



Kontakt: Raum D 307 • Universitätsstraße 10 • 78457 Konstanz • andreas.jungherr@gmail.com
Sekretariat: Birgit Jacob • Raum D 312 • Tel.: 07531-884045 • birgit.jacob@uni-konstanz.de

Using Digital Trace Data in the Social Sciences (SS 2016)

Time:

July 18, 2016 Monday—10:00 to 13:15
July 19, 2016 Tuesday—13:30 to 18:30
July 20, 2016 Wednesday—10:00 to 13:15
July 21, 2016 Thursday—08:15 to 13:15
July 22, 2016 Friday—08:15 to 11:30

Place: E E 405

Start: July 18, 2016

Office Hours: By appointment

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. The course will be offered as a Blockseminar after the end of the regular term.

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-19640-20161—Vertiefungsseminar—6 ECTS

POL-19650-20161—Seminar—7 ECTS

POL-19630-20161—Doktorandenseminar—4 ECTS

Kursseite: <http://andreasjungherr.net>

I will post information, readings, and example scripts for the sessions of this course on my website <http://andreasjungherr.net>. 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* (2016). The tutorial is freely available on the *Social Science Research Network (SSRN)* at http://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.

A good many of the recommended further readings is available through the *Proquest Safari-Books Online-Shelf* (<http://proquest.techbus.safaribooksonline.de>). Access to Safari-Books Online is freely available by using your Uni-Konstanz VPN-access.

Requirements:

1. Regular and active participation.
2. Independent data analysis:
 - Following the course, you will be asked to perform and report an independent data analysis based on data collected on Twitter 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.
 - Presentation: You will be asked to present a research question of interest to you on Thursday, July 21. For this presentation please prepare a short statement introducing your research question, your motivation, your proposed approach, and open questions. This presentation will take approximately 5 minutes and will be followed by a quick round of feedback from the other participants.
 - Paper: 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)* as given here <http://www.apsanet.org/apsrsubmissions2016>.
 - 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 (September 15) electronically at andreas.jungherr@gmail.com and by hardcopy with Birgit Jacob (Raum D 312). The date is mandatory and can only be extended in case of officially certified illness.

Syllabus

Course Overview and Session Details:

<http://andreasjungherr.net/>

Textbook:

- 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. Available at http://papers.ssrn.com/sol3/papers.cfm?abstract_id=2710146

Code Repository:

<https://github.com/trifle/twitterresearch>

Background Readings:

Using Digital Trace Data in the Social Sciences:

- David Donoho. *50 Years of Data Science*. Paper presented at the *Tukey Centennial workshop*, Princeton, NJ. Sept. 18 (2015).
- James Howison, Andrea Wiggins, and Kevin Crowston. “Validity issues in the use of social network analysis with digital trace data”. In: *Journal of the Association for Information Systems* 12.12 (2011), pp. 767–797.
- Andreas Jungherr. *Analyzing Political Communication with Digital Trace Data: The Role of Twitter Messages in Social Science Research*. Cham, CH: Springer, 2015.
- Andreas Jungherr, Harald Schoen, and Pascal Jürgens. “The mediation of politics through Twitter: An analysis of messages posted during the campaign for the German federal election 2013”. In: *Journal of Computer-Mediated Communication* 21.1 (2016), pp. 50.68. doi: 10.1111/jcc4.12143.
- Andreas Jungherr, Harald Schoen, Oliver Posegga, and Pascal Jürgens. “Digital Trace Data in the Study of Public Opinion: An Indicator of Attention Toward Politics Rather Than Political Support”. In: *Social Science Computer Review*. (2016). doi: 10.1177/0894439316631043

Python:

- Swaroop Chitlur: *A Byte of Python*. Available at <http://python.swaroopch.com>
- Nick Eubank (2015) *Data Analysis in Python*. Available at <http://www.data-analysis-in-python.org/index.html>
- Mark Lutz (2013) *Learning Python*, 5th Edition. O'Reilly Media, Inc.
- Wes McKinney (2012) *Python for Data Analysis*. O'Reilly Media, Inc.
- Al Sweigart (2015): *Automate the Boring Stuff with Python: Practical Programming for Total Beginners*. No Starch Press.

R:

- Winston Chang (2012) *R Graphics Cookbook*. O'Reilly Media, Inc.
- Richard Cotton (2013) *Learning R*. O'Reilly Media, Inc.
- Robert I. Kabacoff (2015) *R in Action*. 2nd ed. Manning Publications.

