RUTGERS
School of Communication and Information

Computational Social Science
Doctoral Seminar
16:194:671
Spring 2021

Instructor:	Dr. Katherine Ognyanova
Email:	katya.ognyanova@rutgers.edu
Online Office Hours:	Friday 4:00pm-6:00pm and by appointment
	• Office hours Zoom link
	• Phone call: (609) 759-0896
Course Time:	Monday, 11:50am-2:30pm
Course Zoom:	Class Zoom link
Course Website:	canvas.rutgers.edu

Remote learning

In Spring 2021, Rutgers University will continue to conduct most of its classes online to avoid exposing students, faculty, and staff to health risks. This format could be challenging for people who may be experiencing technological or other issues that make it difficult to participate remotely. Health problems, stress, and economic uncertainty can also take their toll on all of us. We are all in this together, so feel free to request all the accommodations you need. If you encounter any problems with this course, reach out to me as soon as you can, and I will work with you to solve those issues.

Course Description

In a digital age, researchers have access to large-scale data describing complex social systems. This seminar offers an introduction to tools and techniques used to obtain and analyze such data. The course examines computational social science methods and their applications to the study of communication, information, and human behavior. Theoretical discussion is combined with hands-on training using the R programming language.

Learning Objectives

At the end of the course, students will be able to:

- Critically evaluate computational social science research
- Design a study using computational methods to examine human behavior
- Collect, process, and analyze online data (including text and network data)
- Understand and apply ethical digital research practices

Contact Information

My office hours will be held online using Zoom on Friday, 4pm-6pm. You can also call me over the phone at (609) 759-0896. If I do not answer the Zoom or phone call right away, I am likely having a conversation with another student. Please call back in 15 minutes!

If need be, let me know and we can also schedule an appointment outside of office hours.

The most reliable way to reach me during the week is over e-mail. Send any questions, ideas, or concerns related to the class to katya.ognyanova@rutgers.edu. Please include "SCI 671" in the e-mail subject – that ensures your message will be tagged as high-priority mail and will receive prompt attention. I will typically respond within a couple of days of receiving your email.

Required Readings

Readings will be available on the course's Canvas website (canvas.rutgers.edu). Log in using your Rutgers NetID, navigate to the course site, and browse the *Course Readings* page. The readings for each week of our class are also listed in the *Course Outline* section of this syllabus. Recommended books for the course include:

- Salganik, M. J. (2017). *Bit by Bit: Social Research in the Digital Age*. Princeton University Press. Available to read online or purchase on Amazon.
- Wickham, H., & Grolemund, G. (2017). *R for Data Science*. O'Reilly Media. Available to read online or purchase on Amazon.

Required software

The hands-on portion of this class will use the R language for statistical computing. R is widely used in both academic and corporate settings. Open-source and versatile, it has a large user community and a thriving ecosystem of packages that expand its functionality.

All software and services listed here are free or have a free version for non-commercial use. For your work in the course, you will need to do the following:

- Install **R**: A powerful platform for data analysis. (cran.r-project.org)
- Install **RStudio**: a user-friendly interface for R (www.rstudio.com)
- Register a free account at RStudio Cloud, an online environment that allows users to work on collaborative R projects from a browser
- Register a free account at GitHub, a collaborative platform that offers code management and version control

Course Attendance

You are expected to attend all classes. If you are unable to attend classes for longer than one week, you should contact a dean of students who can help verify your circumstances. University policy excuses absences due to religious observance or participation and permits students to make up work missed for that reason. Please notify me at least two weeks in advance if you are unable to come to class or take an exam due to religious observance.

Course Requirements and Evaluation

Participation (200 points)

In this course, we will work together to review and examine critical questions and themes related to computational social science. Much of the class will be discussion-based, with limited lecturing on key points. Everyone is expected to be active in contributing to the conversations we will have in class. You should read all the required materials carefully and thoroughly, reflect critically on their strengths and weaknesses, identify their key points, and be prepared to discuss them.

Lab assignments (300 points)

The course requirements include completing a total of six lab assignments worth 50 points each. Each assignment will ask you to write R code to complete a set of tasks similar to those covered in that week's class. In some cases, you will also have to interpret the results obtained by running the code and write a summary of your findings. Each lab will be due by the end of the week when it was assigned. Late labs can be submitted by May 10, but will have 10 points deducted.

Research outline (100 points)

The first step towards completing your final project for this class will be to submit an outline describing the design of your study. The outline should be 3-5 pages long, excluding references. It should include: (1) a brief theoretical setup grounding your research in existing literature; (2) research questions or hypotheses; and (3) a proposed methodology. The outline should be submitted through Canvas by **March 20**. It is a good idea to stop by during office hours and discuss your research plan with me before that date.

