

MSDS 452: Web and Network Data Science

Syllabus, Spring 2019

Course Description

This course shows how to acquire and analyze information from the web. It provides a comprehensive review of web analytics, including website usage and search performance metrics. The course introduces the mathematics of network science, including properties of graphs, random graph, small world, preferential attachment and diffusion models. Students use methods of network science with applications to information and social networks. They compute a variety of network metrics as they analyze software systems, website structure, and user interactions through electronic communications and social media. This is a case study and project-based course with a strong programming component.

Course Objectives

By the end of this course, you will be able to:

- Recognize the importance of analyzing the web and social media.
 - Demonstrate the appropriate skills for extracting social networks from the web and social media.
 - Analyze web and network data (including semi-structured and unstructured text) using open-source software and appropriate data science models and methods. □
 - Produce static and interactive data visualizations, showing meaningful patterns in web and network data. □
 - Produce research reports for management clients that are consistent with professional standards for expository writing. □
 - Address business issues associated with having an online presence, including networks of various interactions in the web and social media. □
 - Transform data and research results into actionable insights □
-

Prerequisites

- MSDS 420-DL Database Systems and Data Preparation or CIS 417 Database Systems Design and Implementation and

- MSDS 422-DL Practical Machine Learning or CIS 435 Practical Data Science Using Machine Learning.
-

Required Readings and Resources

Required Readings

- Newman, M. E. (2018). *Networks* (2nd ed.). Oxford: Oxford University Press. ISBN-13: 978-0198805090
- Miller, T. W. (2015). *Web and Network Data Science: Modeling Techniques in Predictive Analytics*. Upper Saddle River, N.J.: Pearson. ISBN-13: 978-0-13-388644-3
- Zafarani, R. Abbasi, M.A., and Liu, H. (2014). *Social Media Mining*. New York City, NY: Cambridge University Press. ISBN: 978-1139088510.
- Zinoviev, D., & Tulton, A. O. (2018). *Complex network analysis in Python: Recognize, construct, visualize, analyze, interpret*. Raleigh, NC: The Pragmatic Bookshelf. ISBN-13: 978-1680502695

Required Reference Safari Books Online for free books for students are referenced below:

- Tsvetovat, M., & Kouznetsov, A. (2012). *Social network analysis for startups*. Sebastopol, CA.: O'Reilly. ISBN-13: 978-1449306465
- Russell, M. A. and M. Klassen (2019). *Mining the Social Web: Data Mining Facebook, Twitter, LinkedIn, Instagram, GitHub, and More* (3rd ed.). Sebastopol, CA.: O'Reilly. ISBN-13: 978-1491985045

Course Reserves

Some readings will be available through the Course Reserves in the left navigation menu. Assignment and Discussion forum instructions will note which readings are to be accessed through Course Reserves. For assistance with Course Reserves, e-mail e-reserve@northwestern.edu. To ask a librarian for assistance, visit Northwestern's [Ask A Librarian](#) page.

Required Software

You will need access to a web browser such as Google Chrome, Mozilla Firefox, or Safari. Additionally, Python will need to be downloaded and installed on your computer.

Assignment Overview and Grading Breakdown

The team assignment involves a research project carried out throughout the term. Teams of three to five students are formed based on expressed interests across a number of possible topic areas. The final project, the delivery of the final project report, is due at the end of the course and shared with all students in the class.

Grading and feedback turnaround will be one week from the due date. You will be notified if turnaround will be longer than one week. The Midterm and Final Presentation will be graded and returned within two weeks.

The discussion forums, written assignments, and midterm and final presentation will be graded based on specific criteria or a rubric. The criteria or rubric for each type of assessment will be available in the course. To view the discussion forum rubric, click the gear icon in the upper right corner of the page and choose Show Rubric. The Written Assignment Rubric and Presentation Rubric will automatically appear on the page.

