09-60-M8/9-C: Introduction to Computational Communication

Prof. Yannis Theocharis

Fall Semester, 2018

E-mail: yannis.theocharis@uni-bremen.de
Web: Stud.IP page
Office Hours:
Class Hours:
Office:
Class Room:

Course Description

The rapid digitalization and spread of digital communication is changing our social, political and economic world in unprecedented ways. Digital footprints left behind by humans using digital tools have the potential to provide new angles for understanding social phenomena. Digitally-enabled communication opens up completely new avenues for social and political interaction that have radical effects on political information environments, and the democratic attitudes and behaviours they shape. Computational methods for analyzing communication structures and processes are at the epicentre of these developments. This course is aimed at providing students with a basic understanding of computational methods in communication research, and will train them to deploy them for a variety of purposes and research projects. The core part of the course involves learning about how to access new types of data and developing skills for analyzing them in connection with core theoretical questions in communication research. The course structure is as follows: after being introduced to some basic ideas, opportunities, and challenges emerging from the rise of computational social science and big data, students will receive training in computational tools, and more specifically in the statistical programming platform R and the visualization software Gephi. Students will learn to use R to conduct basic text analysis and build dictionaries, as well as to access, download, analyze and visualize social media data. Students will get the opportunity to collaborate and build group projects on topics of their choice using data they have collected themselves. This learning structure aims to provide a foundation for building skills in computational methods that can be applied above and beyond the context of this course.

Intended learning objectives

By the end of this course students will be able to do the following:

- Identify the major challenges and opportunities for communication research emerging from the proliferation of digital trace data
- Use basic coding skills for a variety of research purposes, such as building dictionaries and conducting basic text analysis
- Utilize R, a software for statistical computing, for a range of tasks, including the collection, management, analysis and visualization of text- and network-based social media data
- Apply text- and network-analytic methods to research questions in media and communication
- Work collaboratively with other students to build research projects that apply computational methods to the study of communication processes

Core readings

Computational social science – and computational communication in particular – is still very much a field under development. While there are various excellent volumes and even more excellent individual scientific articles on this or that aspect of this course, no one source brings together the diverse theoretical and technical aspects required for a solid introduction to computational communication. As such, there is no standardized textbook for this course. We will rely mostly on selected materials from different collections (see weekly schedule and literature), the course slides, as well as scripts for the code. However, segments from the following books, which will be available at the **Computational Communication & Democracy Lab** library and the central University of Bremen library (some are also freely available online), will be used. This material will be of much help to those wishing to delve deeper into both theoretical and technical questions, and expand their knowledge and skills in both the computational tools/coding and the theoretical foundations upon which this course is built.

Theoretical background

- Salganik, M. (2017). *Bit by Bit: Social Research in the Digital Age*. New Jersey: Princeton University Press.
- Jungherr, A. (2015). *Analyzing Political Communication with Digital Trace Data: The Role of Twitter Messages in Social Science Research*. New York: Springer.
- González-Bailón, S.(2018) *Decoding the Social World: Data Science and the Unintended Consequences of Communication*. Cambridge: MIT Press.
- Ackland, R. (2013). *Web Social Science: Concepts, Data and Tools for Social Scientists in the Digital Age*. New York: Sage.
- Stroud, N.J. & McGregor, S. (2018) *Digital Discussions: How Big Data Informs Political Communication*. New York: Routledge.

- Sloan, L. & Quan-Haase, A. (2017). *The SAGE Handbook of Social Media Research Methods*. Sage.

Technical background

- Grolemund, G. & Wickham, H. (2017). *R for Data Science: Import, Tidy, Transform, Visualize, and Model Data*. O'Reilly Media. (also available online)
- Teetor, P. (2011). *R Cookbook: Proven Recipes for Data Analysis, Statistics, and Graphics*. O'Reilly Media.
- Russell, M.A. (2018) *Mining the Social Web*. O'Reilly Media (3rd Edition).
- Steinert-Threlkeld, Z. (2018). *Twitter as Data (Elements in Quantitative and Computational Methods for the Social Sciences)*. Cambridge: Cambridge University Press.

Other sources (or "the truth is out there – but you have to search for it")

If this is your first time coding, or the first time you find yourself working with a command-line interface, you will soon figure out that there are moments that you feel trapped because, while it seems you have typed everything correctly, your code doesn't work. There might also be moments where you are unsure how to proceed and the code the instructor provided in the class does not include this extra step you need to make a more elegant estimation. In this case do not worry. We've all been there. The good thing is that "we" is actually a *huge* number of people. As such, it is highly unlikely that someone didn't have the problem you are facing before you. Crucially, there is usually more than one way to reach your aim when programming. One of the most distinctively helpful things about R is the excellent documentation, online help and vast resources. Most of the times it is sufficient to just google your problem. However, for more focused discussions (and probably quicker solutions too) the following will come in handy:

- StackExchange
- R Bloggers
- GitHub
- Documentation that comes with the package you're using!

