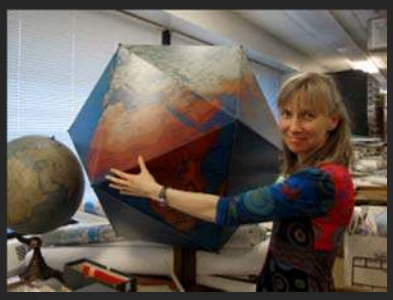
Visual Insights: A Practical Guide to Making Sense of Data

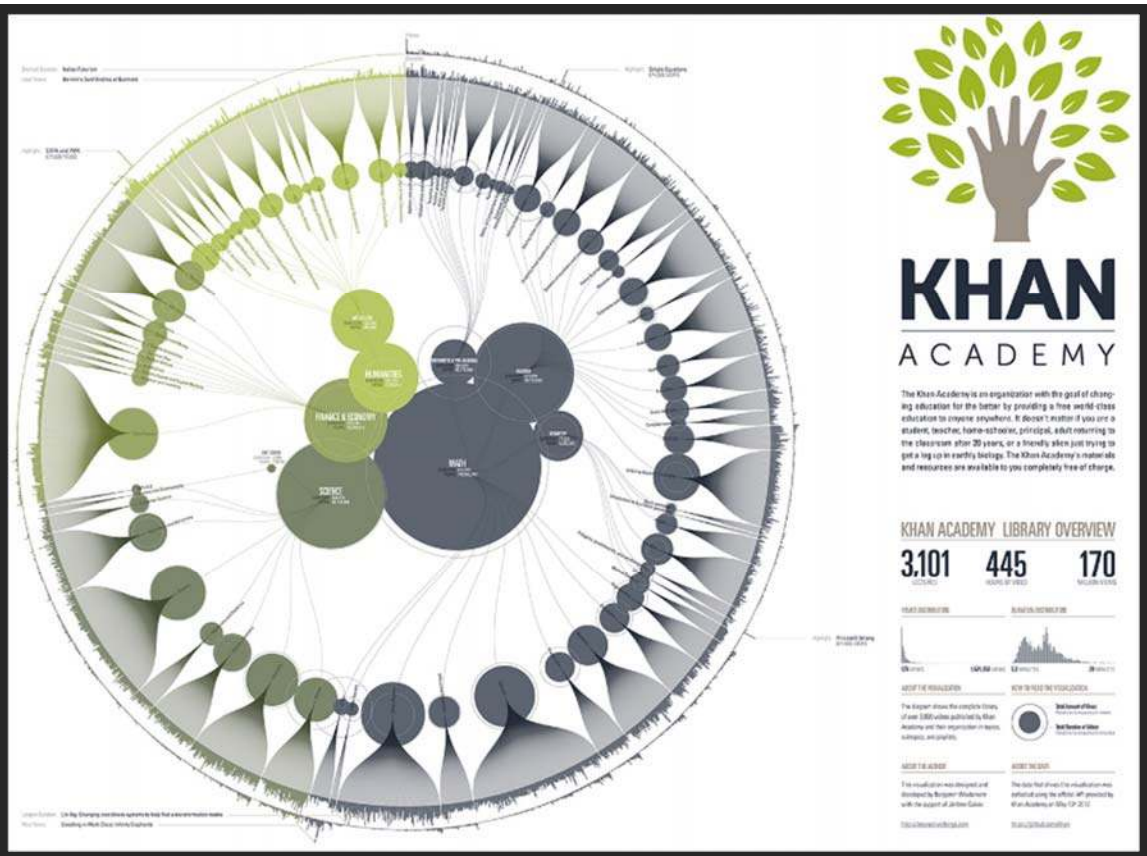


Katy Börner
Cyberinfrastructure for Network Science Center, Director
Information Visualization Laboratory, Director
School of Library and Information Science
Indiana University, Bloomington, IN
katy@indiana.edu

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

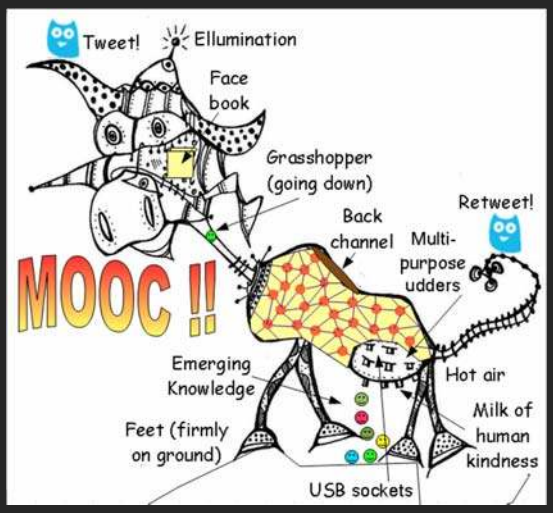


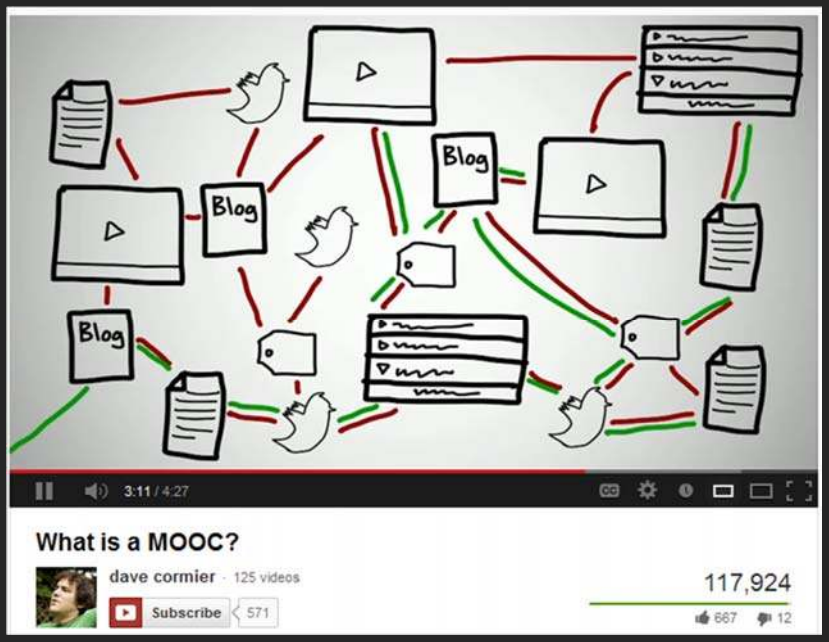
http://scimaps.org/maps/map/khan_academy_library_147 by Benjamin Wiederkehr & Jérôme Cukier

