Northwestern | Kellogg

Workshop Descriptions and Instructions
International Conference on Computational Social Science
Kellogg School of Management at Northwestern University
Thursday, July 12 2018
Workshops Organizer: Laura K. Nelson, Northeastern University
If you have questions about the workshops, email Laura Nelson at l.nelson@northeastern.edu

For social science researchers and data analyst enthusiasts who are new to computational methods or want to add tools to their toolkit, we offer six workshops on a variety of topics such as computational psychometrics, new advances in social network analysis, visual communication, and more.

Below is a description of each workshop, including any pre-work that attendees should have prepared before coming onsite. Before the conference, you should read these descriptions and decide which three workshops you will be attending--one in each session. **To get the most out of these workshops, it is important to bring your own laptop and complete the pre-workshop instructions for each of your three selected workshops. The appropriate software will need to be installed and datasets downloaded before the date of the conference.** Laptops will not be provided to attendees. Information regarding software and datasets for each offered workshop is included in the descriptions below.

Session one: 9:00am-12:00pm

Choose one

Workshop 1a: New Advances in Network Analysis

Instructor: Roger Guimerà, Catalan Institution for Research and Advanced Studies at Universitat Roviral Virgili
Location: Room 1420

Workshop Description:
After decades developing methods to model and characterize social networks, we are now in a position to use rigorous probabilistic approaches to make predictions from network data. For example, using network approaches we can anticipate conflict between team members, or predict whether individuals will cooperate or defect when facing a social dilemma. In this session, we will discuss the fundamentals and learn some Python tools for network inference.

Pre-Workshop Instructions:
You will be using Jupyter notebook, so the following should be installed:

1) Python
2) Ipython/jupyter notebook
3) numpy
4) matplotlib

Additionally, you need to install the graph-tool python module, which is available from:
https://graph-tool.skewed.de/
and can be installed with the following instructions:
https://git.skewed.de/count0/graph-tool/wikis/installation-instructions

The native installation is preferred. If not possible, and you will use Docker installation, you should be sure to know how to run Jupyter notebooks from Docker, as detailed in: https://git.skewed.de/count0/graph-tool/wikis/installation-instructions#jupyter-notebooks

Links to the data and scripts you should download ahead of the workshop can be found:
https://www.dropbox.com/sh/bmtdru43t9zvlrj/AABcMd9ra9HuJWIRD0G85_Pwa?dl=0

Instructor Bio:
Roger Guimerà graduated in Physics at Universitat de Barcelona in 1998, and obtained a PhD in Chemical Engineering from Universitat Rovira i Virgili in 2003. He then moved to Northwestern University, where he worked as a postdoctoral fellow and, later, as a Fulbright Scholar. In 2008 he became a Research Assistant Professor at Northwestern's department of Chemical and Biological Engineering, before accepting his current position at ICREA. Roger's research is devoted to the study of complex systems and, particularly, of the structure of complex networks and the interplay between network structure and dynamics.

Workshop 1b: Computational Psychometrics for Education and Psychological Assessments
Instructors: Alina von Davier, ACTNext by ACT
Pietro Cipresso, Instituto Auxologico Italiano
Location: Room 1430

Workshop Description:
In this workshop, the instructors will use lecture, discussions, and software demos to introduce a new area, Computational Psychometrics (CP; Cipresso, 2015; von Davier, 2015), and the best assessment practices for data logging, data mining (DM), visualization, and machine learning (ML) techniques, as well as methods for evaluating results from the analyses of Big Data, even using virtual reality (VR). The session is designed for researchers with a background in measurement but less experience with data mining or machine learning.

Pre-Workshop Instructions:
Please download WEKA for the workshop - http://www.cs.waikato.ac.nz/ml/weka/downloading.html

- For Mac OS X or Linux users, download and run the corresponding executables
- For Windows users, download the version with Oracle's Java VM (if you don't already have the latest version installed)
    - If you have a new/newer Windows machine, I would recommend weka-3-8-1jre-x64.exe
    - Click "Run" and accept the terms/recommendations
- It's safe to install both Java and Weka, despite the warnings

In addition, please see attached the data set –a subset from the ACT Collaborative Problem Solving (CPS) Study I. We will use it for the demo.

