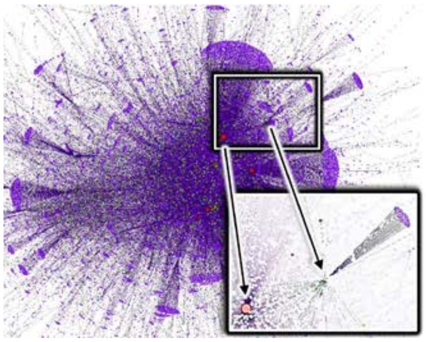
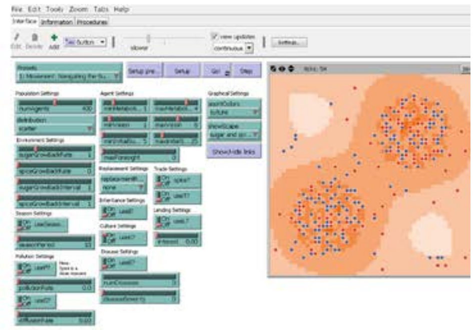
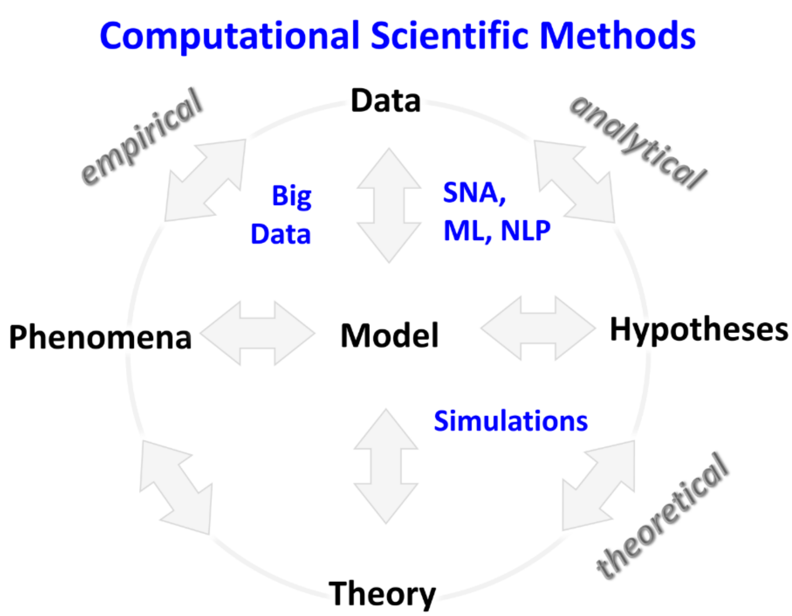
Computational Social Science (CMN150V)



Cross-Campus Enrollment



Digital technology has not only revolutionized society but also the way we can study it. An increasing part of human interaction leaves a massive **digital footprint** behind. Studying it allows us to gain unprecedented insights into what society is and how it works, including its intricate **social networks** that had long been obscure. Computational power allows us to detect hidden patterns through analytical tools like **machine learning** and to **natural language processing**. Finally, **computer simulations** enable us to explore hypothetical situations that may not even exist in reality, but that we would like to exist: a better world.



Computational social science provides us with the tools to explore **new scenarios** in a way that is as intriguing as playing a video game, while at the same time grounding it into the **empirical reality** of the world around us. This course gives an introduction to some of the exciting possibilities of **how to do research**. While no formal requisites are necessary to join this course, at the end you will **web-scrape 'big' data** from the internet, execute a **social network analysis (SNA)**, find hidden patterns with **machine learning (ML)** and **natural language processing (NLP),** and create an **agent-based computer simulation** to explore what might happen if we would change certain things in society.

Given the myriad of possibilities offered by computational social science and by online course delivery, this course draws on the expertise from **faculty members of all 10 UC campuses**, including from faculty from UC Berkeley, UC Davis, UC Irvine, UC Los Angeles, UC Merced, UC Riverside, UC San Diego, UC San Francisco, UC Santa Barbara, and UC Santa Cruz. Hands-on labs and overall coordination is provided by M. Hilbert (UC Davis).