Final project (400 points)

For your final project, you should design and conduct a study using the computational methods discussed in this class. Given the limited time of the course, your work should not incorporate human subject research (e.g., surveys, experiments) that would require an IRB approval.

You should write up your research in a 20-25 page paper formatted in APA style. I recommend that you use the free tool Zotero (www.zotero.com) to manage and format your citations and bibliography. You can submit your work as a Microsoft Word or a PDF file through Canvas.

Your paper should be clearly written and logically organized, with a robust theoretical grounding, relevant data, using an appropriate methodological approach. You should include a discussion of your study's ethical implications, its limitations, and suggested directions for future research.

The following elements of the project will be due during the semester:

- **March 20**   Research outline due
- **April 12**   Data collection due (submit description via Canvas)
- **May 3**   Data analysis due (submit description via Canvas, add code on GitHub)
- **May 10**   Project presentation due (10-15 minutes in class)
- **May 12**   Final project due (submit paper via Canvas, add code on GitHub)

You can work on your project individually or in groups of up to 3 students. Group projects will be held to a higher standard and all members should be fully involved in all stages of the work.

Grade Breakdown & Scale

A and B grades in this course are reserved for outstanding work. To get a high grade, students need to participate actively in class, be thorough and careful in assignments, and demonstrate excellent understanding of the subject, research skills, critical thinking, and originality in their work. The grade breakdown is as follows:

Participation	200 points
Lab assignments	300 points
Research outline	100 points
Final project	400 points
Total:	1000 points

The final grade will be awarded according to the following scale:

A	900-1000 points
B+	850-899 points
B	800-849 points
C+	750-799 points
C	700-749 points
D	600-699 points
F	Below 600 points

Grade appeals

You can appeal individual assignment grades in writing up to 7 days after the grades are announced. In order to be reviewed, your appeal has to be submitted over e-mail. Once the course grades are announced, they are final and will only be changed in case of an error in the computation of the student's score.

Academic Integrity

You are required to complete your own assignments and always acknowledge the sources of contributions, materials, quotes, and ideas that you did not develop yourself. The consequences of scholastic dishonesty in this class and at Rutgers University in general are very serious. For more details, consult the University's academic integrity policy. Any violation will at a minimum result in no credit earned for the assignment in question. Serious violations of academic integrity may prevent students from completing the course or their academic program. If you have questions about issues related to plagiarism or academic integrity, do not hesitate to contact me.

Accommodation

This course will accommodate any student in need of assistance. Students with documented disabilities who need accommodations should contact the Rutgers Disabilities Services Office (see disabilityservices.rutgers.edu for details). You can also speak with a SC&I adviser by visiting the Office of Student Services in the SC&I Building, Room 214 or calling them at 848-932-7500 (dial 2 as your menu choice). Please contact me with information about the requested assistance and present your Letter of Accommodation as early in the semester as possible.

Additional Resources

The university offers a number of resources that you can access if needed:

- If you need a consultation on **research materials** and ways to find them, you can contact the Rutgers University subject specialist librarian for communication.

- The SC&I IT Services can help you with various **technological problems**. You can find them in CI 120, by phone at 848-932-5555, or by email at help@comminfo.rutgers.edu.

- If you encounter a **problem with Canvas**, you can contact the Rutgers Canvas help desk at help@canvas.rutgers.edu or call them at 877-361-1134.

- Student wellness services are available to you at Rutgers. You can contact CAPS for **mental health support** at rhscaps.rutgers.edu or by phone at 848 932 7884.

- The Office for Violence Prevention and Victim Assistance provides confidential crisis intervention, counseling, and advocacy for victims of **sexual and relationship violence**. You can reach VPVA at vpva.rutgers.edu and 848 932 1181.

- The Office of Disability Services can be reached for help with accommodation and facilities for **students with disabilities** at ods.rutgers.edu, or by phone at 848 445 6800.

- On occasion, the university may have to **cancel classes** due to inclement weather. To check if classes are canceled, visit campusstatus.rutgers.edu or call 732-932-7799.