Instructor Bio:
Alina von Davier is the Vice President of ACTNext by ACT, Inc., a Research, Development, and Business Innovation Division, as well as an Adjunct Professor at Fordham University. She earned her PhD in mathematics from the Otto von Guericke University of Magdeburg, Germany, and her MS in mathematics from the University of Bucharest, Romania. At ACT, von Davier and her team of experts are responsible for developing prototypes of research-based solutions and creating a research agenda to support the next generation for learning and assessment systems (LAS). She pioneers the development and application of computational psychometrics and conducts research on blending machine learning algorithms with the psychometric theory. Prior to her employment with ACT, von Davier was a Senior

2211 Campus Drive, Evanston, Illinois 60208 | 847.491.3300 | kellogg.northwestern.edu

Research Director at Educational Testing Service (ETS) where she led the Computational Psychometrics Research Center.

Pietro Cipresso is the senior researcher and Chief Technical Officer at the Applied Technology for Neuro-Psychology Laboratory - ATN-P Lab at Istituto Auxologico Italiano in Milan, Italy, and Assistant Professor in Psychometrics at Catholic University of Milan. Cipresso, who is also Project Manager inside the ATN-P Lab for the Positive Technology area, has coordinated and coinvestigated several National, European and International Projects and has been Visiting Researcher at Massachusetts Institute of Technology (MIT), Cambridge, MA, USA, and at Monash University, Melbourne, Australia. He is the team leader of a forthcoming (2018) EU Project (BodyPass: API-ecosystem for cross-sectorial exchange of 3D personal data).

Session two: 1:00pm-3:00pm
Choose one

Workshop 2a: Now They See it: Visual Communication of the Patterns in your Data
Instructor: Steven Franconeri, Northwestern University
Location: Room 1420

Workshop Description:
Within a well-designed graph or data visualization, the eyes can be a powerful tool for understanding patterns in data. But within a poorly-designed depiction of the same data, the same tasks can be inefficient, or even overwhelming. In this workshop Psychology Professor Steven Franconeri will combine an overview of data visualization techniques with hands-on exercises to illustrate how to clearly present your data to both your research colleagues and to non-technical audiences.

Pre-Workshop Instructions:
In preparation for this workshop, Steve has requested that you submit data visualization(s) that you would like to improve, from your own research (papers, talks, or posters). They can be as simple or complex as you like.

In the workshop, we'll help each other sketch redesigns. If you don't have any available, you could instead submit example(s) from the papers, talks, or posters of others in your field.

Please submit your visualizations to this folder shared below by July 10 at 5pm CST.
https://www.dropbox.com/request/dRG81MZuzpB7vjrMx916

Instructor Bio:
Steven Franconeri is a Professor of Psychology at Northwestern University and Director of the Northwestern Cognitive Science Program. His research is on visual thinking, visual communication, and the psychology of data visualization. He directs the Visual Thinking Laboratory, where a team of researchers explore the power and limits of your visual system, and how better design and pedagogy can help students and scientists understand and use visual representations across paper, screens, and their imagination.

Workshop 2b: Using Web of Science Citation Data for 'science of science' Studies on the Global Research Network
Instructor: Joe Brightbill, Technology Lead, Custom Data
Location: Room 1430

Workshop Description:

2211 Campus Drive, Evanston, Illinois 60208 | 847.491.3300 | kellogg.northwestern.edu

The Web of Science Core Collection is a vast citation network representing the global landscape of science since 1900. For decades, researchers have explored this rich dataset to answer key questions about the nature of scientific discovery and innovation and to empower big data analytics. The Web of Science Core Collection contains over a billion cited references, forming a giant network of interconnections between scientists, their institutions, and their disciplines on a global scale. Big data techniques have opened the analytical possibilities for citation data beyond the traditional bibliometric studies into network analyses that combine citation data with datasets from other disciplines. Citation data are used by a variety of researchers in different fields. We will discuss the characteristics of Web of Science Core Collection data, how it is structured, and common types of analysis conducted.

Pre-Workshop Instructions:
Clarivate Analytics will provide a sample dataset for publications in the Infectious Diseases category in Web of Science, with at least one US address from 1980-2016. We will provide a subset of fields that seem of interest from a sociological perspective (authors, locations, dates, OA status, etc.) for both source records, citing records, and cited records. To get access to the data, you will have to register your interest and acknowledge the data usage terms here: https://www.surveymonkey.com/r/JZTYLG3. The data will be sent to all respondents who express interest and acknowledge the terms on July 9th. If you do not register interest in advance, we can't guarantee we can send the data during the workshop given the size and limited time.