Contributing faculty include (not exhaustive):

UC Berkeley: *Joshua Blumenstock*, Assistant Prof. iSchool

UC Davis: *Martin Hilbert*, Associate Prof., Dpt. of Communication & *Seth Frey*, Assistant Prof., Dpt. of Communication & *Cynthia Gates*, Director of the IRB.

UC Irvine: *Lisa Pearl*, Associate Prof. Cognitive Sciences & *Ines Levin*, Assistant Prof. Dpt. Political Science

UC Los Angeles: *PJ Lamberson*, Assistant Prof. Communication Studies

UC Merced: *Paul Smaldino*, Assistant Prof. Cognitive and Information Sciences

UC Riverside: *Christian Shelton*, Prof. Computer Science & *Dena Plemmons*, Director Research Ethics Education

UC San Diego: *James Fowler*, Prof. Global Public Health and Political Science

UC San Francisco: *Maria Glymour*, Associate Prof. School of Medicine, Social Epidemiology & Biostatistics

UC Santa Barbara: *René Weber*, Prof. Dpt. of Communication & Media Neuroscience Lab

UC Santa Cruz: *Marilyn Walker*, Prof. Computer Science, Director, Computational Media

Course Outline

Each session is three-fold. The **first part** consists of lectures from the coordinating instructor to introduce the **main conceptual ideas**. The **second part** consists of several segments from contributing **UC faculty members**, who provide examples from their research, real world applications, and future outlooks. The **third part** consists of an **applied lab** that explores some research tools hands-on. All software is freely available and none requires concrete programming skills. Lecture videos link to tutorial videos on diverse concepts from the social sciences, programming, and statistics. Previous exposure to scientific research methods is welcome, but no specific prerequisite is required to participate successfully in this course.

(1) **Introduction:** Why Science? Why Social Science? Why Computational Social Science?

The course serves as an introduction to scientific thinking. We ask the very same questions usually asked by traditional introductory courses to social science research methods, just that our answers will be somewhat different. We will answer the question about how to obtain information about people not by learning how to develop surveys, but by web scraping people's digital footprint; we distinguish between different social groups not only on basis who they are, but also on who they are connected with; instead of deciding how to analyzing the obtained data, we let machine learning algorithms find hidden patterns; and we build models not by setting up and solving differential equations but by creating and playing with visually intuitive computer simulations. While we recognize the strengths of the traditional approaches, our emphasis is set on these new and complementary possibilities. The goal stays the same: to better understand, predict and ultimately shape the social world around us.

(2) **Big Data:** digital social footprints

Every digital interaction almost inevitably leaves behind a digital footprint. With over 99% of all information available in digital format, and with a mobile penetration of 98% worldwide, the so-called Big Data paradigm allows us to study human behavior and social interaction to an unprecedented degree of granularity. Traditional social scientists from Adam Smith, over Karl Marx, to Max Weber, who painstakingly collected empirical evidence by hand, would not believe how easy it is nowadays to observe society as a whole in action through its digital footprint, without leaving one's seat. Today's digital footprint allows us to identify even most ephemeral aspects of social life, from communication flows, to psychological personality traits, and the level of happiness of a society. In our **applied lab**, we will obtain some of this digital footprint ourselves by scraping content from a social media site.

(3) **Social Network Analysis:** network structure

A society is a network of individuals, which makes the analysis of social network structure decisive. We review a large variety of social networks and talk about their importance for human and social dynamics. Once we are familiar with networks, we discuss concepts and metrics of social networks, such as different measures of centrality, network partitioning into groups and clusters, homophily, and the implications of stronger and weaker ties. In our **applied lab**, we will start using software to analyze social networks in practice by analyzing the social network of our class with a software.