Prerequisites/Parallel courses

This is a course for 2nd and/or 3rd year BA students. There are no prerequisites for attending. However, this more "hands-on" introduction to computational communication is particularly well suited for students choosing the more theory-oriented course "Political Communication, Social Media and Democracy" as it provides a theoretical framework that can inspire your final projects. For this reason it is recommended that the two courses are taken in parallel, if possible.

Hardware and software requirements

- Students will have to bring a laptop to the classroom for sessions 4 to 12. Students who are not able to do this must get in touch with the instructor well before the beginning of the course.
- For the purposes of this course, we will use R, a free, open-source software for statistical programming. With more than 10.000 packages extending its capabilities, R is a powerful tool for data collection, analysis and visualization of data. Students need to install R and R Studio (in this order) to their laptops. While students are encouraged to install the software themselves, instructions on how to install both R and R Studio will be given in the class.
- Students also need to install Gephi and make sure that they have got it working by session 12. Gephi is a free software for the (basic) analysis and visualization of network data. Normally you will have no problem to get it to work, but if it doesn't work on your computer the usual reason is your Java version. Please upgrade to the latest version.

Course Structure

Class Structure

The course is divided into two complementary sections. The first section "Conceptualization and Design" is aimed at providing students with a theoretical and conceptual background of the major debates around terms such as "big data" and "computational social science". It will discuss the associated challenges and opportunities that come along with them for researchers of media and communication, use examples of data and studies using computational approaches, and discuss various possible research designs. The second section, "Tools and Methods", will introduce students to computational tools used to analyze, explore and visualize communication structures and processes. Students will be learn to use R, and get a basic understanding of data types, structures, and functions and how these can be used for basic quantitative text analysis. Students will also learn how to collect, analyze and visualize Twitter data using social network analysis.

Sessions 1 - 3 involve regular presentations by the instructor, as well as discussions around the core topics. The rest of the sessions are labs. The course philosophy follows a "learning by doing" approach, so both the course leader and the students will work in parallel in R. Towards the end of the course, on Week 13, students will deliver a brief, 5-6 minute presentation with their main project idea. The presentation should be around 2-3 slides and should contain the following:

- Main motivation for the project
- Core literature and theoretical framework
- Research question
- Data collection and method of analysis
- Possible contribution

Schedule

—Conceptualization and Design—

Week 01 (16/10/2018) Introduction to the Course

- Lazer, D. et al. (2009). Life in the network: the coming age of computational social science. *Science*, 323(5915): 721–723.
- King, G. (2016). Preface. In Alvarez, M.R. (Ed.) *Computational Social Science: Discovery and Prediction*. Cambridge: Cambridge University Press.

Week 02 (23/10/2018) Big Data in Communication Research: Challenges and Opportunities

- Jungherr, A. (2018). Normalizing Digital Trace Data. In Stroud, N.J. & McGregor, S. (Eds.) *Digital Discussions: How Big Data Informs Political Communication*. New York: Routledge.
- **Optional reading:** Salganik, M. (2018). Observing behavior. In Salganik, M. *Bit by Bit: Social Research in the Digital Age*. New Jersey: Princeton University Press (pp. 13-39).
- **Optional reading:** González-Bailón, S.(2018). Introduction: Decoding the Social World. In González-Bailón, S., *Decoding the Social World: Data Science and the Unintended Consequences of Communication*. Cambridge: MIT Press (pp. 1-18).

Week 03 (30/10/2018) Research Design in Computational Communication: Finding a Research Question

- Stroud, N.J. & McGregor, S. (2018). Big Data in Political Communication. In Stroud, N.J. & McGregor, S. (Eds.) *Digital Discussions: How Big Data Informs Political Communication*. New York: Routledge.
- Habel, P. & Theocharis, Y. (2018, forthcoming). Citizens, Elites, and Social Media: Methodological Challenges and Opportunities in the Study of Persuasion and Mobilization. In Grofman, B., Suhay, E. & Trechsel, A. (Eds.) *Oxford Handbook of Electoral Persuasion*. Oxford: Oxford University Press.
- **Optional reading:** McCay-Peet, L. & Quan-Haase, A. (2017). What is social media and what questions can social media research help us answer? In Sloan, L. & Quan-Haase, A. *The SAGE Handbook of Social Media Research Methods*. Sage.