MOOCs

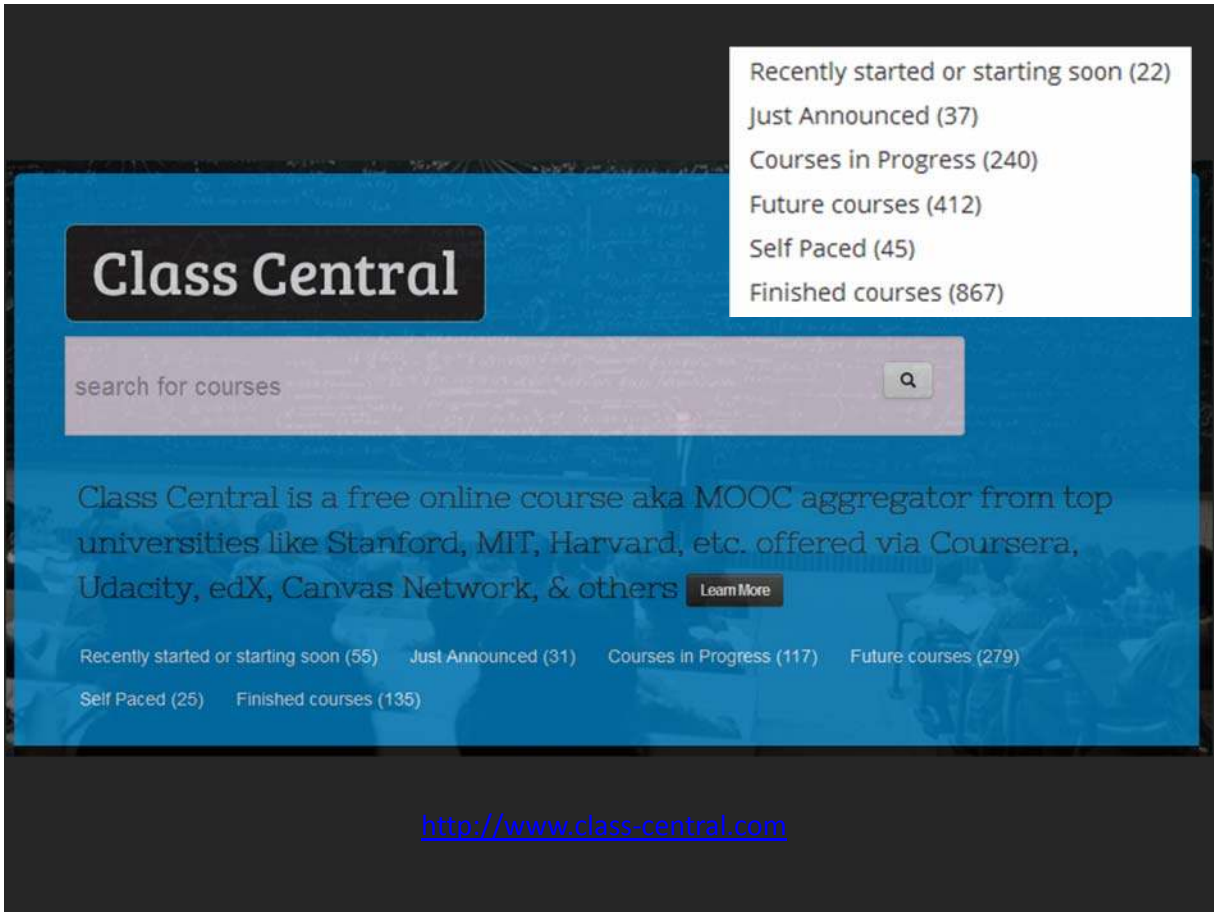
In 2012, Google hosted three massive open online courses (MOOCs) collectively reaching over 400,000 registrants.