Course summary

Week	Dates	Topic	Assignment
1	**Jan 25 (Mon)**	Introduction, course overview	
2	**Feb 1 (Mon)**	What is computational social science?	
3	**Feb 8 (Mon)**	Surveys in a digital age	
4	**Feb 15 (Mon)**	Digital trace data I	**Lab 1 due**
5	**Feb 22 (Mon)**	Digital trace data II	**Lab 2 due**
6	**Mar 1 (Mon)**	Networks I: Network basics	
7	**Mar 8 (Mon)**	Networks II: Tie formation	
	Mar 20 (Sat)	*Spring break*	**Research outline due**
8	**Mar 22 (Mon)**	Networks III: Groups and graphs	
9	**Mar 29 (Mon)**	Networks IV: Structure and behavior	**Lab 3 due**
10	**Apr 5 (Mon)**	Data visualization	**Lab 4 due**
11	**Apr 12 (Mon)**	Text analysis I	**Data collection due**
12	**Apr 19 (Mon)**	Text analysis II	**Lab 5 due**
13	**Apr 26 (Mon)**	Online experiments	**Lab 6 due**
14	**May 3 (Mon)**	Ethics and open science	**Data analysis due**
15	**May 10 (Mon)**	Final presentations	**Final presentation due**
	May 12 (Wed)		**Final project due**

′

Course Outline

The course schedule is subject to change: materials may be added or replaced during the semester. If that happens, the changes will be reflected on Canvas and announced in class/over email.

_____

Week 1 – January 25

Introductions, course overview

Read the course syllabus.

_____

Week 2 – February 1

What is computational social science?

Lazer, D. M. J., Pentland, A., Watts, D. J., Aral, S., Athey, S., Contractor et al. (2020). Computational social science: Obstacles and opportunities. *Science*, *369*(6507), 1060–1062.

van Atteveldt, W., & Peng, T.-Q. (2018). When Communication Meets Computation: Opportunities, Challenges, and Pitfalls in Computational Communication Science. *Communication Methods and Measures*, *12*(2–3), 81–92.

Salganik, M. J. (2017). *Bit by Bit: Social Research in the Digital Age*. Princeton, NJ: Princeton University Press. Ch.2 Observing behavior

Adler, J. (2010). *R in a Nutshell*. O'Reilly Media, Inc. Ch.3,5,6,7

_____

Week 3 – February 8

Surveys in a digital age

Salganik, M. J. (2017). *Bit by Bit: Social Research in the Digital Age*. Princeton, NJ: Princeton University Press. Ch.3 Asking questions

Couper, M. P. (2017). New Developments in Survey Data Collection. *Annual Review of Sociology*, *43*(1), 121–145.

Stier, S., Breuer, J., Siegers, P., & Thorson, K. (2020). Integrating Survey Data and Digital Trace Data: Key Issues in Developing an Emerging Field. *Social Science Computer Review*, *38*(5), 503–516.

Schneider, D., & Harknett, K. (2019). What's to Like? Facebook as a Tool for Survey Data Collection. *Sociological Methods & Research*, 1–33.

Survey Data Analysis with R: https://stats.idre.ucla.edu/r/seminars/survey-data-analysis-with-r

_____

_____

Week 4 – February 15
Digital trace data I

Barbera, P., & Steinert-Threlkeld, Z. C. (2020). How to Use Social Media Data for Political Science Research. In L. Curini & R. Franzese, T*he SAGE Handbook of Research Methods in Political Science and International Relations* (pp. 404–423). New York, NY: SAGE

Hampton, K. N. (2017). Studying the Digital: Directions and Challenges for Digital Methods. *Annual Review of Sociology*, *43*(1), 167–188.

Tufekci, Z. (2014). Big Questions for Social Media Big Data: Representativeness, Validity and Other Methodological Pitfalls. *Proceedings of the International AAAI Conference on Web and Social Media*, *8*(1).

Hargittai, E. (2020). Potential Biases in Big Data: Omitted Voices on Social Media. *Social Science Computer Review*, *38*(1), 10–24.

Wojcik, S., & Hughes, A. (2019). *Sizing Up Twitter Users* (p. 23). Washington, DC: Pew Research Center.

_____

Week 5 – February 22
Digital trace data II

Lazer, D., & Radford, J. (2017). Data ex Machina: Introduction to Big Data. *Annual Review of Sociology*, *43*(1), 19–39.

Freelon, D. (2018). Computational Research in the Post-API Age. *Political Communication*, 1–4.

Vraga, E. K., Bode, L., Smithson, A.-B., & Troller-Renfree, S. (2019). Accidentally Attentive: Comparing visual, close-ended, and open-ended measures of attention on social media. *Computers in Human Behavior*, *99*, 235–244.

Weber, M. S., & Napoli, P. M. (2018). Journalism History, Web Archives, and New Methods for Understanding the Evolution of Digital Journalism. *Digital Journalism*, *6*(9), 1186–1205.

Bradley, A., & James, R. J. E. (2019). Web Scraping Using R. *Advances in Methods and Practices in Psychological Science*, *2*(3), 264–270.