(4) **Machine Learning:** hidden social patterns

Thanks to the massive amounts of data produced in the digital age, machine learning applications are able to identify patterns that allow them to reach conclusions that frequently surpass the quality of any human judgment. As a result, we have endowed algorithms with the tasks of executing 3 out of 4 transactions on U.S. stock market, managing our electric grid, identifying credit card fraud, and deciding which information we are exposed to online. We trust them with our lives by letting them fly our airplanes, drive our metros, and manage the brakes in our cars. We rely on them as translators, as psychological therapists, and by now they are even the successful matchmakers for 1 in 3 marriages in America through online dating. We will discuss the opportunities and threats of artificial intelligence and explore the behavior of some machine learning applications in our **applied lab**.

(5) **Language processing:** from qualitative prose to quantitative science

The digital footprint provides massive amounts of text. The only reasonable way to make sense of this information flood is with automated text analysis, which is known as natural language processing. We will learn about how companies and governments use this automated kind of analysis to make sense of massive amounts of online text. In our **applied lab**, we explore how one of the most well-known artificial intelligences makes use of natural language processing: IBM's Watson, who famously beat the human champions of Jeopardy.

(6) **Computer Simulations (ABM part 1):** agent-based models

Social science can be as intuitive as playing a video game. Much like the well-known video game SimCity is used to build and manage an artificial city, agent-based models are used to creating and study artificial societies. The benefit of modeling social dynamics is that we can explore scenarios that have never existed in the real world, and we can carelessly pursue trial and error techniques without hurting anybody (but maybe our virtual agents). In this sense, computer simulations give us the opportunity to explore how we can make the world a better place, better than it ever existed in reality. In our **applied lab,** we will explore several artificial societies hands-on with a simulation software.

(7) **Computer Simulations (ABM part 2):** growing artificial societies

The generativist motto of computer generated social science is: "If you didn't grow it, you didn't explain it!" We discuss the bigger picture of what computer simulations can teach us about the complexity of social systems. We talk about the practical applications and explore how we can combine hypothetical models with real world data. We discuss the implications and limits of this way of doing science and end up talking about the fundamental limits of science in general. In our **applied lab**, we will grow our own artificial society by coding it up in an agent-based model simulation software.

(8) Bringing it all together: integrative lab

This final week consists of one large lab that brings together the tools that we have explored. We will first scrape data from an online social media site (drawing on week (2)). We will transfer the collected data and analyze the resulting networks (week (3)). We analyze some key aspects of it in depth, using natural language processing (weeks (4) and (5)) and use the gained conclusions to calibrate a computer simulation model that allows us to explore aspects that we did not find in our empirical reality, but that might be interesting to explore in order to improve this aspect of society (drawing on weeks (6) and (7)). The result is the first glimpse at a new way of doing social science in a digital age.

(9) Research Ethics: norms or standards for responsible research

The digital age has also created new ethical challenges. Researchers, companies and governments all face this dilemma and it is truly a dilemma, as there is currently much uncertainty and disagreement about the appropriate conduct of some digital age social research. We will review some of the fundamental principles that guide research and explore some of the borderline cases from academic, commercial and governmental research. As we are all apprentices in this new way of doing research, we have to be aware of the ethical implications.

(10) Exam week: online questionnaire and online ProctorU exam

Selected Related Readings

Cioffi-Revilla (2014). *Introduction to Computational Social Science*. London: Springer.

Easley & Kleinberg (2010). *Networks, Crowds, and Markets*. Cambridge University Press.

Epstein & Axtell (1996). *Growing Artificial Societies: Social Science from the Bottom Up*. Bradford.

Hanneman & Riddle (2005). *Introduction to Social Network Methods*. UC Riverside.

Krippendorff (2004). *Content Analysis: An Introduction to Its Methodology*. SAGE.

Lazer, Pentland, Adamic, Aral, Barabási… (2009). *Computational Social Science*. Science, 323(5915), 721–3.

Mayer-Schönberger & Cukier (2013). *Big data: a revolution that will transform how we live, work and think*.

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

Schelling (2006). *Micromotives and Macrobehavior*. W. W. Norton & Company.

Wilensky & Rand (2015). *An Introduction to Agent-Based Modeling… with NetLogo*. The MIT Press.

Evaluation and grade composition

Video attendance questions	20 %	Mid-exam	10 %
Assignments from labs	30 %	Final exam	40 %