
Katy Börner
Cyberinfrastructure for Network Science Center, Director
Information Visualization Laboratory, Director
School of Library and Information Science
Indiana University, Bloomington, IN
katy@indiana.edu

With special thanks to the members of the Cyberinfrastructure for Network Science Center

Keynote for NSF’s CS10K initiative to support high school computer science teachers around the country who are teaching or preparing to teach Computer Science Principles (CSP).

December 6, 2013
MOOCs

In 2012, Google hosted three massive open online courses (MOOCs) collectively reaching over 400,000 registrants.

By the end of 2013 more than 250 courses will be run using the Google, Coursera, Udacity, EdX, and other platforms.

http://www.youtube.com/watch?feature=player_embedded&v=eW3qM6qcZQc#at=128
Since Spring 2013, more than 2000 students from 100+ countries registered via http://ivmooc.cns.iu.edu/2013

The course will re-start with new materials on January 27, 2014 and registration will soon open at http://ivmooc.cns.iu.edu

Take IVMOOC for IU credits, see Data Science Certificate http://bit.ly/1fBolkZ
Instructors

**Katy Börner – Theory Parts**
Instructor, Professor at SLIS

**David E. Polley – Hands-on Parts**
CNS Staff, Research Assistant with MIS/MLS
Teaches & Tests Sci2 Tool

**Scott B. Weingart – Client Work**
Assistant Instructor, SLIS PhD student
Course Schedule

Course started on January 22, 2013

- Session 1 – Workflow design and visualization framework
- Session 2 – “When:” Temporal Data
- Session 3 – “Where:” Geospatial Data
- Session 4 – “What:” Topical Data

Mid-Term

Students work in teams with clients.

- Session 5 – “With Whom:” Trees
- Session 6 – “With Whom:” Networks
- Session 7 – Dynamic Visualizations and Deployment

Final Exam

Grading

All students are asked to create a personal profile to support working in teams.

Final grade is based on Midterm (30%), Final (40%), Client Project (30%).

- Weekly self-assessments are not graded.
- Homework is graded automatically.
- Midterm and Final test materials from theory and hands-on sessions are graded automatically.
- Client work is peer-reviewed via online forum.

All students that receive more than 80% of all available points get an official certificate/badge.
Scholarly Database at Indiana University

http://sdb.wiki.cns.iu.edu

Supports federated search of 26 million publication, patent, clinical trials, and grant records. Results can be downloaded as data dump and (evolving) co-author, paper-citation networks.

Register for free access at http://sdb.cns.iu.edu

OSGi & Cyberinfrastructure Shell (CIShell)

- CIShell ([http://cishell.org](http://cishell.org)) is an open source software specification for the integration and utilization of datasets, algorithms, and tools
- It extends the Open Services Gateway Initiative (OSGi) ([http://osgi.org](http://osgi.org)), a standardized, modularized service platform
- CIShell provides “sockets” into which algorithms, tools, and datasets can be plugged using a wizard-driven process

Unit Structure

The course and each unit has three components:

**Theory:** Videos and Slides  
Self-Assessment (not graded)

**Hands-on:** Videos and Slides & Wiki pages with workflows  
Homework (not graded)

**Client Work:** Using Drupal Forum (graded)
Theory Unit Structure

Each theory unit comprises:
• Examples of best visualizations
• Visualization goals
• Key terminology
• General visualization types and their names