_____

_____

Week 6 – March 1

Networks I: Network basics

Green, H. D., & Wasserman, S. (2015). Network Analysis: A Definitional Guide to Important Concepts. *Network Analysis*, 35.

Himelboim, I. (2017). Social Network Analysis (Social Media). In The International Encyclopedia of Communication Research Methods (pp. 1–15). American Cancer Society. doi: 10.1002/9781118901731.iecrm0236

Miller, P. R., Bobkowski, P. S., Maliniak, D., & Rapoport, R. B. (2015). Talking Politics on Facebook: Network Centrality and Political Discussion Practices in Social Media. *Political Research Quarterly*, *68*(2), 377–391.

Patty, J. W., & Penn, E. M. (2015). Analyzing Big Data: Social Choice and Measurement. *PS: Political Science & Politics*, *48*(01), 95–101.

Wasserman, S., & Faust, K. (1994). Ch.5 Centrality and Prestige. In *Structural Analysis in the Social Sciences*: *Vol. 8*. *Social Network Analysis: Methods and Applications*. New York, NY: Cambridge University Press.

_____

Week 7 – March 8

Networks II: Tie formation

Kitts, J. A., & Quintane, E. (2019). Rethinking Social Networks in the Era of Computational Social Science. *Oxford Handbook of Social Networks.*, 29.

McPherson, M., Smith-Lovin, L., & Cook, J. M. (2001). Birds of a Feather: Homophily in Social Networks. *Annual Review of Sociology*, *27*(1), 415–444.

Rivera, M. T., Soderstrom, S. B., & Uzzi, B. (2010). Dynamics of Dyads in Social Networks: Assortative, Relational, and Proximity Mechanisms. *Annual Review of Sociology*, *36*(1), 91–115.

Malik, P., & Lee, S. (2020). Follow Me Too: Determinants of Transitive Tie Formation on Twitter. *Social Media + Society*, *6*(3), 1-12.

Wasserman, S., & Faust, K. (1994). Ch.6 Structural Balance and Transitivity. In *Structural Analysis in the Social Sciences*: *Vol. 8*. *Social Network Analysis: Methods and Applications*. New York, NY: Cambridge University Press.

_____

Deadline: March 20

Research outline due

Week 8 – March 22

Networks III: Groups and graphs

Golder, S. A., & Macy, M. W. (2014). Digital Footprints: Opportunities and Challenges for Online Social Research. *Annual Review of Sociology*, *40*(1), 129–152.

Kilduff, M., & Tsai, W. (2003). Ch.3 Is There Social Network Theory? A Critical Examination of Theoretical Foundations. In *Social Networks and Organizations*. Thousand Oaks, CA: SAGE Publications.

Valente, T. W. (2010). Ch.6 Groups. In *Social Networks and Health: Models, Methods, and Applications*. New York, NY: Oxford University Press.

Luke, D. A. (2015). *A User's Guide to Network Analysis in R*. Cham Hildesheim New York: Springer. Ch. 8 Subgroups

Borgatti, S. P., Everett, M. G., & Johnson, J. C. (2013). *Analyzing Social Networks* (1 edition). Los Angeles: SAGE Publications Ltd. Ch. 8 Testing hypotheses.

Week 9 – March 29

Networks IV: Structure and behavior

Centola, D. (2019). The Truth About Behavioral Change. *MIT Sloan Management Review*, (Winter Issue).

Mai, B., Liu, J., & Gonzalez-Bailon, S. (2015). Network Effects in the Academic Market: Mechanisms for Hiring and Placing PhDs in Communication (2007-2014): Hiring and Placement in Communication. *Journal of Communication*, *65*(3), 558–583.

Lusher, D., Koskinen, J., & Robins, G. (Eds.). (2012). *Exponential Random Graph Models for Social Networks: Theory, Methods, and Applications* (Illustrated edition). Cambridge: Cambridge University Press. Ch. 2, 3, 4, 5.

_____

Week 10 – April 5

Data visualization

Healy, K. (2018). Ch.1 Look at Data. In *Data Visualization: A Practical Introduction*. Princeton, NJ: Princeton University Press.

Fox, P., & Hendler, J. (2011). Changing the Equation on Scientific Data Visualization. *Science*, *331*(6018), 705–708.

Pfeffer, J. (2017). Visualization of Political Networks. In *The Oxford Handbook of Political Networks*. New York, NY: Oxford University Press

Foucault Welles, B., & Meirelles, I. (2015). Visualizing Computational Social Science: The Multiple Lives of a Complex Image. *Science Communication*, *37*(1), 34–58.