—Computational Tools and Methods—

Week 04 (06/11/2018) Introduction to R

- Leeper, T. (2016). Really Introductory Introduction to R. or/and,
- Teetor, P. (2011). Getting Started and Getting Help. In Teetor, P. *R Cookbook: Proven Recipes for Data Analysis, Statistics, and Graphics*. O'Reilly Media.
- **Optional reading:** Teetor, P. (2011). Navigating the Software. In Teetor, P. *R Cookbook: Proven Recipes for Data Analysis, Statistics, and Graphics*. O'Reilly Media.

Week 05 (13/11/2018) Working with R: Data types and Data Structures

- Kabacoff, R.I. (2011). Creating a dataset. In Kabacoff, R.I. *R in Action: Data Analysis and Graphics with R*. Shelter Island: Manning.
- **Optional reading:** Teetor, P. (2011). Some Basics. In Teetor, P. *R Cookbook: Proven Recipes for Data Analysis, Statistics, and Graphics*. O'Reilly Media.
- **Optional reading:** Teetor, P. (2011). Data structures. In Teetor, P. *R Cookbook: Proven Recipes for Data Analysis, Statistics, and Graphics*. O'Reilly Media.

Week 06 (20/11/2018) Introduction to Quantitative Text Analysis with Quanteda

- Welbers, K., van Atteveldt, W. & Benoit, K. (2017). Text Analysis in R. *Communication Methods and Measures*, 11 (4): 245-265
- **Additional sources for building skills on Quanteda:** Watanabe, K. & Müller, S. Quanteda Tutorials.

Week 07 (27/11/2018) News Articles as Data Sources: Dictionary Methods

- Grimmer, J. & Stewart, B.M. (2013). Text as Data: The Promise and Pitfalls of Automatic Content Analysis Methods for Political Texts. *Political Analysis*, 21 (3): 267-297.
- **Additional sources for applying dictionary methods:** Watanabe, K. & Müller, S. Quanteda Tutorials: Targeted Dictionary Analysis.

Week 08 (04/12/2018) Understanding APIs and Using Twitter as a Source of Data

- Steinert-Threlkeld, Z. (2018). *Twitter as Data (Elements in Quantitative and Computational Methods for the Social Sciences)*. Cambridge: Cambridge University Press (pp. 1 - 29).

Week 09 (11/12/2018) Collecting data from Twitter's streaming API

- For documentation using the `rtweet` package see here.
- For documentation using the `tweetscores` package see here.

Week 10 (18/12/2018) Collecting data from Twitter's REST API

- For documentation using the `rtweet` package see here.
- For documentation using the `tweetscores` package see here.

Week 11 (08/01/2019) Basic Ideas in Social Network Analysis: Centrality and Community

- Prell, C. (2011). Becoming Familiar with Social Networks. In Prell, C. *Social Network Analysis: History, Theory and Methodology*. Sage.
- **Optional reading:** Kadushin, C. (2012). Introduction. In Kadushin, C. *Understanding Social Networks: Theories, Concepts and Findings*. Oxford: Oxford University Press.
- **Optional reading:** Watts, D. (2007). A twenty-first century science. *Nature*, 445 (1).

Week 12 (15/01/2019) Social Network Analysis with Twitter Data

- Ackland, R. (2013). Social Media Networks. In Ackland, R. *Web Social Science Concepts, Data and Tools for Social Scientists in the Digital Age*. Sage.

Week 13 (22/01/2019) Group Project Presentations

Week 14 (29/01/2019) Q & A and Course Feedback

Assessment and grading

The final grade for this course will be determined by the following assignments:

- Reading quiz: **10%** (session 2, in class)
- R challenge 1: **15%** (in the form of homework after session 5)
- R challenge 2: **15%** (in the form of homework after session 8)
- Group project: **60%**

Reading quiz (**10%**)

The Reading Quiz will take place at the beginning of session 2. This Quiz is only meant to be a basic check up that the students have understood the main ideas behind the challenges and opportunities of Big Data research (a decisively important thing for grasping the material to follow in this course). The seminar leader will provide students with the quiz in paper format in the class. The quiz will include two questions, and students will only have to briefly respond to one of them.

R challenges (**15% + 15%**)

The R Challenges are small assignments in which the students will have to demonstrate that they are cultivating their R coding skills. Each assignment counts 15% towards the final grade. Students will upload their assignment in the form of R script online, into the depository created on the course page at Stud.IP.