• Workflow design
  – Read data
  – Analyze
  – Visualize

• Discussion of specific algorithms

Different Question Types

Terabytes of data

Find your way

Identify trends

Find collaborators, friends

Descriptive & Predictive Models
Different Levels of Abstraction/Analysis

<table>
<thead>
<tr>
<th>Type of Analysis</th>
<th>Micro/Individual (1-100 records)</th>
<th>Meso/Local (101-10,000 records)</th>
<th>Macro/Global (10,000 &lt; records)</th>
</tr>
</thead>
<tbody>
<tr>
<td><strong>Statistical Analysis/Profiling</strong></td>
<td>Individual person and their expertise profiles</td>
<td>Larger labs, centers, universities, research domains, or states</td>
<td>All of NSF, all of USA, all of science.</td>
</tr>
<tr>
<td><strong>Temporal Analysis (When)</strong></td>
<td>Funding portfolio of one individual</td>
<td>Mapping topic bursts in 20 years of PNAS</td>
<td>113 years of physics research</td>
</tr>
<tr>
<td><strong>Geospatial Analysis (Where)</strong></td>
<td>Career trajectory of one individual</td>
<td>Mapping a state’s intellectual landscape</td>
<td>PNAS publications</td>
</tr>
<tr>
<td><strong>Topical Analysis (What)</strong></td>
<td>Base knowledge from which one grant draws.</td>
<td>Knowledge flows in chemistry research</td>
<td>VxOrd/Topic maps of NIH funding</td>
</tr>
<tr>
<td><strong>Network Analysis (With Whom?)</strong></td>
<td>NSF Co-PI network of one individual</td>
<td>Co-author network</td>
<td>NIH’s core competency</td>
</tr>
</tbody>
</table>
## Type of Analysis vs. Level of Analysis

<table>
<thead>
<tr>
<th></th>
<th>Micro/Individual (1-100 records)</th>
<th>Meso/Local (101–10,000 records)</th>
<th>Macro/Global (10,000 &lt; records)</th>
</tr>
</thead>
<tbody>
<tr>
<td><strong>Statistical Analysis/Profiling</strong></td>
<td>Individual person and their expertise profiles</td>
<td>Larger labs, centers, universities, research domains, or states</td>
<td>All of NSF, all of USA, and all of science</td>
</tr>
<tr>
<td><strong>Temporal Analysis (When)</strong></td>
<td>Funding portfolio of one individual</td>
<td>Geo-spatial analysis of PNAS research bursts</td>
<td>113 years of PNAS research</td>
</tr>
<tr>
<td><strong>Geospatial Analysis (Where)</strong></td>
<td>Career trajectory of one individual</td>
<td>Mapping an intellectual landscape</td>
<td>Knowledge flows in chemistry research</td>
</tr>
<tr>
<td><strong>Topical Analysis (What)</strong></td>
<td>NSF-funded network of publications</td>
<td>PNAS publications</td>
<td>NIH’s core competency</td>
</tr>
<tr>
<td><strong>Network Analysis (With Whom?)</strong></td>
<td>NSF Co-PI network of one individual</td>
<td>Co-author network of one individual</td>
<td>NIH’s core competency</td>
</tr>
</tbody>
</table>

### Individual Co-PI Network

*Ke & Börner, (2006)*
Mapping Indiana’s Intellectual Space

Identify

- Pockets of innovation
- Pathways from ideas to products
- Interplay of industry and academia

Mapping the Evolution of Co-Authorship Networks

Compare R01 investigator based funding with TTURC Center awards in terms of number of publications and evolving co-author networks.

Zoss & Börner, forthcoming.

Supported by NIH/NCI Contract HHSN261200800812
### Type of Analysis vs. Level of Analysis

<table>
<thead>
<tr>
<th>Statistical Analysis/Profiling</th>
<th>Micro/Individual (1-100 records)</th>
<th>Meso/Local (101–10,000 records)</th>
<th>Macro/Global (10,000 &lt; records)</th>
</tr>
</thead>
<tbody>
<tr>
<td>Temporal Analysis (When)</td>
<td>Individual person and their expertise profiles</td>
<td>Larger labs, centers, universities, research domains, or states</td>
<td>All of NSF of science, all L labs, all L centers, all L universities, all L research domains, or all L states</td>
</tr>
<tr>
<td>Geospatial Analysis (Where)</td>
<td>Funding portfolio of one individual</td>
<td></td>
<td></td>
</tr>
<tr>
<td>Topical Analysis (What)</td>
<td>Career trajectory of one individual</td>
<td>Knowledge flows in chemistry research</td>
<td></td>
</tr>
<tr>
<td>Network Analysis (With Whom?)</td>
<td>NSF co-PI network of one individual</td>
<td>NIH's core competency</td>
<td></td>
</tr>
</tbody>
</table>
Needs-Driven Workflow Design

Types and levels of analysis determine data, algorithms & parameters, and deployment

Validation
Interpretation

DEPLOY

Visually encode data
Overlay data
Select visualiz. type

Graphic Variable Types
Modify reference system, add records & links
Visualization Types (reference systems)

Stakeholders

Data
READ ANALYZE

VISUALIZE

Types and levels of analysis determine data, algorithms & parameters, and deployment

Needs-Driven Workflow Design
Needs-Driven Workflow Design

Stakeholders

Types and levels of analysis determine data, algorithms & parameters, and deployment