Attendees should have their own laptops. There is no specific pre-work that needs to be done, but it is suggested that you have something to work with the data (MySQL, Python, R, etc.) and have loaded or familiarized yourself with the data in advance given the size. The data will come with a more detailed readme file so that attendees should have enough information to start playing around.

Some examples of research projects that have used similar Web of Science data are http://advances.sciencemag.org/content/advances/3/4/e1601315.full.pdf and http://journals.sagepub.com/doi/10.1177/0306312705052359.

Instructor Bio:
Joe Brightbill is the technology lead for Custom Data projects in the Scientific and Academic Research division of Clarivate Analytics. He has been with the company for seven years, working in detail with bibliometric data from the Web of Science, InCites, Essential Science Indicators, and Journal Citation Reports. He graduated from Drexel University in 2012 with a BS in computer science.

Session three: 3:30pm-6:00pm
Choose one

Workshop 3a: Creating Interactive Visualizations with R Shiny
Instructor: Christina Maimone, Northwestern University Information Technology
Location: Room 1420

Workshop Description:
Shiny is an R package that lets you build interactive web applications that can stand alone on a web page, function as a dashboard, or be incorporated into R Markdown documents. You can share your data and research in engaging ways or create tools to help students explore statistical concepts and data analysis. In this workshop we'll explore what you can do with Shiny, and you'll create your first Shiny application.

Pre-Workshop Instructions:

2211 Campus Drive, Evanston, Illinois 60208 | 847.491.3300 | kellogg.northwestern.edu

You will need to be familiar with R and bring your own laptop. You can either install R, RStudio, and the packages tidyverse, shiny, markdown, DT, and shinythemes on your own laptop, or use RStudio Cloud. Details, instructions, and workshop materials are available at https://github.com/nuitrcs/rshiny.

Instructor Bio:
Christina provides data science support and training to researchers as a part of Northwestern IT's Research Computing Services team. After leading and advising on quantitative and computationally intensive social science research projects for over a decade, she now enjoys helping researchers overcome computational and technical obstacles in their research. While she uses a wide variety of programming languages and tools in her work, R is a favorite; she is the founder of the Northwestern R User Group. Christina has a PhD in Political Science and a MS in Statistics from Stanford University.

Workshop 3b: Building Multi-level Agent-based Models with NetLogo and LevelSpace
Instructors: Bryan Head, Northwestern University
Arthur Hjorth, Northwestern University
Location: Room 1430

Workshop Description:
Multi-level Agent-Based Modeling (ML-ABM) enables modelers to easily expand on models by connecting them to other models. Typical use cases for ML-ABM include modeling interactions between levels by delegating each representational level to each own model, each represented by a model; zooming in on particular event by designing higher spatial or temporal granularity event-specific models; or connecting different types of models like agent based models and systems dynamics models. In this workshop, participants will learn about ML-ABM, and use NetLogo and LevelSpace to build and/or expand on models. We will run the workshop to accommodate a wide range of experience levels. No programming or model experience is necessary.

Pre-Workshop Instructions:
In preparation for the Building multi-level agent-based models with NetLogo and LevelSpace-workshop, please bring a laptop with NetLogo 6.0.4 installed. Also please fill out this form, so we can adapt our materials to your specific research and learning interests.

We believe that learning happens best when participants have hands-on experiences with our modeling tools, and participants will be building and extending multi-level models in this workshop. If you are new to NetLogo, or if you have not used NetLogo in a while, we recommend spending an hour in the company of our programming guide, (re)acquainting yourself with the language.

Instructor Bio:
Bryan Head is a PhD candidate at the Center for Connected Learning and Computer-Based Modeling (CCL). His research focuses on the design and analysis of multi-level agent-based models. His work focuses on combining machine learning with agent-based modeling, and applies cutting edge techniques to both the design and the analysis of agent-based models.

Arthur Hjorth is a postdoctoral fellow at the Center for Connected Learning and Computer-Based Modeling (CCL), and at the Center for Prevention and Implementation Methodology (CePIM) at Northwestern University. In his work at the CCL, Arthur works on the use of ABMs as teaching tools, specifically for helping learners think about social issues as complex systems. In his work at CePIM, Arthur builds research models for HIV prevention at the Feinberg School of Medicine at Northwestern University. He received his PhD in Learning Sciences at Northwestern University.

2211 Campus Drive, Evanston, Illinois 60208 | 847.491.3300 | kellogg.northwestern.edu