Your final grade will be determined as follows:

| | |
|---|----------------------------|
| Class Participation - Discussion Posts Weekly class participation includes posting to discussion forums. You will be expected to provide quality insights in class discussion forums in a timely fashion. | 15% or 150 Points |
| Introduction and Statement of Interest During the first week, students are to rank topics, based on personal interest, for their group project. These ratings will be used to create student teams. This is done via discussion board at the same time as personal introductions. | 10% Or 100 Points |
| Individual Assignments Individual assignments in the course involve the collection and analysis of web and social media data. Students will be assigned a data type and they will be requested to conduct a particular analysis (corresponding to the modules of the course). The analysis will be implemented with Python in a Jupyter notebook. For each assignment, a template Jupyter notebook together with jump start code will be provided and may be used, though in certain cases it must be expanded upon to make one's own work. | 50% Or 500 Points |

| | |
|---|------------------------------|
| Team Final Project Students will be assigned groups and a topic the first week of this course. After refining their topic to a category of interest, the group will collect data and conduct the computations of their topic using Pythor (in Jupyter notebooks). Additional details on this project will be available in the course. | 25% Or 250 Points |
| TOTAL | 100% Or 1000 Points |

Grading Scale

| Letter Grade | % and Points |
|--------------|----------------------------|
| A | 94-100% or 940-1000 Points |
| A- | 90-93% or 900-939 Points |
| B+ | 87-89% or 870-899 Points |
| B | 84-86% or 840-860 Points |

| | |
|----|-------------------------------|
| B- | 80-83% or 800-839 Points |
| C+ | 77-79% or 770-799 Points |
| C | 74-76% or 740-769 Points |
| C- | 70-73% or 700-739 Points |
| F | Below 70% or Below 700 Points |

Late Work

Late work will be accepted only in the event of an instructor-approved absence. Contact your instructor as soon as possible, at least 24 hours in advance.

Online Communication Expectations

Discussion Forums

The purpose of the discussion boards is to allow students to freely exchange ideas. It is imperative to remain respectful of all viewpoints and positions and, when necessary, agree to respectfully disagree. While active and frequent participation is encouraged, cluttering a discussion board with inappropriate, irrelevant, or insignificant material will not earn additional points and may result in receiving less than full credit. Frequency matters, but contributing content that adds value is paramount. Please remember to cite all sources—when relevant—in order to avoid plagiarism. Please post your viewpoints first and then discuss others' viewpoints.

The quality of your posts and how others view and respond to them are the most valued. A single statement mostly implying “I agree” or “I do not agree” is not counted as a post. Explain, clarify, politely ask for details, provide details, persuade, and enrich communications for a great discussion experience. Please note, there is a requirement to respond to at least two fellow class members posts. Also, remember to cite all sources—when relevant—in order to avoid plagiarism.

Online Communication Etiquette

Beyond interacting with your instructor and peers in discussions, you will be expected to communicate by Canvas message, email, and sync session. Your instructor may also make themselves available by phone or text. In all contexts, keep your communication professional and respect the instructor’s posted availability. To learn more about professional communication, please review the [Communicating Effectively with Faculty](#) guide.

Just as you expect a response when you send a message to your instructor, please respond promptly when your instructor contacts you. Your instructor will expect a response within two business days. This will require that you log into the course site regularly and set up your notifications to inform you when the instructor posts an announcement, provides feedback on work, or sends you a Canvas message. For guidance on setting your notifications, please review [How do I set my Canvas notification settings as a student?](#) It is also recommended that you check your u.northwestern e-mail account regularly, or forward your u.northwestern e-mail to an account you check frequently.

Participation and Attendance

This course will not meet at a particular time each week. All course goals, session learning objectives, and assessments are supported through classroom elements that can be accessed at any time. To measure class participation (or attendance), your participation in threaded discussion boards is required, graded, and paramount to your success in this course. Please note that any scheduled synchronous meetings are optional. While your attendance is highly encouraged, it is not required and you will not be graded on your attendance or participation.

Student Support Services

AccessibleNU

This course is designed to be welcoming to, accessible to, and usable by everyone, including students who are English-language learners, have a variety of learning styles, have disabilities, or are new to online learning. Be sure to let me know immediately if you encounter a required element or resource in the course that is not accessible to you. Also, let me know of changes I can make to the course so that it is more welcoming to, accessible to, or usable by students who take this course in the future.

