General Information

School of Education EDUC 390X
Computer Science CS 424M
Winter 2012: Thursdays, 2:15 - 5:05pm
Room: TBD

Teaching team

Paulo Blikstein
Assistant Professor
520 Galvez Mall - CERAS 232
Tel. (650) 736-0966
paulob@stanford.edu

Marcelo Worsley
Ph.D. candidate
520 Galvez Mall - CERAS 102
Tel. (650) 924-2672
mworsley@stanford.edu

Teaching assistant

TA: TBD
Office hours: TBD

Course Web Site & email lists

Class: http://learninganalytics.stanford.edu
Blog: http://learninganalyticsblogs.stanford.edu
Email list for course instructors: lacmfaculty@lists.stanford.edu
Email list for all course members (students and faculty): lacm2013@lists.stanford.edu
Course Abstract

Traditional analytical techniques are insufficient to investigate and predict complex dynamic social phenomena such as classroom and online interactions, social networks, online gaming, diffusion of innovation, opinion dynamics, and other complex adaptive systems. Data mining, machine learning, and computational modeling are dramatically changing the ways scholars conduct research in these fields. In this course, we will learn about how agent-based modeling, network theory, and basic learning analytics/data-mining can support novel research in the cognitive and social sciences, with a particular focus on learning (online and offline), cognitive development, and educational policy.

Course Description

Over the last two decades, the use of computer modeling, machine learning, and large-scale data-mining in the natural sciences has exploded. This trend has recently reached the social and behavioral sciences as well, and the exponential growth of data sources and modalities has deeply transformed the work of natural and social scientists. As a result, many traditional analysis techniques have become insufficient for a wide range of research questions – especially when they require the understanding of many elements interacting in a multidimensional space, or billions of data points. Furthermore, as social interaction and learning start to take place in several different settings and media, and classrooms start transitioning from tens of students to hundreds of thousands, these new methodological tools are becoming a crucial part of the researcher’s toolbox.

For example, without computational tools, collecting, modeling, and analyzing data from social networks, online games, or massive online open courses would be unfeasible. Similarly, much of the contemporary research on mobile device usage, market dynamics, diffusion of innovation, human and business networks, consumer behavior, and opinion dynamics, to mention a few, would have been impossible without the use of computational methods and tools.

The goal of this course is to introduce students to these new methods and techniques. We will explore computational methods and data-mining techniques to support research in the learning, cognitive,
and social sciences, with a particular focus on issues of (offline and online) learning, cognitive development, and educational policy.

The course content and assignments will be centered on two primary methodological tools. One primary tool will be multi-agent computer modeling (also called agent-based modeling, or ABM). This modeling technique helps explain systems where a large number of independent entities simultaneously interact with each other, without central control, generating non-linear behaviors, adaptation, and learning, and resulting in emergent macroscopic patterns for the larger system. In addition to agent-based modeling, we will also focus on constructing models using techniques from Learning Analytics. Some of the specific techniques include web-mining, text analytics and natural language processing - which allows for the automated analysis of large corpora of verbal data, such as transcriptions. Additionally, we will also explore network analysis and bio-sensing (galvanic skin response, heart rate variability, eye tracking), as these offer additional approaches for both gathering and studying information from complex systems.

The combination of computational modeling and learning analytics will allow students to bring real-world data into their models, explore results using a variety of paradigms, and ultimately enhance understanding of cognition, learning, and other social phenomena.

Multi-agent Modeling. Methods such as agent-based modeling are part of the emergent field of the “complexity sciences.” If you have heard of a social phenomenon that exhibits a “biological behavior,” an idea that is “spreading virally,” a stock market in panic, or just watched flocking birds or an ant nest, you already know a lot about complexity sciences. The complexity sciences use novel investigation methods and frameworks to understand natural and social systems which cannot be described well with conventional models and formulas. Those systems are better understood starting from the simple behaviors of its microelements (people, schools, businesses, or birds), instead of just examining the overall macroscopic system (a society, a classroom, a market, or a flock). Typical of complex phenomena is that the aggregate behaviors at the macroscopic level are not premeditated by any of the micro behaviors. For example, flocking birds do not intend to construct an arrow-shaped structure,
and stockbrokers do not want to crash the market. Rather, each element (or “agent”) follows its local rules (following another bird, for example, or trying to get the better price for a stock), and the overall pattern emerges from these multiple local behaviors.