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

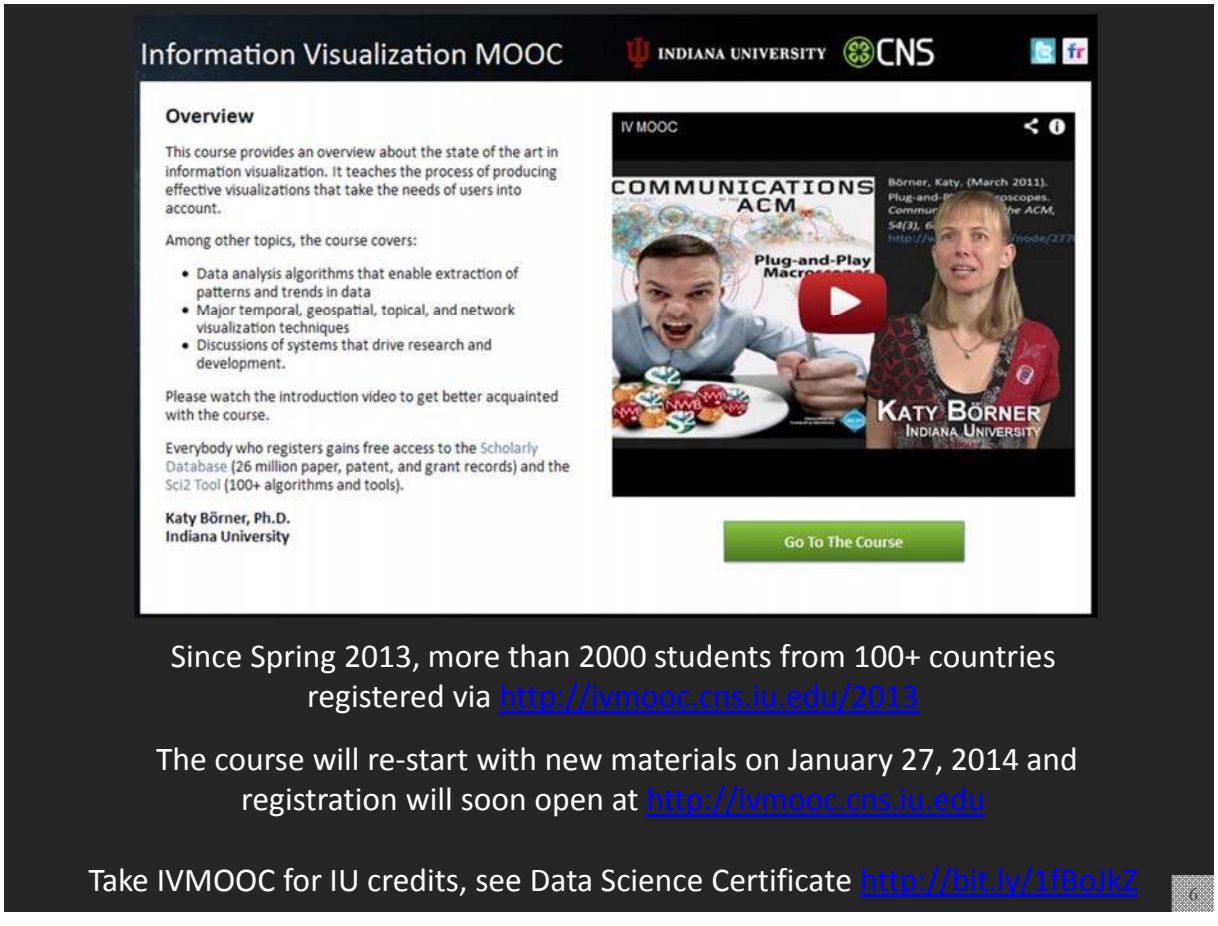




http://www.youtube.com/watch?feature=player_embedded&v=eW3gMGocZQc#at=128



http://www.class-central.com

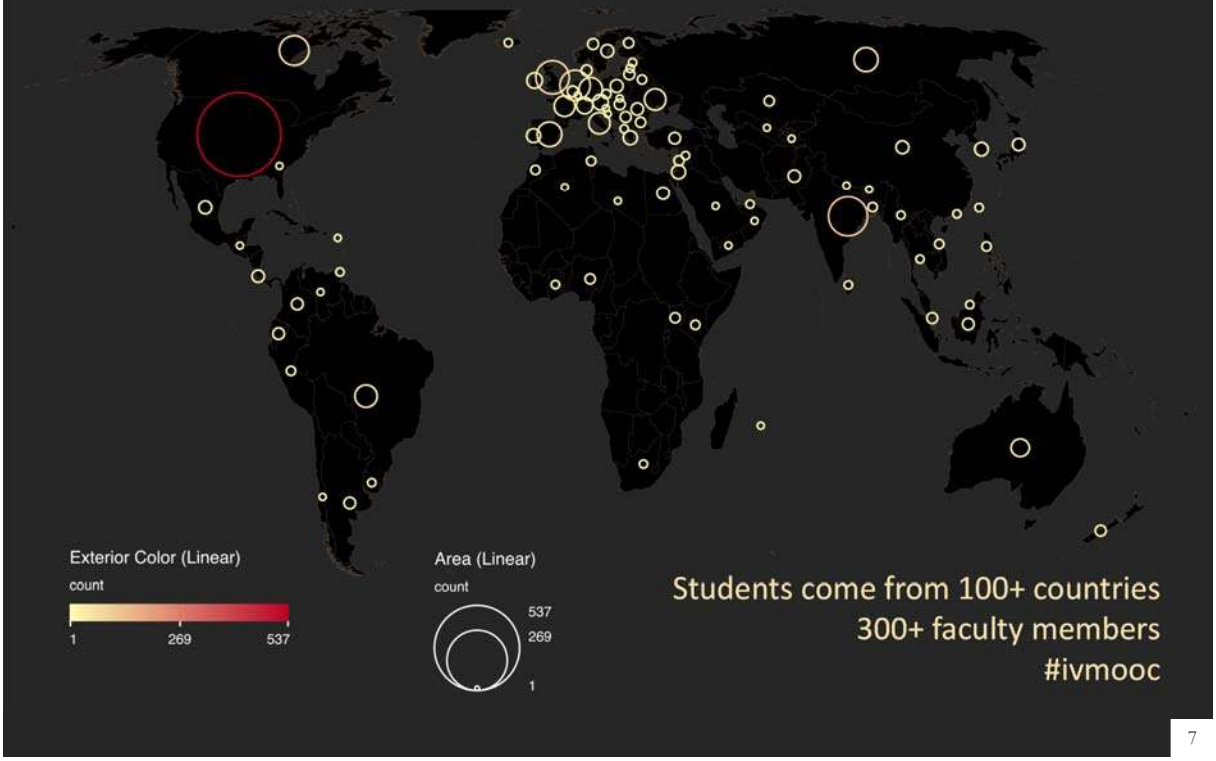


Since Spring 2013, more than 2000 students from 100+ countries registered via http://ivmooc.cns.iu.edu/2013

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/1fBoJkZ

The Information Visualization MOOC
ivmooc.cns.iu.edu



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%**).

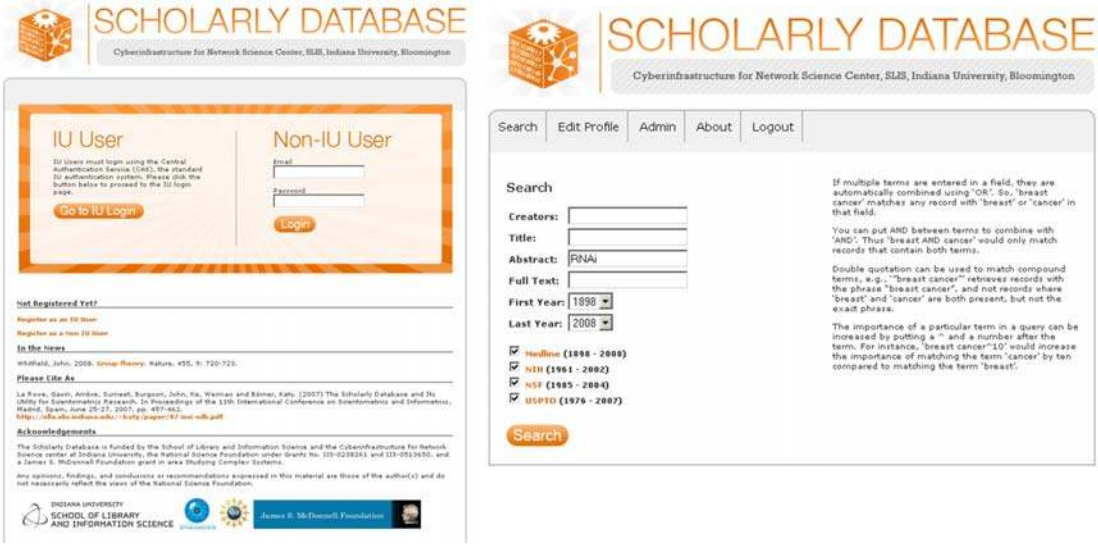
- Weekly self-assessments are not graded.
- Homework is graded automatically.
- Midterm and Final test materials from theory and hands-on sessions are graded automatically.
- Client work is peer-reviewed via online forum.

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.



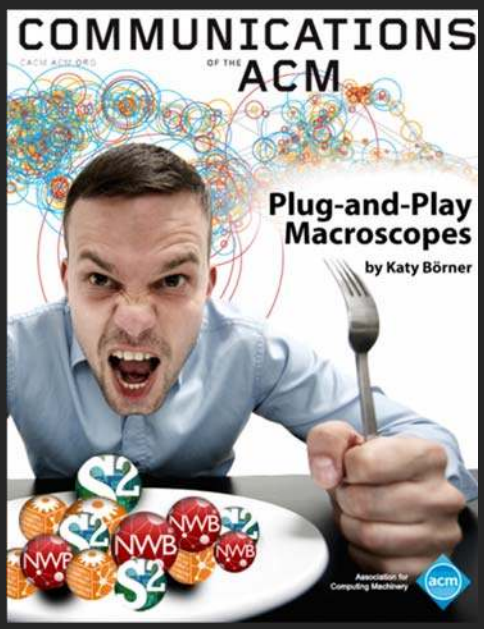
Register for free access at http://sdb.cns.iu.edu

Plug-and-Play Macroscopes
cishell.org

NetworkWorkbench
A Workbench for Network Scientists
cyberinfrastructure for NETWORK SCIENCE CENTER
CIShell Powered

