürgens and Andreas Jungherr. 2016. *A Tutorial for Using Twitter-Data in the Social Sciences: Data Collection, Preparation, and Analysis*. Social Science Research Network (SSRN). doi:10.2139/ssrn.2710146.

Background Readings:
– Grant Allen and Mike Owens. 2010. *The Definitive Guide to SQLite*. 2nd ed. New York, NY: Apress.

– Jay A. Kreibich. 2010. *Using SQLite: Small. Fast. Reliable. Choose Any Three.* Sebastopol, CA: O'Reilly Media.

– Charles Leifer. 2018. *pewee Documentation*. http://peewee.readthedocs.io/en/latest/.

4.9 Week 9: Sample Analyses: Counts & Time Series (June 14)

Required Readings:
– Pages 42-67 from Pascal Jürgens and Andreas Jungherr. 2016. *A Tutorial for Using Twitter-Data in the Social Sciences: Data Collection, Preparation, and Analysis*. Social Science Research Network (SSRN). doi:10.2139/ssrn.2710146.

Background Readings:
– Janet M. Box-Steffensmeier, John R. Freeman, Matthew P. Hitt, and Jon C. W. Pevehouse. 2014. *Time Series Analysis for the Social Sciences*. New York, NY: Cambridge University Press.

– Winston Chang. 2018. *R Graphics Cookbook*. 2nd ed. Sebastopol, CA: O'Reilly Media.

– Robert I. Kabacoff. 2015. *R in Action: Data Analysis and Graphics with R*. 2nd ed. Shelter Island, NY: Manning Publications Co.

– Hadley Wickham. 2016. *ggplot2: Elegant Graphics for Data Analysis*. 2nd ed. Cham, CH: Springer.

– Hadley Wickham and Garrett Grolemund. 2017. *R for Data Science*. Sebastopol, CA: O'Reilly Media. http://r4ds.had.co.nz.

4.10 Week 10: Sample Analyses: Networks (June 21)

Required Readings:
– Pages 68-79 from Pascal Jürgens and Andreas Jungherr. 2016. *A Tutorial for Using Twitter-Data in the Social Sciences: Data Collection, Preparation, and Analysis*. Social Science Research Network (SSRN). doi:10.2139/ssrn.2710146.

Background Readings:
– David Easley and Jon Kleinberg. 2010. *Networks, Crowds, and Markets: Reasoning About a Highly Connected World*. Cambridge, UK et al.: Cambridge University Press.

– Matthew O. Jackson. 2008. *Social and Economic Networks*. Princeton, NJ: Princeton University Press.

– Eric D. Kolaczyk. 2009. *Statistical Analysis of Network Data: Methods and Models*. Cham, CH: Springer. doi:10.1007/978-0-387-88146-1.

– Eric D. Kolaczyk and Gábor Csárdi. 2014. *Statistical Analysis of Network Data with R*. Cham, CH: Springer.

– Katherine Ognyanova. 2017. *Network visualization with R*. http://kateto.net/network-visualization.

4.11 Week 11: Independent Study and Preparation of Presentations (June 28)—no meeting

4.12 Week 12: Student Presentations I. (July 5)

In Week 6, students commit to a date for their presentation.

4.13 Week 13: Student Presentations II. (July 12)

4.14 Week 14: Student Presentations III. & Where to take it from here? Discussion of Open Questions and Paper (July 19)

Using other data sources:
– Ryan Mitchell. 2018. *Web Scraping with Python: Collecting More Data from the Modern Web*. 2nd ed. Sebastopol, CA: O'Reilly Media.

– Matthew A. Russell. 2018. *Mining the Social Web*. 3rd ed. Sebastopol, CA: O'Reilly Media.

– Matthew J. Salganik. 2017. *Bit By Bit: Social Research in the Digital Age*. Princeton, NJ: Princeton University Press.

Extending your analytical skill set:
– Mario Callegaro, Katja Lozar Manfreda, and Vasja Vehovar. 2015. *Web Survey Methodology*. London, UK: SAGE Publications.

– David Donoho. 2015. "50 Years of Data Science". In *Paper Presented at the Tukey Centennial Workshop*. Princeton, NJ. http://courses.csail.mit.edu/18.337/2015/

docs/50YearsDataScience.pdf.

– Bradley Efron and Trevor Hastie. 2016. *Computer Age Statistical Inference: Algorithms, Evidence, and Data Science*. Cambridge, UK: Cambridge University Press.

– Peter Flach. 2012. *Machine Learning: The Art and Science of Algorithms that Make Sense of Data*. New York, NY: Cambridge University Press.

– Alan S. Gerber and Donald P. Green. 2012. *Field Experiments: Design, Analysis, and Interpretation*. New York, NY: W. W. Norton & Company.

– Ian Goodfellow, Yoshua Bengio, and Aaron Courville. 2016. *Deep Learning*. Cambridge, MA: The MIT Press.

– Wes McKinney. 2017. *Python for Data Analysis: Data Wrangling with Pandas, NumPy, and IPython*. 2nd ed. Sebastopol, CA: O'Reilly Media.

– Diana C. Mutz. 2011. *Population-Based Survey Experiments*. Princeton, NJ: Princeton University Press.

– Sebastian Raschka and Vahid Mirjalili. 2017. *Python Machine Learning: Machine Learning and Deep Learning with Python, scikit-learn, and TensorFlow*. 2nd ed. Birmingham, UK: PACKT Publishing.

– Matthew J. Salganik. 2017. *Bit By Bit: Social Research in the Digital Age*. Princeton, NJ: Princeton University Press.

How might you employ these skills outside of academia:

– Fred Benenson. 2016. "On to the next 2,271 days..." *Medium: Hackernoon* (). https://hackernoon.com/on-to-the-next-2-271-days-309d6ba672d7.

– Thomas H. Davenport and D. J. Patil. 2012. "Data Scientist: The Sexiest Job of the 21st Century". *Harvard Business Review* (). https://hbr.org/2012/10/data-scientist-the-sexiest-job-of-the-21st-century.

– Olivia Lau and Ian Yohai. 2016. "Using Quantitative Methods in Industry". *PS: Political Science & Politics* 49 (3): 524–526. doi:10.1017/S1049096516000901.

– David W. Nickerson and Todd Rogers. 2014. "Political Campaigns and Big Data". *The Journal of Economic Perspectives* 28 (2): 51–74. doi:10.1257/jep.28.2.51.

– Andrew Therriault. 2016. "Finding a Place in Political Data Science". *PS: Political Science & Politics* 49 (3): 531–534. doi:10.1017/S1049096516000925.