Data Collection Online:

- Matthew A. Russell (2013) *Mining the Social Web: Data Mining Facebook, Twitter, LinkedIn, Google+, GitHub, and More*, 2nd Edition. O'Reilly Media, Inc.

Sessions Overview

Session 1: Introduction and Conceptual Issues in the Use of Digital Trace Data in Social Science, Computational Social Science, Digital Methods, and Big Data

July 18, 2016 Monday—10:00 to 11:30

Session 2: Set Up and Introduction to Collecting Data on Twitter

July 18, 2016 Monday—11:45 to 13:15

Session 3: Introduction to Python

July 19, 2016 Tuesday—13:30 to 15:00

Session 4: Collecting Data Through Twitter's API

July 19, 2016 Tuesday—15:15 to 16:45

Session 5: Data Lab

July 19, 2016 Tuesday—17:00 to 18:30

Session 6: Loading Twitter Data Into a Database

July 20, 2016 Wednesday—10:00 to 11:30

Session 7: Extracting Data for Typical Analyses

July 20, 2016 Wednesday—11:45 to 13:15

Session 8: Presentation and Discussion of Students' Research Projects Pt. 1

July 21, 2016 Thursday—08:15 to 09:45

Session 9: Presentation and Discussion of Students' Research Projects Pt. 2

July 21, 2016 Thursday—10:00 to 11:30

Session 10: Presentation and Discussion of Students' Research Projects Pt. 3

July 21, 2016 Thursday—11:45 to 13:15

Session 11: Data Lab

July 22, 2016 Friday—08:15 to 09:45

Session 12: Where to take it from here? Discussion of Open Questions and Paper

July 22, 2016 Friday—10:00 to 11:30

Detailed Session Plan and Suggested Readings

Session 1: Introduction and Conceptual Issues in the Use of Digital Trace Data in Social Science, Computational Social Science, Digital Methods, and Big Data

July 18, 2016 Monday—10:00 to 11:30

Required Readings:

- 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. Available at http://papers.ssrn.com/sol3/papers.cfm?abstract_id=2710146 (pp. 7-14)

Background Readings:

- David Donoho. *50 Years of Data Science*. Paper presented at the Tukey Centennial workshop, Princeton, NJ. Sept. 18 (2015).
- Deen Freelon. "On the interpretation of digital trace data in communication and social computing research". In: *Journal of Broadcasting & Electronic Media* 58.1 (2014), pp. 59–75. doi: 10.1080/08838151.2013.875018.
- Scott A. Golder and Michael W. Macy. "Digital Footprints: Opportunities and Challenges for Online Social Research". In: *Annual Review of Sociology* 40 (2014), pp. 129–152. doi: 10.1146/annurevsoc071913043145.
- James Howison, Andrea Wiggins, and Kevin Crowston. "Validity issues in the use of social network analysis with digital trace data". In: *Journal of the Association for Information Systems* 12.12 (2011), pp. 767–797.
- Andreas Jungherr. *Analyzing Political Communication with Digital Trace Data: The Role of Twitter Messages in Social Science Research*. Cham, CH: Springer, 2015.
- Andreas Jungherr and Pascal Jürgens. "Forecasting the pulse: How deviations from regular patterns in online data can identify offline phenomena". In: *Internet Research* 23.5 (2013), pp. 589–607. doi: 10.1108/IntR-06-2012-0115.
- Andreas Jungherr, Harald Schoen, and Pascal Jürgens. "The mediation of politics through Twitter: An analysis of messages posted during the campaign for the German federal election 2013". In: *Journal of Computer-Mediated Communication* 21.1 (2016), pp. 50.68. doi: 10.1111/jcc4.12143.
- Andreas Jungherr, Harald Schoen, Oliver Posegga, and Pascal Jürgens. "Digital Trace Data in the Study of Public Opinion: An Indicator of Attention Toward Politics Rather Than Political Support". In: *Social Science Computer Review*. (2016). doi: 10.1177/0894439316631043.
- David Lazer et al. "Computational social science". In: *Science* 323.5915 (2009), pp. 721–723. doi: 10.1126/science.1167742.
- David Lazer et al. "The Parable of Google Flu: Traps in Big Data Analysis". In: *Science* 343.6176 (2014), pp. 1203–1205. doi: 10.1126/science.1248506.
- Viktor Mayer-Schönberger and Kenneth Cukier. *Big Data: A Revolution that Will Transform How We Live, Work, and Think*. New York, NY: Houghton Mifflin, 2013.
- Richard Rogers. *Digital Methods*. Cambridge, MA: The MIT Press, 2013.
- Derek Ruths and Jürgen Pfeffer. "Social media for large studies of behavior". In: *Science* 346.6213 (2014), pp. 1063–1064. doi: 10.1126/science.346.6213.1063.
- Markus Strohmaier and Claudia Wagner. "Computational Social Science for the World Wide Web". In: *IEEE Intelligent Systems* 29.5 (2014), pp. 84–88. doi: 10.1109/MIS.2014.80.