Sci² Tool
A Tool for Science of Science Research & Practice
cyberinfrastructure for NETWORK SCIENCE CENTER
CIShell Powered

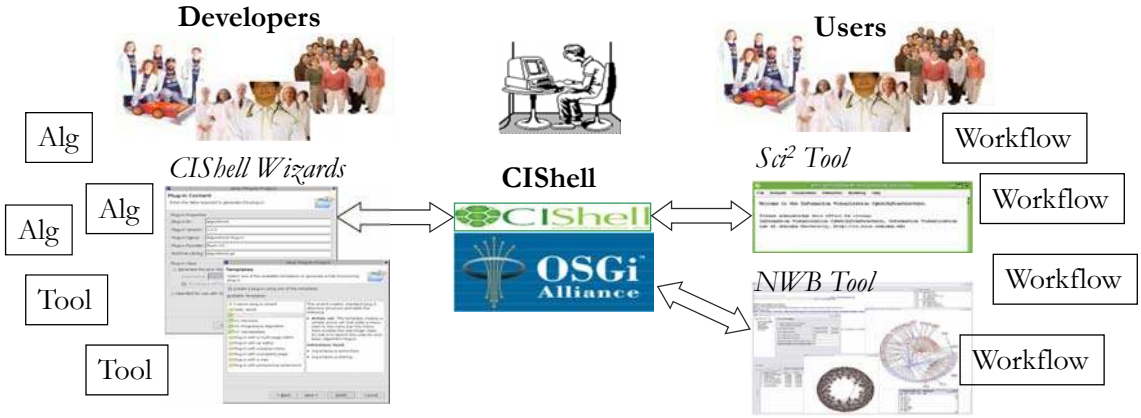
EpiC
cyberinfrastructure for NETWORK SCIENCE CENTER
CIShell Powered



Börner, Katy. (March 2011). Plug-and-Play Macroscopes. *Communications of the ACM, 54(3), 60-69.* http://www.scivee.tv/node/27704

OSGi & Cyberinfrastructure Shell (CIShell)

- CIShell (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), 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

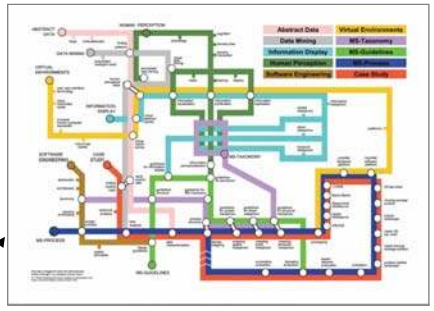
- Workflow design
  - Read data
  - Analyze
  - Visualize

- Discussion of specific algorithms

Different Question Types

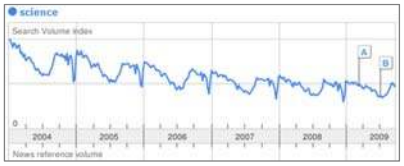
Terabytes of data

Descriptive & Predictive Models



Find your way

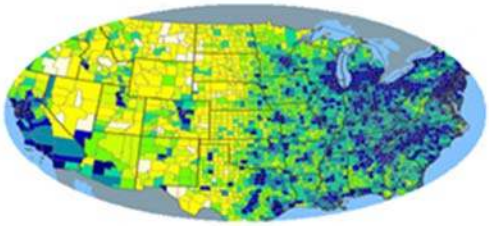
Find collaborators, friends



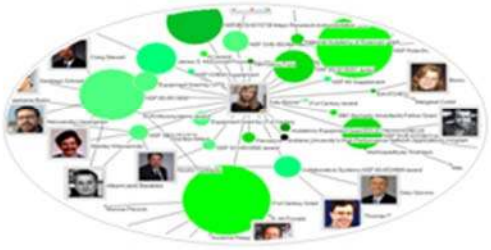
Identify trends

Different Levels of Abstraction/Analysis

Macro/Global
Population Level



Meso/Local
Group Level



Micro
Individual Level

Type of Analysis vs. Level of Analysis

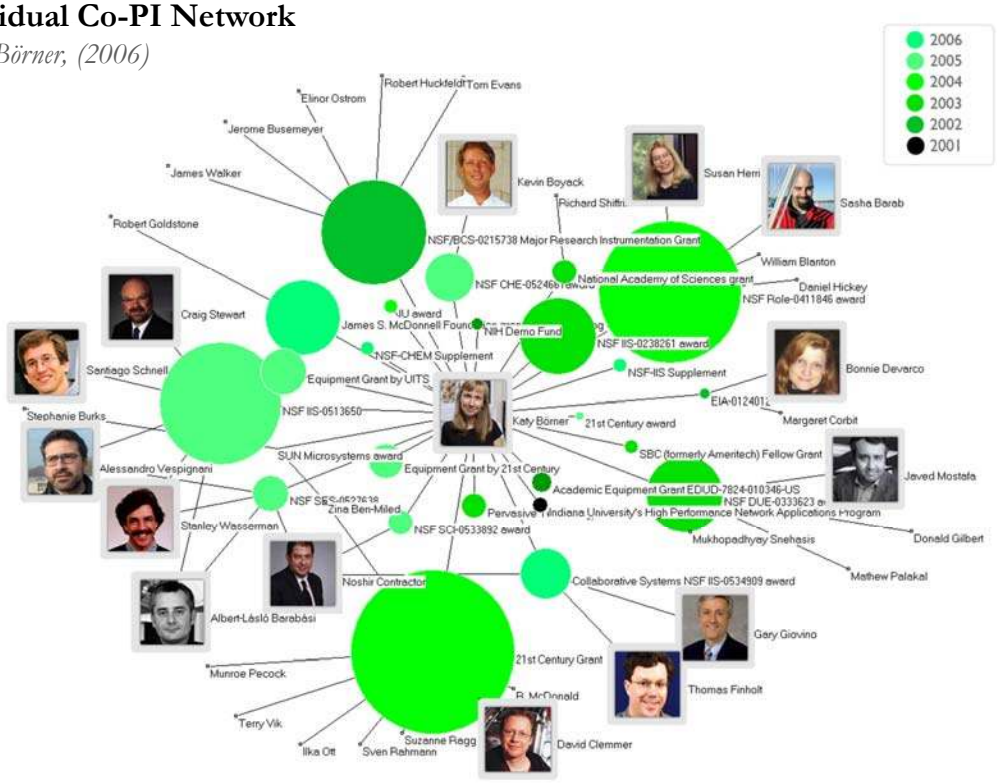
	Micro/Individual (1-100 records)	***Meso/Local (101–10,000 records)***	***Macro/Global (10,000 < records)***
Statistical Analysis/Profiling	Individual person and their expertise profiles	Larger labs, centers, universities, research domains, or states	All of NSF, all of USA, all of science.
Temporal Analysis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