The first challenge will be focused on data types and structures. It will be given to the students at the end of session 5. The **deadline** for uploading the R script to Stud.IP is one week later, i.e. Tuesday 20 November, till 23:59.

The second challenge will focus on data collection and descriptive analysis of Twitter data. It will be given to the students at the end of session 8. The **deadline** for uploading the R script to Stud.IP is one week later, i.e. Tuesday 11 December, till 23:59.

A successful assignment is one in which the code provided to the seminar leader produces the requested answer to the challenge. Please note that there are different ways to correctly respond to the challenges, so don't be afraid to experiment!

Group project (**60%**)

The final project, which counts 60% towards the final grade, is a larger assignment whose purpose is to assess the students' capacity to *synthesize* what they have learned in the course (theory *and* practice) into a solid project, as well as to collaborate with each other. As such the final project is a group project, whereby each group will consist of **three** students, making a total of 10 groups. Each group is free to select the topic of their choice, and explore it through the lens of the communication theory they find most suitable for its investigation. The groups will collect and analyze data appropriate for answering their RQ, and write a report of 12 pages plus references. The project should contain:

- Introduction + literature review (2 pages)
- Theoretical framework and research question (2-3 pages)
- Data collection and methodology (2 pages)
- Analysis and discussion of results (4 pages)
- Conclusion (1-2 pages)

- References (7 or more)
- Appendix I (includes extra tables and figures)
- Appendix II (includes the R code used to collect and analyze the data)

The project will be evaluated using the following criteria:

- Structure and organization of the project
- Appropriate theoretical framework and soundness of research question/puzzle
- Choice of method of data collection and analysis
- Sound argument and analysis, including type of inferences made and material chosen for answering the RQ convincingly
- Adequacy and range of bibliographic information

The **deadline** for submitting the project is three weeks after the end of the semester, i.e. Friday 22 February 2019.

Grading rules

To pass the course, students need to get an average of at least 4.0 across all the assignments. Students who fail to attend the reading quiz and who will also not hand in any of the assignments, cannot participate further in the group projects and will thus fail the course. Students who are absent during the reading quiz for medical reasons can re-schedule the quiz with the instructor.

Unless an extended deadline has been authorized by the instructor, late submission of an assignment will receive a penalty. Penalties for handing in the assignment late are 1 grade per business day (i.e. a grade of 2 becomes a 3, a 3 becomes a 4, etc.).

Course Policies

Office hours

To arrange a meeting please email Ms. Fadil-Kerstein at fadil@uni-bremen.de. Office hours are on Tuesdays, between 10 and 11:30. I am happy to see anyone who comes to my office hours without an appointment, but priority will be given to those who have booked a slot through Ms. Fadil-Kerstein.

Group work

For your final project you will have to pair with fellow students. Group work is an essential part of the course, and for good reasons. Venturing into the unexplored field of computational communication with your fellow students will not only help you learn how to communicate and collaborate with others, but it will be essential for getting help when you are stuck, and gaining a better understanding through collectively thinking through problems that may appear (be it in relation to the theory or in relation to the coding aspect of the course). To this end, once you settle on a topic organize meetings with your fellow team-member well in advance. Setting deadlines and meeting outside the class will ensure working more efficiently. Try to distribute the workload equally and according to each other's strong points. If you have questions or are unsure about how to proceed ask me in the classroom (others probably have similar questions too) or come to my office hours. Remember: questions related to R are almost always likely to have already found an answer online!

Laptop/smartphone/tablet or other electronic device policy

Use of laptops is essential for this course in order to conduct the data collection, analysis and visualization in R. Please be respectful of your instructor and fellow students by using your computers *only for class-related purposes*. Please also make certain that you have put your phone away before class starts and not take it out during class. Laptops, smartphones, tablets, and other electronic devices *may not* otherwise be used during class without the permission of the instructor. Please understand that this policy is applied for educational purposes. There is growing scientific research which shows that permitting laptops and tablets in the classroom has clear negative effects on academic performance, while laptops during class hinder learning also for those who are around those who use them. Multitasking in general is highly distracting and will also have a negative effect on your performance. Exceptions will be made for students with disabilities who need to use a laptop.

Academic integrity

Students are responsible for understanding the academic integrity rules at the University of Bremen, including how to cite sources appropriately. Ignorance of the Academic Honor Principle or appropriate citation practices will not be considered an excuse if a violation occurs. Any student who is found to have cheated or plagiarized on any assignment will receive a failing grade in the class.