Northwestern University and [AccessibleNU](#) are committed to providing a supportive and challenging environment for all undergraduate, graduate, professional school, and professional studies students with disabilities who attend the University. Additionally, the University and AccessibleNU work to provide students with disabilities and other conditions requiring accommodation a learning and community environment that affords them full participation, equal access, and reasonable accommodation. The majority of accommodations, services, and auxiliary aids provided to eligible students are coordinated by AccessibleNU, which is part of the [Dean of Students Office](#).

SPS Student Services

The Department of [Student Services](#) supports the academic and professional growth of SPS students. The Student Services team guides students through academic planning, policies, and administrative procedures, and promotes a supportive environment to foster student success. Students are encouraged to actively make use of the resources and staff available to assist them: Academic and Career Advisers, Counseling and Health Services, Student Affairs, Legal Services, Financial Aid and Student Accounts, among other services. For a comprehensive overview of course and program processes and policies and helpful student resources, please refer to your [SPS Student Handbook](#).

Academic Support Services

Northwestern University Library

As one of the leading private research libraries in the United States, Northwestern University Library serves the educational and information needs of its students and faculty as well as scholars around the world. Visit the [Library About](#) page for more information or contact Distance Learning Librarian Tracy Coyne at 312-503-6617 or tracy-coyne@northwestern.edu.

Library Resources

- [Connectivity: Campus Wireless and Off-Campus Access to Electronic Resources](#)
- [Reserve a Library Study Room](#)
- [Sign up for an in-person or online Research Consultation Appointment](#)
- [Getting Available Items: Delivery to Long-Distance Patrons](#)
- [Social Science Data Resources](#)
- [Resources for Data Analysis](#)

The Writing Place

SPS sponsors a free tutoring service available on the Chicago campus. Students can meet with a tutor either in person or online. For more information or to book an appointment, see [The](#)

[Writing Place, Chicago Campus](#). SPS students can also set up appointments at [The Writing Place, Evanston Campus](#).

The Math Place

The Math Place is a free tutorial service provided to students currently enrolled in Northwestern University's School of Professional Studies courses or in other Northwestern University courses. Students of all levels can benefit from the individual tutoring provided from this service, whether they are taking undergraduate or graduate level courses. To book an appointment, go to [The Math Place](#) website.

Academic Integrity at Northwestern

Students are required to comply with University regulations regarding academic integrity. If you are in doubt about what constitutes academic dishonesty, speak with your instructor or graduate coordinator before the assignment is due and/or examine the University website. Academic dishonesty includes, but is not limited to, cheating on an exam, obtaining an unfair advantage, and plagiarism (e.g., using material from readings without citing or copying another student's paper). Failure to maintain academic integrity will result in a grade sanction, possibly as severe as failing and being required to retake the course, and could lead to a suspension or expulsion from the program. Further penalties may apply. For more information, visit [The Office of the Provost's Academic Integrity page](#). Some assignments in SPS courses may be required to be submitted through Turnitin, a plagiarism detection and education tool. You can find [an explanation of the tool here](#).

Course Technology

This course will involve a number of different types of interactions. These interactions will take place primarily through the Canvas system. Please take the time to navigate through the course and become familiar with the course syllabus, structure, and content and review the list of resources below.

Canvas

The [Canvas Student Center](#) includes information on communicating in Canvas, navigating a Canvas course, grades, additional help, and more. The [Canvas at Northwestern](#) website provides information of getting to know Canvas at Northwestern and getting Canvas support. The [Canvas Student Guide](#) provides tutorials on all the features of Canvas. For additional Canvas help and support, you can always click the Help icon in the lower left corner to begin a live chat with Canvas support or contact the Canvas Support Hotline. The [Canvas Accessibility Statement](#) and [Canvas Privacy Policy](#) are also available.

BlueJeans

We will use BlueJeans for optional synchronous meetings. The [Northwestern IT YouTube channel on Blue Jeans Video conferencing](#) and the [Canvas Learning Center BlueJeans page](#) provide additional guidance for using BlueJeans.

The [Blue Jeans Privacy Policy](#) and the [Accessibility Features on BlueJeans](#) are also available. Please note that any scheduled synchronous meetings are optional. While your attendance is highly encouraged, it is not required and you will not be graded on your attendance or participation. These synchronous sessions will be recorded, so you will be able to review the session afterwards.