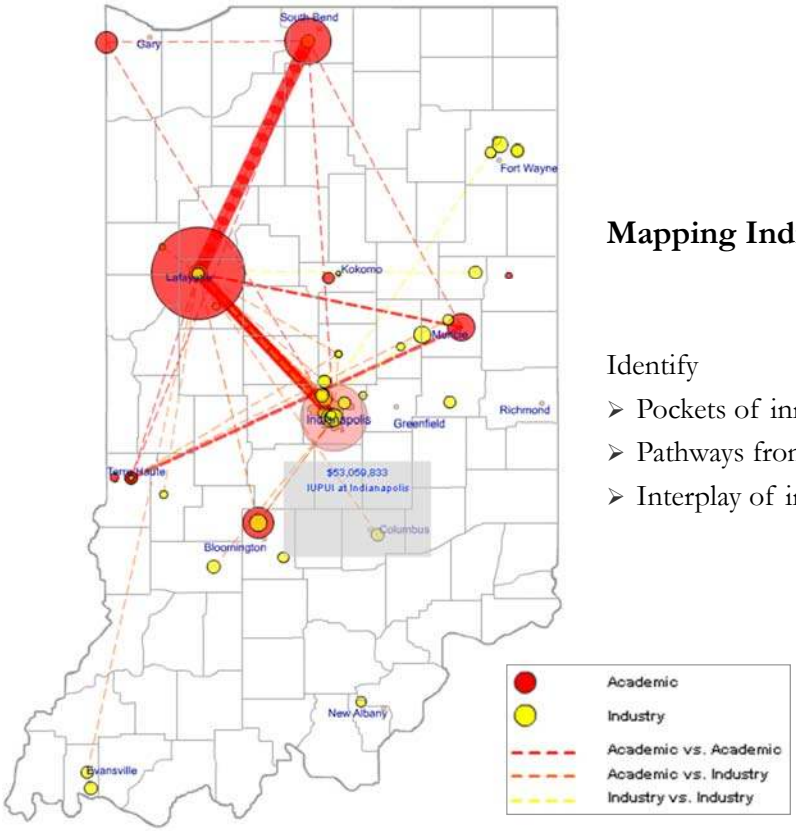
Type of Analysis vs. Level of Analysis

	Micro/Individual (1-100 records)	***Meso/Local (101–10,000 records)***	***Macro/Global (10,000 < records)***
Statistical Analysis/Profiling	Individual person and their expertise profiles	Larger labs, centers, universities, research domains, or states	All of NSF, NIH, NEA, all of science
Temporal Analysis (When)	Funding portfolio of one individual	Publication topic bursts of *PNAS*	113 years of physics research
Geospatial Analysis (Where)	Career trajectory of one individual	Mapping a scholar's intellectual landscape	*PNAS* publications
Topical Analysis (What)		Knowledge flows in chemistry research	
Network Analysis (With Whom?)	NSF network of one		NIH's network

Individual Co-PI Network

Ke & Börner, (2006)





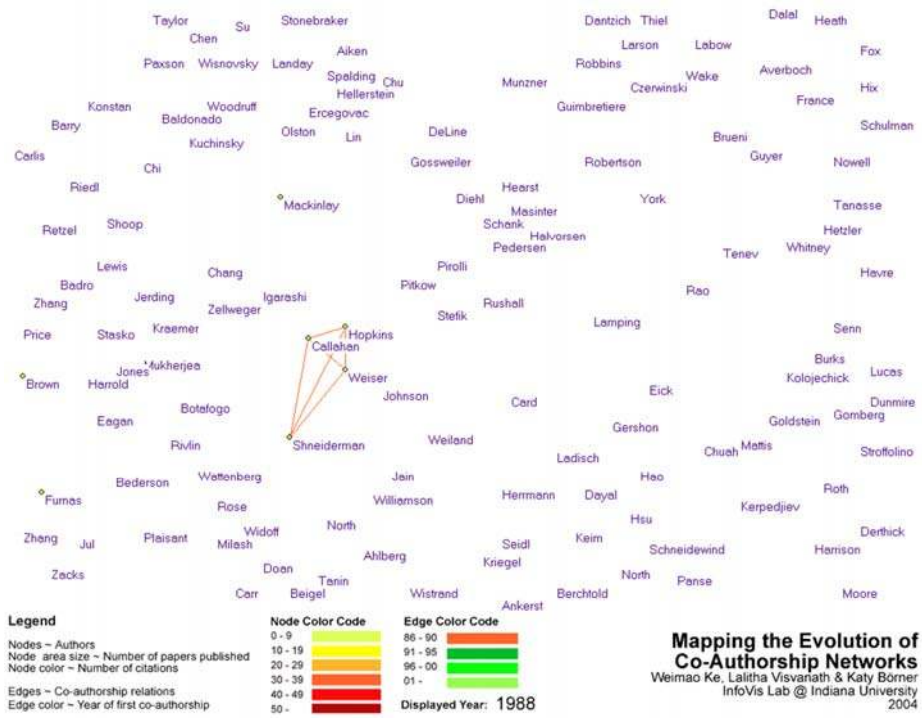
Mapping Indiana's Intellectual Space

Identify

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

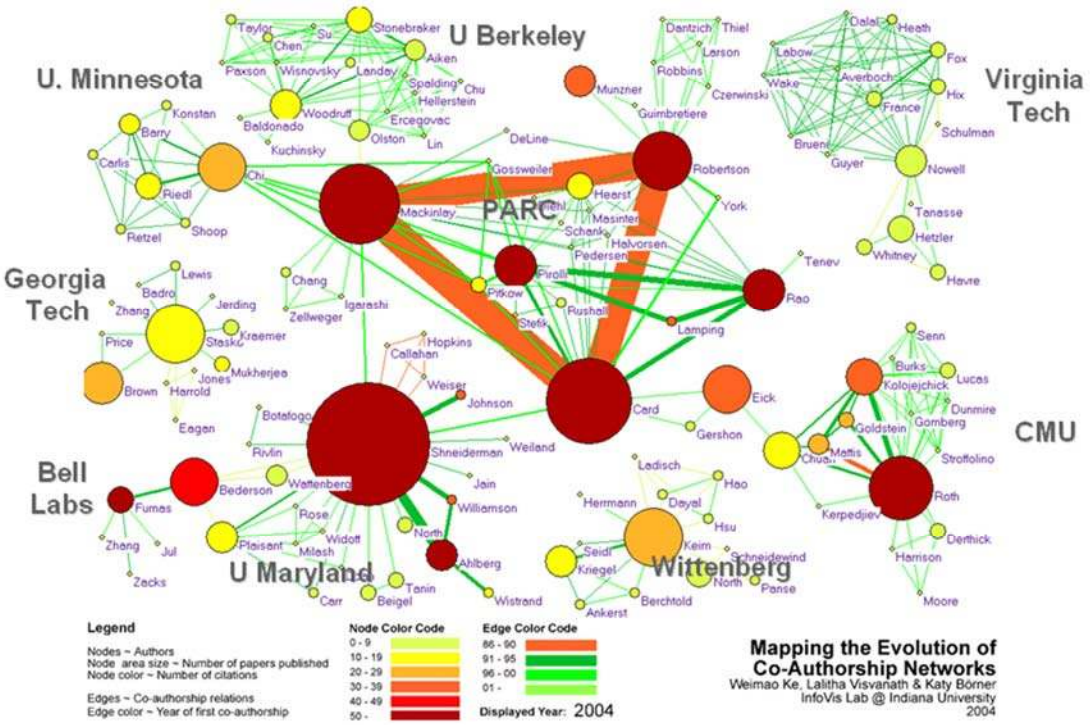
Mapping the Evolution of Co-Authorship Networks

Ke, Visvanath & Börner, (2004) Won 1st price at the IEEE InfoVis Contest.