Wickham, H., & Grolemund, G. (2017). R for Data Science: Import, Tidy, Transform, Visualize, and Model Data. O'Reilly Media. Ch.1.

 Optional: Wickham, H. (2009). [Ggplot2: Elegant Graphics for Data Analysis](). Springer.

_____

Week 11 – April 12

Text I: Sentiment

Benoit, K. (2020). Text as Data: An Overview. In L. Curini & R. Franzese (Eds.), *The SAGE Handbook of Research Methods in Political Science and International Relations* (1st edition). SAGE Publications Ltd.

Curini, L., & Fahey, R. A. (2020). Sentiment Analysis and Social Media. In L. Curini & R. Franzese (Eds.), *The SAGE Handbook of Research Methods in Political Science and International Relations* (1st edition). SAGE Publications Ltd.

Young, L., & Soroka, S. (2012). Affective News: The Automated Coding of Sentiment in Political Texts. Political Communication, 29(2), 205–231.

 Bode, L., Budak, C., Ladd, J. M., et al. (2020). Words That Matter: How the News and Social Media Shaped the 2016 Presidential Campaign. Washington, DC.
Brookings Institution Press. Ch.5

Arnold, T. (2015). Humanities Data in R: Exploring Networks, Geospatial Data, Images, and Text. Springer Media. Ch.9 NLP

_____

_____

Week 12 – April 19

Text II: Topic models

Jacobi, C., Atteveldt, W. van, & Welbers, K. (2016). Quantitative analysis of large amounts of journalistic texts using topic modelling. *Digital Journalism*, *4*(1), 89–106.

Maier, D., Waldherr, A., Miltner, P., et al (2018). Applying LDA Topic Modeling in Communication Research: Toward a Valid and Reliable Methodology. Communication Methods and Measures, 12(2–3), 93–118.

Song, H., Eberl, J.-M., & Eisele, O. (2020). Less Fragmented Than We Thought? Toward Clarification of a Subdisciplinary Linkage in Communication Science, 2010–2019. *Journal of Communication*, *70*(3), 310–334.

Guo, L., Vargo, C. J., Pan, Z., Ding, W., & Ishwar, P. (2016). Big Social Data Analytics in Journalism and Mass Communication: Comparing Dictionary-Based Text Analysis and Unsupervised Topic Modeling. Journalism & Mass Communication Quarterly, 93(2), 332–359.

Arnold, T. (2015). Humanities Data in R: Exploring Networks, Geospatial Data, Images, and Text. Springer Media. Ch.10 Text analysis

Optional: Silge, J., & Robinson, D. (2017). Text Mining with R: A Tidy Approach. O'Reilly.

_____

Week 13 – April 6

Online experiments

Salganik, M. J. (2017). *Bit by Bit: Social Research in the Digital Age*. Princeton, NJ: Princeton University Press. Ch. 4 Running experiments

Kramer, A. D. I., Guillory, J. E., & Hancock, J. T. (2014). Experimental evidence of massive-scale emotional contagion through social networks. *Proceedings of the National Academy of Sciences*, *111*(24), 8788–8790.

Bail, C. A., Argyle, L. P., Brown, T. W. et al. (2018). Exposure to opposing views on social media can increase political polarization. *Proceedings of the National Academy of Sciences*.

Gross, K., Porter, E., & Wood, T. J. (2018). Identifying Media Effects Through Low-Cost, Multiwave Field Experiments. *Political Communication*, 1–16.

Munger, K. (2017). Tweetment Effects on the Tweeted: Experimentally Reducing Racist Harassment. *Political Behavior*, *39*(3), 629–649.

_____

_____

Week 14 – May 3

Ethics and open science

Salganik, M. J. (2017). *Bit by Bit: Social Research in the Digital Age*. Princeton, NJ: Princeton University Press. Ch. 6 Ethics

Crowcroft, J., Haddadi, H., & Henderson, T. (2018). Responsible Research on Social Networks: Dilemmas and Solutions. In B. F. Welles & S. González-Bailón (Eds.), *The Oxford Handbook of Networked Communication*. Oxford University Press.

Menchen-Trevino, E. (2018). Digital Trace Data and Social Research: A Proactive Research Ethics. In B. F. Welles & S. González-Bailón (Eds.), *The Oxford Handbook of Networked Communication*. Oxford University Press.

Bowman, N. D., & Keene, J. R. (2018). A Layered Framework for Considering Open Science Practices. *Communication Research Reports*, *35*(4), 363–372.

Freese, J., & Peterson, D. (2017). Replication in Social Science. *Annual Review of Sociology*, *43*(1), 147–165.

_____

Week 15 – May 10

Final presentations

_____

Deadline: May 12

Final project due

_____