Panopto

Videos in this course may be hosted in Panopto. If you have not used Panopto in the past, you may be prompted to login to Panopto for the first time and authorize Panopto to access your Canvas account. You can learn more about using Panopto and login to Panopto directly by visiting the Panopto guide on the [Northwestern IT Resource Hub](#). Depending on the assignment requirements of this course, you may be asked to create videos using Panopto in addition to viewing content that your instructor has provided through Panopto.

The Panopto [Privacy Policy](#) and the [Accessibility Features](#) on Panopto are also available.

Lynda.com

Some required resources used in this course will be accessed via Lynda.com. Northwestern offers access to [Lynda.com](#) courses for students, faculty, and staff. Courses cover a wide range of technology and business topics including databases, programming, and data analysis. For more information, see the [Human Resources Lynda.com page](#).

The [Privacy Policy](#) and [Accessibility Statement](#) are also available.

Read&Write Gold

Read&Write Gold is an optional text reading and writing program with numerous beneficial features. Originally developed to assist users with print disabilities, such as visual impairments, dyslexia, ADHD, etc., this program provides a wide array of tools to assist with reading, writing, and notetaking. One of the most useful tools is the text-to-speech function, which students may use to convert digital text into an audio format. Read&Write Gold is available for free to all Northwestern students, faculty, and staff. Visit the [Northwestern IT site on Read&Write Gold](#) for more information about the software, as well as instructions on how to download it.

Python

Python is a programming language that can read data frames and create visualizations.

To download and/or obtain additional information, visit the [Python website \(Links to an external site.\)](#).

The [Privacy Policy for Python can be found here \(Links to an external site.\)](#).

Jupyter Notebook

Part of the Anaconda Navigator package, Jupyter Notebook can be used as an IDE for Python coding. Spyder is another tool that may be preferable if you are accustomed to the design of RStudio.

To download Anaconda and/or obtain additional information, visit the [Anaconda website \(Links to an external site.\)](#)

The [Privacy Policy for Anaconda can be found here \(Links to an external site.\)](#)

Git/GitHub for Sparknotes

Git and GitHub have been designed for knowledge workers—designers, document authors, software developers, and data scientists. Git helps us to maintain a work history, saving all the intermediate versions of our creations locally. If we want to go back to an earlier version, we can do that with ease. Git runs offline on personal computers, and GitHub provides a server-based storage facility for our work, a collection of repositories that can be shared with others. GitHub repositories reside on the World Wide Web, and we can use GitHub to facilitate collaboration among team members.

To download and/or obtain additional information, visit [GitHub \(Links to an external site.\)](#).

The [Privacy Policy for GitHub can be found here \(Links to an external site.\)](#).

Minimum Required Technical Skills

Students in an online program should be able to do the following:

- Communicate via email and Canvas discussion forums.
- Use web browsers and navigate the World Wide Web.
- Use the learning management system Canvas.
- Use integrated Canvas tools (e.g., BlueJeans, Panopto, Course Reserves).
- Use applications to create documents (e.g., Microsoft Word).
- Use applications to share files (e.g., Box, Google Drive).

Systems Requirements for Distance Learning

Students and faculty enrolled in SPS online classes should have access to a computer with the [Minimum System Requirements](#).

Technical Help and Support

The [SPS Help Desk](#) is available for Faculty, Students and Staff to support their daily IT needs. For additional technical support, contact the [Northwestern IT Support Center](#).

Permissions

Instructional Materials

This course was developed in partnership with Distance Learning staff in the School of Professional Studies at Northwestern University. Every effort has been made to responsibly acquire instructional materials for this class, by adhering to copyright law, obtaining permission from copyright holders, selecting Open Educational Resources (OERs) and Creative Commons (CC) materials, and using citations to credit the work of others.

The same is expected of students in this course. Please review the Academic Integrity statement for more information.

Sharing Course Content

Content within this course—including assignment descriptions, exam questions, and other course components—may not be distributed outside of the course, either to other students or on the Internet more broadly.

Student Ownership of Content

Students retain ownership of all content developed while completing this course, as dictated by the university [Copyright Policy](#) (“copyright ownership resides with the Creator(s) of copyrightable works”).