Data

READ

ANALYZE

VISUALIZE

DEPLOY

Visually encode data

Overlay data

Select visualiz. type

Validation

Interpretation

READ ANALYZE VISUALIZE DEPLOY

Validation

Interpretation

READ ANALYZE VISUALIZE DEPLOY

Visually encode data

Overlay data

Select visualiz. type

Types and levels of analysis determine data, algorithms & parameters, and deployment

Stakeholders
## Visualization Types vs. Data Overlays

<table>
<thead>
<tr>
<th>Visualization Type</th>
<th>Chart</th>
<th>Table</th>
<th>Graph</th>
<th>Geospatial Map</th>
<th>Network Graph</th>
</tr>
</thead>
<tbody>
<tr>
<td>Modify / <strong>visually encode</strong> base map.</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Place and <strong>visually encode</strong> records/nodes.</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Place and <strong>visually encode</strong> links.</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
</tbody>
</table>

Plus, add a title, labels, legend, explanatory text, and author info.
Course Schedule

Course started on January 22, 2013
- Session 1 – Workflow design and visualization framework
- Session 2 – “When:” Temporal Data
- Session 3 – “Where:” Geospatial Data
- Session 4 – “What:” Topical Data

Mid-Term

Students work in teams with clients.
- Session 5 – “With Whom:” Trees
- Session 6 – “With Whom:” Networks
- Session 7 – Dynamic Visualizations and Deployment

Final Exam

Dynamic Visualizations and Deployment

http://www.youtube.com/watch?feature=player_embedded&v=m_TwZxZrkg
Clients

http://ivmooc.cns.iu.edu/ivmooc_clientprojects.html

Diogo Carmo
National Science Foundation Awards to Institutions Conducting "Global Warming" Research from 1979-2009

Legend
- Solid Square: $10,000-99,999
- Solid Circle: $100,000-149,999
- Solid Diamond: $150,000-199,999
- Solid Triangle: $200,000-249,999
- Solid X: $250,000-299,999
- Solid Star: $300,000-349,999
- Solid Octagon: $350,000-399,999
- Solid Pentagon: $400,000-449,999
- Solid Hexagon: $450,000-499,999
- Solid Circle: $500,000-599,999

How to Read this Map
This map shows NSF Awards by name and amount. Each unique award is represented as a label located at the middle of the award amount. The inner color of the label shows the award amount in thousands of 2005 dollars. The outer color of the label represents the name and amount of the award. The inner and outer color values are given in the legend.

Data retrieved from Scholarly Database (http://findb.csx.io). 
Choropleth generated by Sandra Chung (2013) using "So2." 
Recently, a number of high school students took the IVMOOC.

We are interested to discuss in how far IVMOOC learning modules could be integrated into the high school curriculum.

Visualizing IVMOOC Data

- Empowering teachers: How to make sense of the activities of thousands of students? How to guide them?
- Empowering administrators: What courses have the highest success rates are most profitable, etc.?
- Supporting students: How to navigate learning materials and develop successful learning collaborations across disciplines and time zones?
- Informing platform designers: What technology helps and what hurts?
- Conducting research: What teaching and learning works online?
Visualizing IVMOOC Data

Data was collected from different sources:
- 1,901 students registered via GCB (1215 male/557 female)
- 52,557 slide downloads from our server
- 18,893 video views via YouTube
- 193 accounts made 730 tweets
- 134 students took 183 exams in GCB
- 674 remarks on 215 different forum threads in Drupal
- 64 students submitted projects via Drupal
1215 male students
557 female students
1215 male students
557 female students
Visualizing IVMOOC Data

- **Empowering teachers**: How to make sense of the activities of thousands of students? How to guide them?
- **Empowering administrators**: What courses have the highest success rates are most profitable, etc.?
- **Supporting students**: How to navigate learning materials and develop successful learning collaborations across disciplines and time zones?
- **Informing platform designers**: What technology helps and what hurts?
- **Conducting research**: What teaching and learning works online?
What questions would you have when learning/teaching online?

References


All papers, maps, tools, talks, press are linked from http://cns.iu.edu
These slides are at http://cns.iu.edu/docs/presentations/2013-borner-visualinsights-cs10k.pdf
CNS Facebook: http://www.facebook.com/cnscenter
Mapping Science Exhibit Facebook: http://www.facebook.com/mappingscience