Session 2: Set Up and Introduction to Collecting Data on Twitter

July 18, 2016 Monday—11:45 to 13:15

Required Readings:

- Jürgens & Jungherr (2016) (pp. 15-20)

Background Readings:

- Nick Eubank (2015) *Data Analysis in Python*. Available at <http://www.data-analysis-in-python.org/index.html>
- Matthew A. Russell. *Mining the Social Web*. 2nd ed. Sebastopol, CA: O'Reilly Media, 2014.

Session 3: Introduction to Python

July 19, 2016 Tuesday—13:30 to 15:00

Required Readings:

- Nick Eubank (2015) *Data Analysis in Python*. Available at <http://www.data-analysis-in-python.org/index.html>

Background Readings:

- Swaroop Chitlur. *A Byte of Python*. Available at <http://python.swaroopch.com>
- Mark Lutz *Learning Python*, 5th ed. Sebastopol, CA: O'Reilly Media, 2013.
- Wes McKinney. *Python for Data Analysis*. Sebastopol, CA: O'Reilly Media, 2013.
- Sebastian Raschka. *Python Machine Learning*. 2015. PACKT Publishing.
- Al Sweigart (2015): *Automate the Boring Stuff with Python: Practical Programming for Total Beginners*. No Starch Press.

Session 4: Collecting Data Through Twitter's API

July 19, 2016 Tuesday—15:15 to 16:45

Required Readings:

- Jürgens & Jungherr (2016) (pp. 21-28)

Background Readings:

- Matthew A. Russell. *Mining the Social Web*. 2nd ed. Sebastopol, CA: O'Reilly Media, 2014.

Session 5: Data Lab

July 19, 2016 Tuesday—17:00 to 18:30

Session 6: Loading Twitter Data Into a Database

July 20, 2016 Wednesday—10:00 to 11:30

Required Readings:

- Jürgens & Jungherr (2016) (pp. 29-41)

Background Readings:

- Grant Allen and Mike Owens. *The Definitive Guide to SQLite*. 2nd. New York, NY: Apress, 2010.
- Jay A. Kreibich. *Using SQLite: Small. Fast. Reliable. Choose Any Three*. Sebastopol, CA: O'Reilly Media, 2010.
- Charles Leifer. *peewee Documentation*. 2016. Available at <http://peewee.readthedocs.io/en/latest/>

Session 7: Extracting Data for Typical Analyses

July 20, 2016 Wednesday—11:45 to 13:15

Required Readings:

- Jürgens & Jungherr (2016) (pp. 42-79)

Background Readings:

- Winston Chang (2012) *R Graphics Cookbook*. O'Reilly Media, Inc.
- Garrett Grolemund and Hadley Wickham (2016) *R for Data Science*. O'Reilly Media, Inc. Available at <http://r4ds.had.co.nz>
- Robert Kabacoff (2015) *R in Action*. 2nd ed. Manning Publications.

- Eric D. Kolaczyk and Gábor Csárdi (2014) *Statistical Analysis of Network Data with R*. Springer.
- Janet M. Box-Steffensmeier et al. (2014) *Time Series Analysis for the Social Sciences*. New York, NY: Cambridge University Press.
- Hadley Wickham (2016) *ggplot2: Elegant Graphics for Data Analysis*. 2nd ed. Springer.

Session 8: Presentation and Discussion of Students' Research Projects Pt. 1

July 21, 2016 Thursday—08:15 to 09:45

Required Readings:

- Andreas Jungherr. "Twitter Use in Election Campaigns: A Systematic Literature Review". In: *Journal of Information Technology & Politics* 13.1 (2016). doi: 10.1080/19331681.2015. 1132401.

Background Readings:

How to Find a Research Questions:

Howard S. Becker. *Tricks of the Trade: How to think about your research while you're doing it* (1998). Chicago, IL: University of Chicago Press.