Per the Family Educational Rights and Privacy Act ([FERPA](#)), if your instructor wishes to share your work with future students, your permission must be obtained in writing.

Your instructor may limit access to the course after a cutoff date. When you complete the course, please ensure that you have saved all work. You may not be able to return to the course to download your submissions.

Course Schedule

In the course schedule you will find a list of each module’s readings and resources and the assignments due. Links to videos, handouts, and other resources can be found on each module’s Readings and Resources page throughout the course.

- Module 1
- Module 2
- Module 3
- Module 4
- Module 5
- Module 6
- Module 7
- Module 8
- Module 9
- Module 10

Module 1: *Introduction to Web and Network Science*

Learning Objectives

After this module, students will be able to:

- Discuss what is Web and Network Science. (Discussion Board)
- Talk about information resources from the Web. (Discussion Board)
- Analyze type of web, networks and social media user activities. (Progress Bid and Statement of Interest)
- Devise a method of navigating in digital data from the web and social media. (Individual Assignment)

Readings and Resources

Required Resources

- Miller, T. W. (2015). *Web and Network Data Science: Modeling Techniques in Predictive Analytics*. Upper Saddle River, N.J.: Pearson. ISBN-13: 978-0-13-388644-3 (Chapter 1: Getting Technical, pages 1–24) □
- Pastor-Satorras, R., and Vespignani, A. (2004). *Evolution and Structure of the Internet: A Statistical Physics Approach*. New York City, NY: Cambridge University Press. ISBN-13: 978-0521826983. (Chapter 7, pages 140-165, and Chapter 8, pages 166-179)
- Latour, B., Jensen, P., Venturini, D., Grawin, S., & Boullier, D. (2012). [The whole is always smaller, than its parts \(Links to an external site.\)Links to an external site.](#). *The British Journal of Sociology*, 63 (4), 590-615. (pages 590-595)

Assignments

- [Discussion board](#), participation due Sunday evening 11:55 pm Central.
- [Project Bid and Statement of Interest](#)
- [Individual Assignment](#)

Module 2: *Gathering Data from the Web*

Learning Objectives

After this module, students will be able to:

- Analyze mechanisms for gathering data from the web. (Discussion Board)
- Examine data sources using web crawlers and web scrapers. (Individual Assignment)
- Discuss application programming interfaces (APIs) for gathering social media data. (Discussion Board)

- Familiarize yourselves with human readable data exchange formats (XML, JSON, etc.). (Discussion Board)

Readings and Resources

Required Resources

- Miller, T. W. (2015). *Web and Network Data Science: Modeling Techniques in Predictive Analytics*. Upper Saddle River, N.J.: Pearson. ISBN-13: 978-0-13-388644-3 (Chapter 1: Getting Technical, pages 25-42)
- Russell, M. A. and M. Klassen (2019). *Mining the Social Web: Data Mining Facebook, Twitter, LinkedIn, Instagram, GitHub, and More* (3rd ed.). Sebastopol, CA.: O'Reilly. ISBN-13: 978-1491985045 (Chapter 5, pages 181-224)
- Dale, K. (2016). *Data Visualization with Python & JavaScript: Scrape, Clean, Explore, and Transform Your Data*. Sebastopol, CA: O'Reilly. (Chapter 3, pages 73-104, and Chapter 5, pages 145-174).

Assignments

- [Discussion board](#), participation due Sunday evening 11:55 pm Central.
- [Individual Assignment](#)

Module 3: *Web and Social Media Analytics Essentials*

Learning Objectives

After this module, students will be able to:

- Analyze web traffic and visitor statistics. (Discussion Board)
- Examine search engines and link analysis. (Discussion Board)
- Discuss influence and behavior analytics. (Discussion Board)
- Devise a simple recommendation procedure in social media. (Individual Assignment)

Readings and Resources

Required Resources

- Miller, T. W. (2015). *Web and Network Data Science: Modeling Techniques in Predictive Analytics*. Upper Saddle River, N.J.: Pearson. ISBN-13: 978-0-13-388644-3 (Chapter 1: Getting Technical, pages 43-67)

- Zafarani, R. Abbasi, M.A., and Liu, H. (2014). *Social Media Mining*. New York City, NY: Cambridge University Press. ISBN: 978-1107512818. (Chapters 9 and 10, pages 245-293)