Mapping the Evolution of Co-Authorship Networks

Ke, Visvanath & Borner, (2004) Won 1st price at the IEEE InfoVis Contest

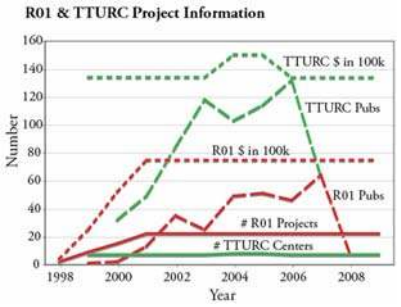


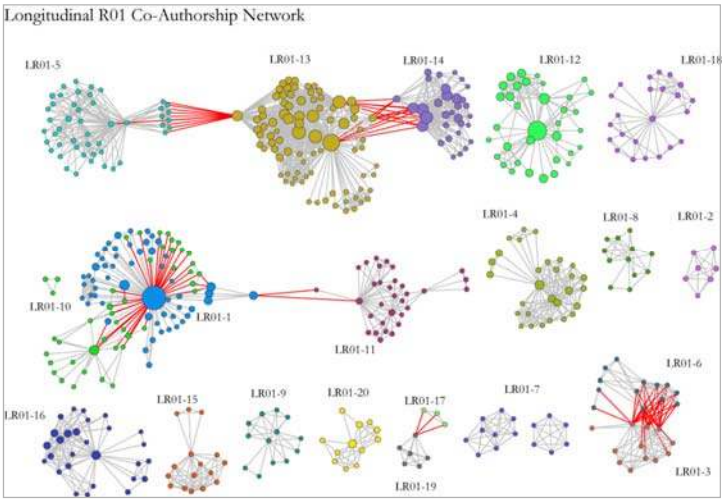
Mapping Transdisciplinary Tobacco Use Research Centers Publications

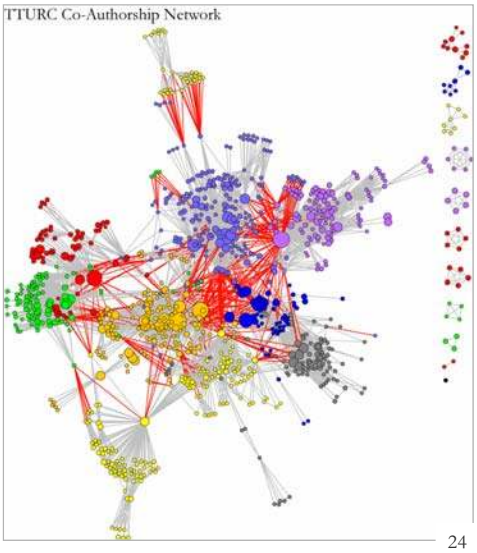
Compare R01 investigator based funding with TTURC Center awards in terms of number of publications and evolving co-author networks.

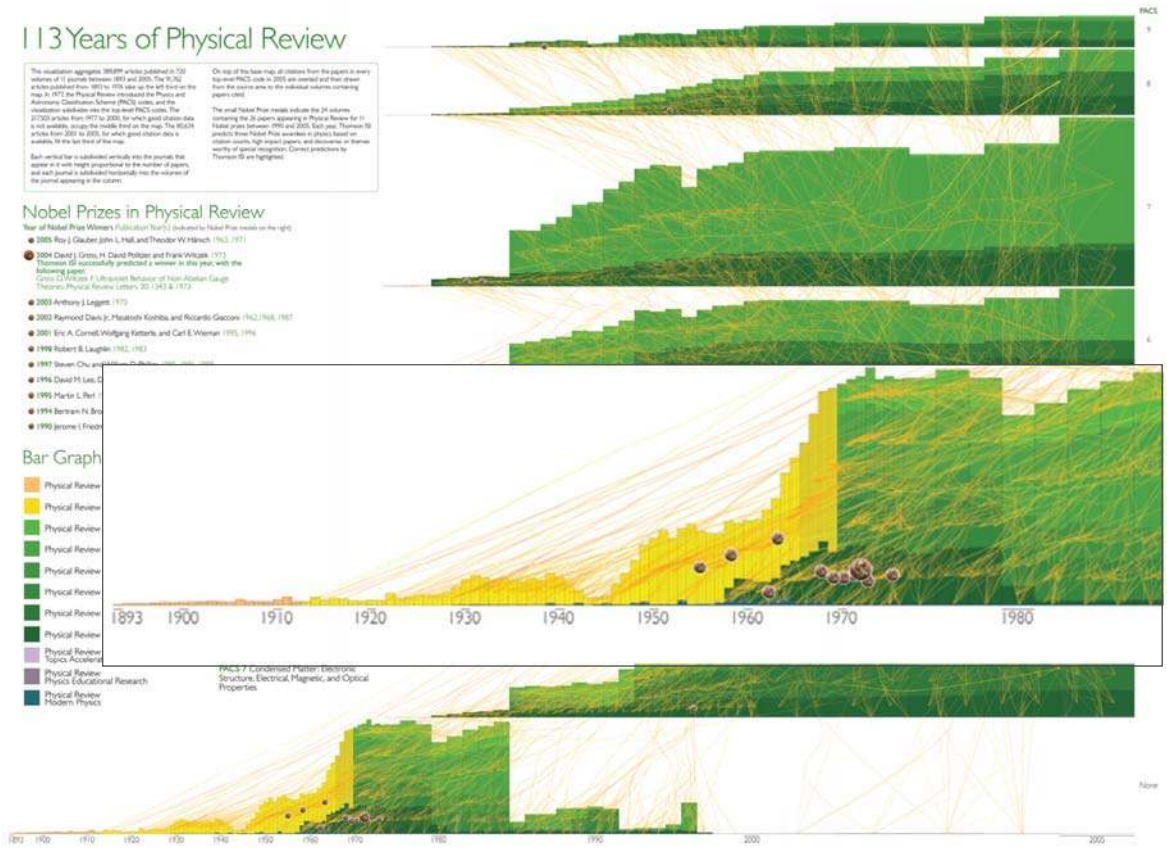
Zoss & Börner, forthcoming.