Conceptual Issues in Working with Digital Trace Data:

- Fernando Diaz et al. "Online and social media data as a flawed continuous panel survey". In: *PLoS ONE* 11.1 (2016). e0145406. doi:10.1371/journal.pone.0145406.
- James Howison, Andrea Wiggins, and Kevin Crowston. "Validity issues in the use of social network analysis with digital trace data". In: *Journal of the Association for Information Systems* 12.12 (2011), pp. 767–797.
- Lilli Japiec, Frauke Kreuter, Marcus Berg, Paul Biemer, Paul Decker, Cliff Lampe, Julia Lane, Cathy O'Neil, and Abe Usher. "Big Data in Survey Research: AAPOR Task Force Report". In: *Public Opinion Quarterly* 79.4 (2015), pp. 839-880. doi: 10.1093/poq/nfv039
- Andreas Jungherr, Harald Schoen, and Pascal Jürgens. "The mediation of politics through Twitter: An analysis of messages posted during the campaign for the German federal election 2013". In: *Journal of Computer-Mediated Communication* 21.1 (2016), pp. 50.68. doi: 10.1111/jcc4.12143.
- Andreas Jungherr, Harald Schoen, Oliver Posegga, and Pascal Jürgens. "Digital Trace Data in the Study of Public Opinion: An Indicator of Attention Toward Politics Rather Than Political Support". In: *Social Science Computer Review*. (2016). doi: 10.1177/0894439316631043
- David Lazer et al. "The Parable of Google Flu: Traps in Big Data Analysis". In: *Science* 343.6176 (2014), pp. 1203–1205. doi: 10.1126/science.1248506.
- Filipe N Ribeiro, Matheus Araújo, Pollyanna Gonçalves, Marcos André Gonçalves, and Fabrício Benevenuto. "SentiBench - a benchmark comparison of state-of-the-practice sentiment analysis methods". *EPJ Data Science* 5.23 (2016) doi: 10.1140/epjds/s13688-016-0085-1
- Michael F. Schober, Josh Pasek, Lauren Guggenheim, Cliff Lampe, and Frederick G. Conrad. "Social Media Analyses for Social Measurement". *Public Opinion Quarterly* 80.1 (2016), pp. 180-211. doi: 10.1093/poq/nfv048

Case Studies Illustrating Different Approaches to the Use of Twitter data

- Pablo Barberá. "Birds of the same feather tweet together: Bayesian ideal point estimation using Twitter data". In: *Political Analysis* 23.1 (2015), pp. 76–91. doi: 10.1093/pan/mpu011.
- Marco T. Bastos, Dan Mercea, and Arthur Charpentier. "Tents, Tweets, and Events: The Interplay Between Ongoing Protests and Social Media". In: *Journal of Communication* 65.2 (2015), pp. 320–350. doi: 10.1111/jcom.12145.
- Michael D. Conover et al. "Political polarization on Twitter". In: *ICWSM 2011: Proceedings of the 5th International AAAI Conference on Weblogs and Social Media*. Ed. by Nicolas Nicolov et al. Menlo Park, CA: Association for the Advancement of Artificial Intelligence (AAAI), 2011, pp. 89–96.
- Peter Sheridan Dodds, Kameron Decker Harris, Isabel M. Kloumann, Catherine A. Bliss, Christopher M. Danforth. "Temporal Patterns of Happiness and Information in a Global Social Network: Hedonometrics and Twitter". *PLoS ONE* 6.12 (2011), e26752. doi: 10.1371/journal.pone.0026752