Assignments

- [Discussion board](#), participation due Sunday evening 11:55 pm Central.
- [Individual Assignment](#)

Module 4: *Social Network Analysis: Essentials from Graph Theory*

Learning Objectives

After this module, students will be able to:

- Discuss about empirical networks. (Discussion Board)
- Analyze in Python undirected and directed networks (simple and multigraphs). (Individual Assignment)
- Examine two-mode networks (bipartite graphs) and trees. (Discussion Board)
- Discuss network connectivity and assortativity. (Discussion Board)

Readings and Resources

Required Readings

- Newman, M. E. (2018). *Networks* (2nd ed.). Oxford: Oxford University Press. ISBN-13: 978-0198805090 (Chapter 4, pages 47-69, Chapter 6, pages 105-157 and Chapter 10, pages 304-340)
- Zafarani, R. Abbasi, M.A., and Liu, H. (2014). *Social Media Mining*. New York City, NY: Cambridge University Press. ISBN: 978-1107512818. (Chapters 2, pages 13-50)

Assignments

- [Discussion board](#), participation due Sunday evening 11:55 pm Central.
- [Individual Assignment](#)

Module 5: *Social Network Analysis: Patterns and Measures*

Learning Objectives

After this module, students will be able to:

- Explain triangles, cycles and cliques. (Discussion Board)
- Explore subgraphs and egocentric-subgraphs. (Discussion Board)
- Compute network measures and metrics. (Individual Assignment)
- Discuss network visualization. (Discussion Board)

Readings and Resources

Required Resources

- Newman, M. E. (2018). *Networks* (2nd ed.). Oxford: Oxford University Press. ISBN-13: 978-0198805090 (Chapter 7, pages 158-217, Chapter 6, pages 105-157)
- Tsvetovat, M., & Kouznetsov, A. (2012). *Social network analysis for startups*. Sebastopol, CA.: O'Reilly. ISBN-13: 978-1449306465 (Chapter 4, pages 61-92)
- Borgatti, S.P., Everett, M.G., & Johnson, J.C. (2018). *Analyzing Social Networks* (2nd ed.). Sage. ISBN-13: 978-1526404107 (Chapter 7, pages 115-142, Chapter 10, pages 189-207, and Chapter 15, pages 305-330)

Assignments

- [Discussion board](#), participation due Sunday evening 11:55 pm Central.
- [Individual Assignment](#)

Module 6: Social Network Analysis: Algorithms and Models

Learning Objectives

After this module, students will be able to:

- Explore shortest Paths and Traversals. (Discussion Board)
- Compute in Python network clustering and partitions (communities). (Individual Assignment)
- Discuss network models (random graphs, preferential attachment, small worlds). (Discussion Board)
- Discuss statistical analysis of network data (link prediction, network inference, ERGMs). (Discussion Board)

Readings and Resources

Required Resources

- Newman, M. E. (2018). *Networks* (2nd ed.). Oxford: Oxford University Press. ISBN-13: 978-0198805090 (Chapter 11, pages 342-368, Chapter 13, pages 434-492, and Chapter 14, 494-568)
- Barabási, A.L. (2016). *Network Science*. New York City, NY: Cambridge University Press ISBN-13: 978-1107076266. (Chapter 4, pages 112-163, and Chapter 5, pages 164-201)
- Watts, D. J. (2003). *Six Degrees: The Science of a Connected Age*. New York: W. W. Norton. (Chapter 3: Small Worlds, pages 69–100, and Chapter 4: Beyond the Small World, pages 101– 129)

Assignments

- [Discussion board](#), participation due Sunday evening 11:55 pm Central.
- [Individual Assignment](#)

Module 7: Network Sciences: Application and Processes

Learning Objectives

After this module, students will be able to:

- Compute simple examples of social influence, contagion and diffusion. (Individual Assignment)
- Discuss percolation and network resilience. (Discussion Board)
- Discuss network epidemics. (Discussion Board)
- Examine spread of social behavior. (Discussion Board)