Supported by NIH/NCI Contract HHSN261200800812





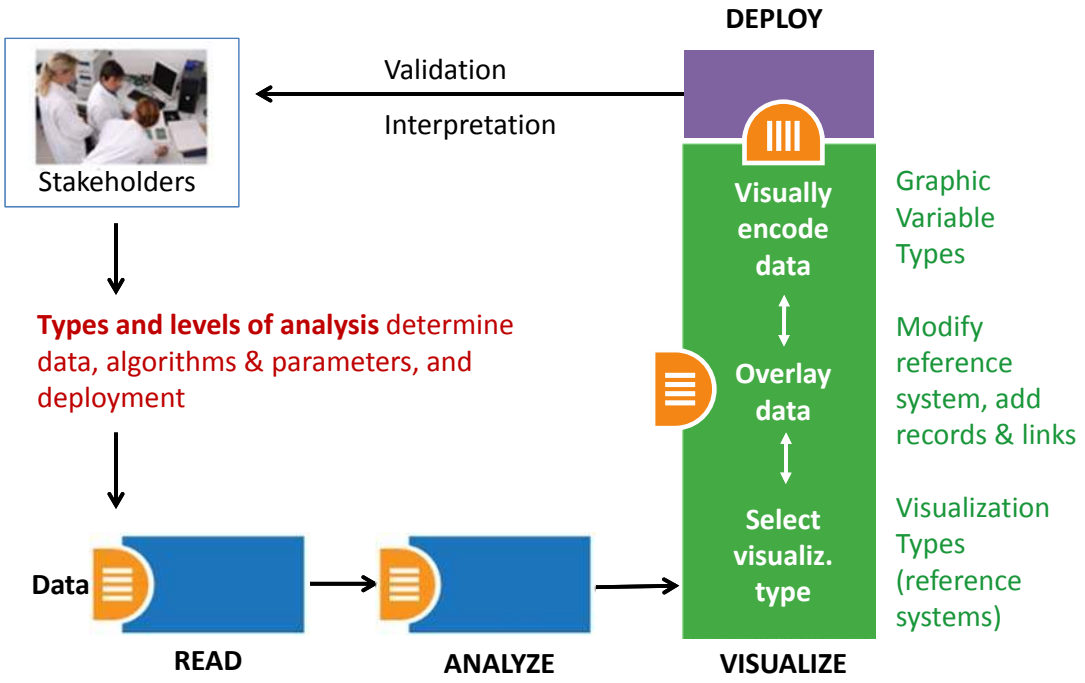




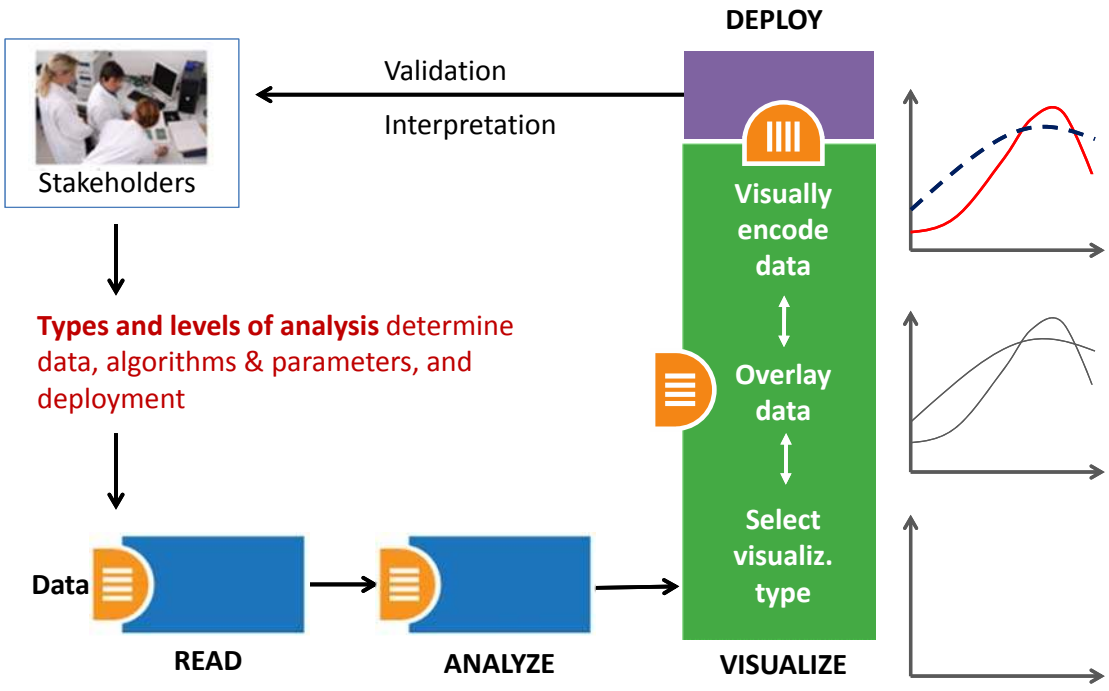
Type of Analysis vs. Level of Analysis

	Micro/Individual (1-100 records)	***Meso/Local (101–10,000 records)***	***Macro/Global (10,000 < records)***
Statistical Analysis/Profiling	Individual person and their expertise profiles	Larger labs, centers, universities, research domains, or states	All of NSF, all of scienc
Temporal Analysis (When)	Funding portfolio of one individual	ic bursts of *PNAS*	113 years of pl research
Geospatial Analysis (Where)	Career trajectory of one individual	Mapping a s intellectual la	*PNAS* p
Topical Analysis (What)	s.	Knowledge fl chemistry research	
Network Analysis (With Whom?)	NSF ork of one	k	NIH's cy