**Learning Analytics.** Educational Data-Mining and Learning Analytics comprise a suite of techniques in data science that are concerned with the intersection of the Learning Sciences and data driven analytics. At a high level, learning analytics enables researchers to use staple qualitative research data, to closely study participant behavior, but in a manner that more easily scales to large populations. To this end, Learning Analytics is concerned with measuring, analyzing and interpreting data from complex data sources, like images, text, speech, collaboration, server logs, interaction logs and more.

The course will have five main of activities:

- **Lectures and class discussions:** here we will get in touch with the theoretical foundations in the complexity sciences and learning analytics, as well as relevant literature in the cognitive and learning sciences.
- **Tool review:** we will review existing tools for computer modeling, data-mining, network analysis, and data visualization, and learn the cutting-edge of what’s “out there”.
- **Invited speakers** from universities and industry will tell students about the state-of-the-art of the field.
- **Mini-tutorials:** since this course is for both novices and experts, we will have mini-tutorials to get everyone up to speed on the tools, software packages and programming languages we will use.
- **Project work:** students will have an immersive, hands-on experience in creating a computer model, preferably by bringing their own data or research questions to the class.

**I can’t program. Should I enroll? (“YES!”)**

No previous programming or technology background is assumed. Even if you have never programmed in your life, you can still enroll! This class is a good fit for people both with and without a technical background, and students who had never written code created wonderful projects.
Summary Schedule

<table>
<thead>
<tr>
<th>Class</th>
<th>Topic</th>
</tr>
</thead>
<tbody>
<tr>
<td>Class 1</td>
<td>Introduction, agent-based modeling, learning analytics, complex systems</td>
</tr>
<tr>
<td>Class 2</td>
<td>Complex systems applied to the learning sciences, case studies</td>
</tr>
<tr>
<td>Class 3</td>
<td>Participatory simulations, collecting data from large groups</td>
</tr>
<tr>
<td>Class 4</td>
<td>Verification and validation of computer models in the social sciences</td>
</tr>
<tr>
<td>Class 5</td>
<td>Learning analytics, theoretical grounding, data mining</td>
</tr>
<tr>
<td>Class 6</td>
<td>Web data mining, Python programming</td>
</tr>
<tr>
<td>Class 7</td>
<td>Text Analytics and natural language processing</td>
</tr>
<tr>
<td>Class 8</td>
<td>Biosensing</td>
</tr>
<tr>
<td>Class 9</td>
<td>Networks and network analysis</td>
</tr>
<tr>
<td>Class 10</td>
<td>Final presentations</td>
</tr>
</tbody>
</table>

Detailed schedule and readings

There will be no printed course reader. All materials will be online as PDFs. As a result, we will save one third of a tree over the ten weeks, and a lot of electrical energy. This will make us a bit less guilty for using so much energy to run lots of simulations for many hours.

Class 1 - Intro to Learning Analytics, Complex Systems, Computational Modeling

- Overview of the course & logistics
- What is learning analytics? Why use it?
- What is a complex system? What is emergence? What is a model? Why are models important?
- Chaos, emergence, order, self-organization.

Modeling and simulation in the social and cognitive science: examples of current research.

Class activity: Introduction to NetLogo: creating your first agent-based model.

Readings:


Class 2 – Complex systems applied to the learning sciences, case studies

Examining your own research questions through computational modeling: how to think like a modeler, and why you should care – a new lingua franca for research.

Stationary and movable agents; interactions between agents; agent topologies, artificial life, comparison with systems dynamics models.

Readings:


---

**Class 3 – Participatory simulations, collecting data from large groups**

- Participatory simulations, real time emergence in social systems.
- Collecting data from large groups, experiments.
- Combining human and virtual agents

**Class activity:** creating your first participatory simulation (in groups).

**Readings:**


Class 4 – Agent-Based Modeling, Verification, replication, and validation.