Readings and Resources

Required Resources

- Tsvetovat, M., & Kouznetsov, A. (2012). *Social network analysis for startups*. Sebastopol, CA.: O'Reilly. ISBN-13: 978-1449306465 (Chapter 6, pages 109-136)
- Pastor-Satorras, R., and Vespignani, A. (2004). *Evolution and Structure of the Internet: A Statistical Physics Approach*. New York City, NY: Cambridge University Press. ISBN-13: 978-0521826983 (Chapter 6, pages 112-139, and Chapter 9, pages 180-210)
- Barrat, A., Barthélemy M., & Vespignani, A. (2008). *Dynamical Processes on Complex Networks*. New York City, NY: Cambridge University Press. ISBN-13: 978-0521879507. (Chapter 6, pages 116-135, Chapter 9, pages 180-215, and Chapter 10, 216-241)
- Newman, M. E. (2018). *Networks* (2nd ed.). Oxford: Oxford University Press. ISBN-13: 978-0198805090 (Chapter 15, pages 569-606, and Chapter 16, 607-674)

Assignments

- [Discussion board](#), participation due Sunday evening 11:55 pm Central.
- [Individual Assignment](#)

Module 8: *Social Network Analysis of Data Extracted from the Web*

Learning Objectives

After this module, students will be able to:

- Navigate on networks of hyperlinks among web pages. (Discussion Board)
- Examine networks of Wikipedia pages and Wikipedia topics. (Discussion Board)
- Compute with Python a network of emails. (Individual Assignment)
- Explore co-occurrence and intersection networks (“wordnets” and “co-word networks”) extracted from textual data in the Web. (Discussion Board)

Readings and Resources

Required Resources

- Russell, M. A. and M. Klassen (2019). *Mining the Social Web: Data Mining Facebook, Twitter, LinkedIn, Instagram, GitHub, and More* (3rd ed.). Sebastopol, CA.: O'Reilly. ISBN-13: 978-1491985045 (Chapter 5, pages 181-224, and Chapter 6, pages 225-278)
- Diesner, J., Frantz, T., & Carley, K.M. (2005). [Communication networks from the Enron email corpus “It’s always about the people Enron is no different” \(Links to an external site.\)Links to an external site.](#) *Computational and Mathematical Organization Theory*, 11(3), 201-228
- Boudourides, M. (2017). [The citation network among Wikipedia pages on Dynamical Systems and Mechanics \(Links to an external site.\)Links to an external site.](#)
- Boudourides, M. (2017). [Jumbling up two assemblages: The networks of the Deleuze and Meillassoux dictionaries \(Links to an external site.\)Links to an external site.](#)

Assignments

- [Discussion board](#), participation due Sunday evening 11:55 pm Central.
- [Individual Assignment](#)

Module 9: *Social Network Analysis of Data Extracted from Social Media*

Learning Objectives

After this module, students will be able to:

- Discuss social media analytics (focused on Twitter). (Discussion Board)
- Compute with Python networks of co-occurrent hashtags from Twitter data. (Individual Assignment)
- Compute with Python networks of users' mentions and users' retweets from Twitter data. (Individual Assignment)
- Examine sentiment analysis of the textual content of Twitter data. (Discussion Board)

Readings and Resources

Required Resources

- Russell, M. A. and M. Klassen 2019. Mining the Social Web: Data Mining Facebook, Twitter, LinkedIn, Instagram, GitHub, and More (third ed.). Sebastopol, Calif.: O'Reilly. [ISBN-13: 978-1491985045] (Chapter 1, pages 5-44, and Chapter 9, pages 329-380)
- Russell, M.A. 2011. 21 Recipes for Mining Twitter: Distilling Rich Information from Messy Data. O'Reilly Media [ISBN-13: 978-1449303167] (passim)
- Siapera, E., Boudourides, M., Lenis, S., & Suiter, J. 2018. [Refugees and Network Publics on Twitter: Networked Framing, Affect, and Capture. Social Media + Society](#)

Assignments

- [Discussion board](#), participation due Sunday evening 11:55 pm Central.
- [Individual Assignment](#)

Module 10: *Course Conclusion*

Assignments

- [Discussion board](#), participation due Sunday evening 11:55 pm Central.
- [Team Final Project](#), due Sunday by 11:55 pm Central.