- Elizabeth Dubois and Devin Gaffney. “The multiple facets of Influence: Identifying political influentials and opinion leaders on Twitter”. *American Behavioural Scientist* 58.10 (2014), pp. 1260-1277. doi: 10.1177/0002764214527088
- Sharad Goel et al. “The Structural Virality of Online Diffusion”. In: *Management Science* 62.1 (2015), pp. 180-196. doi: 10.1287/mnsc.2015.2158.
- Todd Graham et al. “Between broadcasting political messages and interacting with voters: The use of Twitter during the 2010 UK general election campaign”. In: *Information, Communication & Society* 16.5 (2013), pp. 692–716. doi: 10.1080/1369118X.2013.785581.
- Mark Edward Huberty. “Can we vote with our tweet? On the perennial difficulty of election forecasting with social media”. In: *International Journal of Forecasting* 31.3 (2015), pp. 992–1007. doi: 10.1016/j.ijforecast.2014.08.005.
- Andreas Jungherr. “The logic of political coverage on Twitter: Temporal dynamics and content”. In: *Journal of Communication* 64.2 (2014), pp. 239–259. doi: 10.1111/jcom.12087.
- Andreas Jungherr. *Analyzing Political Communication with Digital Trace Data: The Role of Twitter Messages in Social Science Research* (2015). Cham: Springer. doi: 10.1007/978-3-319-20319-5
- Andreas Jungherr and Pascal Jürgens. “Through a glass, darkly: Tactical support and symbolic association in Twitter messages commenting on Stuttgart 21”. In: *Social Science Computer Review* 32.1 (2014), pp. 74–89. doi: 10.1177/0894439313500022
- Pascal Jürgens, Andreas Jungherr, and Harald Schoen. “Small Worlds with a Difference: New Gatekeepers and the Filtering of Political Information on Twitter.” In *WebSci '11: Proceedings of the 3rd International Web Science Conference*. New York, NY: ACM, 2011. doi: 10.1145/2527031.2527034
- Daniel Kreiss. “Seizing the Moment: The Presidential Campaigns’ Use of Twitter During the 2012 Electoral Cycle”. *New Media & Society* (2014). (Online First). doi: 10.1177/1461444814562445
- Yu-Ru Lin et al. “Rising tides or rising stars? Dynamics of shared attention on Twitter during media events”. In: *PLoS One* 9.5 (2014), e94093. doi: 10.1371/journal.pone.0094093.
- Panagiotis Takis Metaxas, Eni Mustafaraj, and Daniel Gayo Avello. “How (not) to predict elections”. In: *SocialCom 2011: The 3rd IEEE International Conference on Social Computing*. Washington, DC: IEEE, 2011, pp. 165–171. doi: 10.1109/PASSAT/SocialCom.2011.98.
- W. Russell Neuman, Lauren Guggenheim, S. Mo Jang, and Soo Young Bae. “The Dynamics of Public Attention: Agenda-Setting Theory Meets Big Data”. *Journal of Communication* 64.2 (2014), pp. 193–214. doi: 10.1111/jcom.12088
- Richard Rogers. “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 et al. New York, NY: ACM, 2013, pp. 356–365. doi: 10.1145/2464464.2464511.
- David A. Shamma, Lyndon Kennedy, and Elizabeth F. Churchill. “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 et al. New York, NY: ACM, 2011, pp. 355–358. doi: 10.1145/1958824.195887810.1145/1958824.1958878.
- Damian Trilling. “Two Different Debates? Investigating the Relationship Between a Political Debate on TV and Simultaneous Comments on Twitter”. In: *Social Science Computer Review* 33.3 (2015), pp. 259–276. doi: 10.1177/0894439314537886.

Session 9: Presentation and Discussion of Students’ Research Projects Pt. 2

July 21, 2016 Thursday—10:00 to 11:30

Session 10: Presentation and Discussion of Students’ Research Projects Pt. 3

July 21, 2016 Thursday—11:45 to 13:15

Session 11: Data Lab

July 22, 2016 Friday—08:15 to 09:45

Session 12: Where to take it from here? Discussion of Open Questions and Paper

July 22, 2016 Friday—10:00 to 11:30

Background Readings:

Using other data sources:

- Matthew A. Russell. *Mining the Social Web*. 2nd ed. Sebastopol, CA: O'Reilly Media, 2014.

Extending your analytical skill set:

- Mario Callegaro, Katja Lozar Manfreda, and Vasja Vehovar. *Web Survey Methodology*. 2015. SAGE.
- David Donoho. *50 Years of Data Science*. Paper presented at the Tukey Centennial workshop, Princeton, NJ. Sept. 18, 2015.
- Peter Flach. *Machine Learning: The Art and Science of Algorithms that Make Sense of Data*. 2012. Cambridge, UK: Cambridge University Press.
- Alan S. Gerber and Donald P. Green. *Field Experiments: Design, Analysis, and Interpretation*. 2012. W.W. Norton.
- Wes McKinney. *Python for Data Analysis*. Sebastopol, CA: O'Reilly Media. 2013.
- Sebastian Raschka. *Python Machine Learning*. 2015. PACKT Publishing.

How might you employ these skills outside of academia:

- Fred Benenson. *On to the next 2,271 days...* January 12, 2016.
- Olivia Lau and Ian Yohai. "Using Quantitative Methods in Industry". In: *PS: Political Science & Politics* 49.3. 2016. pp. 524-526. doi: 10.1017/S1049096516000901
- Andrew Therriault. "Finding a Place in Political Data Science". In: *PS: Political Science & Politics* 49.3. 2016. pp. 531-533. doi: 10.1017/S1049096516000925