- Methods for verification and validation of models
- Model replication
- Systematic exploration of model parameter space, sensitivity analysis
- **Class activity:** design a verification scheme for your model.
- **Readings:**
Class 5 – Learning Analytics, theoretical grounding, data mining

- Presentation of mid-term projects
- Introduction to Learning Analytics
- Modalities for analysis
- Basic concepts in machine learning
- Developing theoretically-grounded data driven analyses
- **Class activity:** Operationalizing a theory into a learning analytics study
- **Readings:**

Classes 6 – Web data mining, Python Programming

- Web data mining
- Python data structures and syntax
- Basic file reading and manipulation
- Tools for mining web data: Beautiful Soup, Rapid Miner
- **Class Activity:** Mine data from a NetLogo log or the Web
- **Readings:**


- (optional) Open Dapper Tutorial Video

### Classes 7 – Text Analytics and natural language processing

- Introduction to Statistical Natural Language Processing
- Automatically counting words
- Natural Language Toolkit (NLTK)
- Sentiment Analysis
- Paraphrase Detection

- **Class Activity:** Sentiment Analysis Activity

- **Readings:**

  - Selections from: Manning, C. and Schütze, H. 1999. Foundations of Statistical Natural Language Processing. MIT
  

Classes 8 – Biosensing

- Introduction to biosensing
- Visualizing and analyzing data
- Incorporating biosensor data into computational models
- **Class Activity:** Collect and depict biosensing data
- **Readings:**

Class 9 – Networks and Network analysis

- Networks, properties of networks, network metrics, agents in networks.
- The small-world problem
- **Class activity:** exploring social network data, adding social network data to your agent-based models.
- **Readings:**

Class 10 – Final presentations

Assignments

- **Due on Week 2**: Your first agent-based model
- **Due on Week 3**: A short case study of a learning problem using computer modeling
- **Due on Week 4**: Participatory simulation project (in groups)
- **Due on Week 5**: Mid-term project: your own agent-based model
- **Due on Week 6**: Introduction to Python activity
- **Due on Week 8**: Log or web mining activity
- **Due on Week 9**: Collect (fun) bio-sensing data
- **Due on Week 10**: Final project

Final projects

The final project is a big part of the course, and consists of a model, a set of results, and a paper. You can choose a theme close to your own research as a final project topic, and thus use data and literature that you already have. There are five main possibilities for final projects:

(a) Build agent-based models and/or use learning analytics to examine behaviors of learners, whether individuals or cohorts, e.g., to model patterns of classroom/online collaboration, teamwork, individual cognition, self-efficacy, motivation, “peer-pressure,” etc.

(b) Create simulations or data analysis schemes that foreground and problematize central tenets of education-research theoretical models, e.g., pitting Piagetian and Vygotskian perspectives, or issues in other fields, such as social sciences, linguistics, political science, etc. You can also create simulations that foreground and problematize educational policy research questions, such as school choice, vouchers, achievement gap, school funding, etc.

(c) Use learning analytics to examine a corpus of text, web, log, or sensory data from a research
project that you are already developing.

(d) Develop creative combinations thereof.

The timeline for the final project is:

- **Due on Week 4:** One-page idea
- **Due on Week 5:** Revised one-page idea
- **Due on Week 6:** Individual meetings with instructors
- **Due on Week 7:** Final project proposal
- **Due on Week 10:** Presentation
- **Due on Week 11:** Final paper

**Software/hardware platforms used**

- **NetLogo:** a multi-agent version of Logo, this language is tuned for constructing models of dynamic systems. It is useful for creating models of ecological systems, chemical systems, economic trade, social behavior, etc.
- **GoGo Board** (interface for physical sensors).
- **Natural Language Toolkit** (text-mining, text-parsing)
- **Sonia, NetworkX, Pajek, CFinder** (network visualization)
- **RapidMiner, Weka** (statistical analysis, machine learning)
- **R** (statistical analysis).

**Summary of Requirements**

This course is designed to be somewhere between a class and a working group. We will work together to make sense of readings, and, for most of the class projects, you will be working in small groups. The requirements for everyone are:

- Set up your personal blog for the course.
- Keep up with the readings and participate in class, both in person and virtually. You will be expected to post a short comment on each week’s reading by Wednesday at 12pm.
☑ Complete and present (mostly group) assignments.

☑ Design and implement your final project, and give a presentation during the last week of the course.

In addition, due to the group project nature of the class, you are asked to email us in advance if you cannot make a particular class meeting, and are responsible for communicating with your project partners.

**Grading**

All assignments will be graded as either incomplete, complete or outstanding. If an assignment is judged incomplete, you will have an opportunity to complete it or redo it the following week. You will also be assessed on your class participation both in class and online.

Overall, the final project will be worth 50% of the final grade, assignments will be worth 25%, and 25% will be given for class participation. A specific rubric and the evaluation standards for the final project will be distributed during the course.

**Additional foundational (& fun) literature (optional)**


