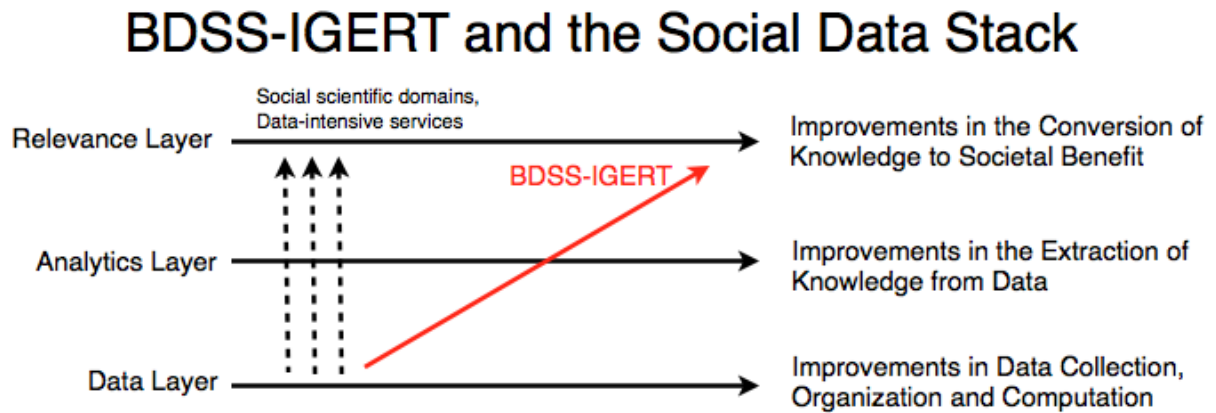


Social Data Analytics Core Seminars, SoDA 501 and SoDA 502

Overview

The core seminars are organized around the metaphor used in BDSS-IGERT of the social data stack. The social data stack consists of three fuzzily bounded layers: the “data layer,” the “analytics layer,” and the “relevance layer” (Figure 1).

Figure 1.



The data layer is comprised of the processes and technologies by which human interactions are translated into data about human interactions. This is material discussed in SoDA 501, offered in the spring semester, and typically taken in the fourth semester of the PhD.

The analytics layer is comprised of the processes and technologies by which social data are translated into knowledge about society. This is material discussed in SoDA 502, offered in the fall semester, and typically taken in the fifth semester of the PhD. (SoDA students must also take a focused seminar in a core approach to analytics: machine learning / statistical learning / data mining, or visual analytics.)

The relevance layer is comprised of the processes and technologies by which knowledge about society is translated into value for science or society. This is addressed in the SoDA seminars through a project orientation, both in evaluation and in the examples discussed with our guests throughout the courses.

Recommended Background for Social Data Analytics Coursework

Students will have a variety of backgrounds. Prior to beginning interdisciplinary coursework to fulfill Social Data Analytics degree requirements, including SoDA 501 and 502, students are expected to have advanced (graduate) training in at least one of the component areas of Social Data Analytics, and a familiarity with basic concepts in the others.

With regard to specialization, students are expected to have advanced (graduate) training in ONE of the following.

- (a) quantitative social science methodology and a discipline of social science (as would be the case for a second-year PhD student in Political Science, Sociology, Criminology, Human Development and Family Studies, or Demography); OR
- (b) statistics (as would be the case for a second-year PhD student in Statistics); OR
- (c) information science or informatics (as would be the case for a second-year PhD student in Information Science & Technology, or a second-year PhD student in Geography specializing in GIScience); OR
- (d) computer science (as would be the case for a second-year PhD student in Computer Science and Engineering).

This requirement is met as a matter of meeting home program requirements for students in the dual-title PhD, but may require additional coursework on the part of students in other programs wishing to pursue the graduate minor.

With regard to general preparation, students are expected to have ALL of the following technical knowledge:

- (a) basic programming skills (e.g., CMPSC 121 or equivalent); AND
- (b) basic knowledge of relational databases and/or geographic information systems (e.g., IST 210 or GEOG 160, or equivalent); AND
- (c) basic knowledge of probability, applied statistics, or social science research design (e.g., STAT 200, PLSC 309, SOC 207, or equivalent); AND
- (d) basic familiarity with a substantive or theoretical area of social science (e.g., 300-level coursework in political science, sociology, criminology, human development, psychology, economics, communication, anthropology, human geography, social informatics, or similar fields).

It is not unusual for students in to have one or more gaps in this preparation. Students must work with Social Data Analytics advisers to develop a plan for timely remediation of any deficiencies, which generally will not require formal coursework for students whose training and interests are otherwise appropriate for pursuit of the Social Data Analytics degree. Where possible this will be addressed at time of application to the Social Data Analytics program.

SODA 501, Big Social Data: Approaches & Issues

This seminar addresses the interdisciplinary integration of computational, informational, statistical, visual analytic and social scientific approaches to the creation of data that are both "social" (about, or arising from, human interactions) and "big" (of sufficient scale, variety, or complexity to strain the informational, computational, or cognitive limits of conventional social scientific approaches to data collection or analysis).

Examples include text, image, audio, video, intensive spatial &/or longitudinal data, data with complex network, hierarchical &/or other relational information, data from distributed sensors and mobile devices, digitized archival data, and data exhaust from sources like social media. Issues include sources of social data, data structures and formats for social data, data collection and manipulation technologies, data provenance, data linkage and alignment, ethics and scientific responsibility in human subjects research, experimental and observational data collection design for causal inference, sampling design, measurement of latent social concepts, reliability and validity, search and information retrieval, nonrelational and distributed databases, and standards for data preservation and sharing.

Half of each week (1-1/2 hours per week) is allotted to discussion with a guest member of the Graduate Faculty in Social Data Analytics, drawn from the full range of participating disciplines, about an active research project that touches on one or more areas of concern in the course. Half of each week (1-1/2 hours per week) is allotted to discussion of readings / assignments focused on exposure to, and integration of, alternative disciplinary perspectives on a particular topic within the data layer.

The primary objective of the seminar is substantive interdisciplinary engagement with and integration of the tools, practices, language, and standards used in the collection and management of data in the component disciplines of the Social Data Analytics field.

SODA 502, Social Data Analytics: Approaches & Issues

This seminar addresses the interdisciplinary integration of computational, informational, statistical, visual analytic and social scientific approaches to learning from data that are both "social" (about, or arising from, human interactions) and "big" (of sufficient scale, variety, or complexity to strain the informational, computational, or cognitive limits of conventional social scientific approaches to data collection or analysis).

Topics include alternative scientific models for learning from data (Bayesian inference, causal inference, statistical / machine learning, visual analytics, measurement modeling), analytics issues with big data (variable selection, parallel computing, algorithmic scaling, ensemble modeling, validation), analytics issues with particular structures and channels of social data (network data, geospatial data, intensive longitudinal data, text data), and issues of scientific responsibility and ethics in big social data.

Half of each week (1-1/2 hours per week) is allotted to discussion with a guest member of the Graduate Faculty in Social Data Analytics, drawn from the full range of participating disciplines, about an active research project that touches on one or more areas of concern in the course. Half of each week (1-1/2 hours per week) is allotted to discussion of readings / assignments focused on exposure to, and integration of, alternative disciplinary perspectives on a particular topic within the data layer.

The primary objective of the seminar is substantive interdisciplinary engagement with and integration of the tools, practices, language, and standards used to learn from data in the component disciplines of the Social Data Analytics field.