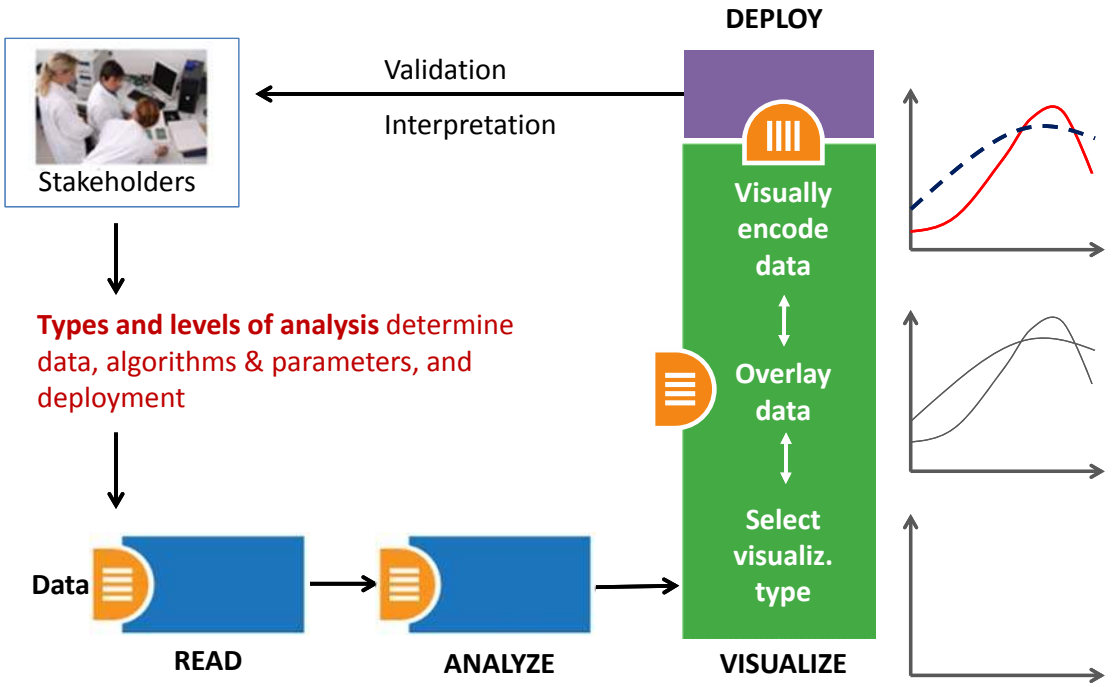
Needs-Driven Workflow Design



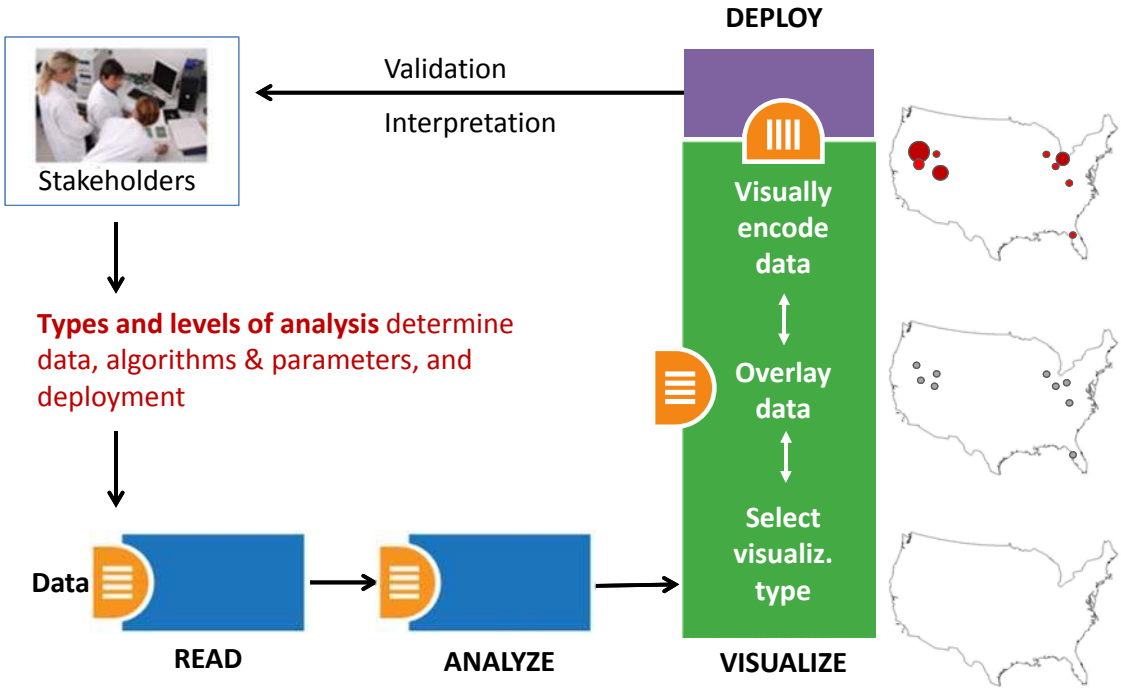
Needs-Driven Workflow Design



Needs-Driven Workflow Design



Needs-Driven Workflow Design

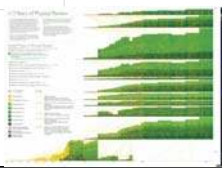
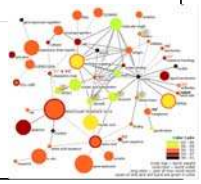


Visualization Types vs. Data Overlays

Visualization Type	Chart	Table	Graph	Geospatial Map	Network Graph
Modify / **visually encode** base map.					
Place and **visually encode** records/nodes.					
Place and **visually encode** links.					

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

Visualization Types vs. Data Overlays

Visualization Type	Chart	Table	Graph	Geospatial Map	Network Graph
Modify / **visually encode** base map.					
Place and **visually encode** records/nodes.					
Place and **visually encode** links.					

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

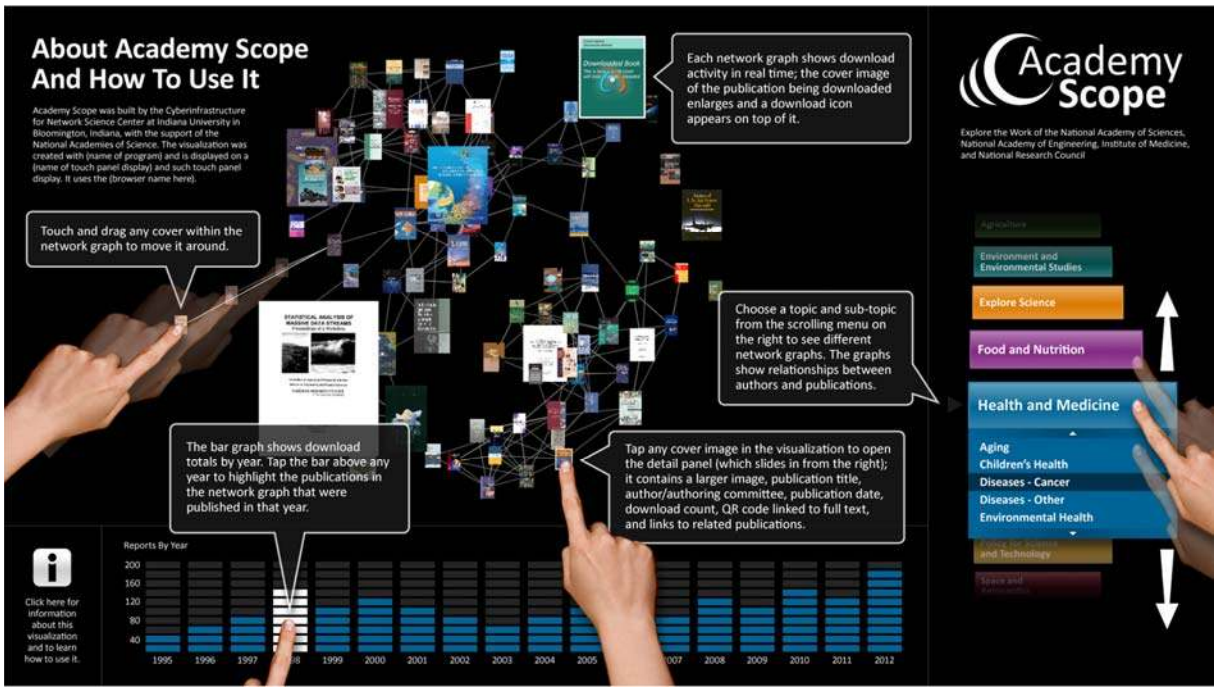
Mid-Term

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_TwZXnZrkg

Clients

Information Visualization MOOC | INDIANA UNIVERSITY | CNS

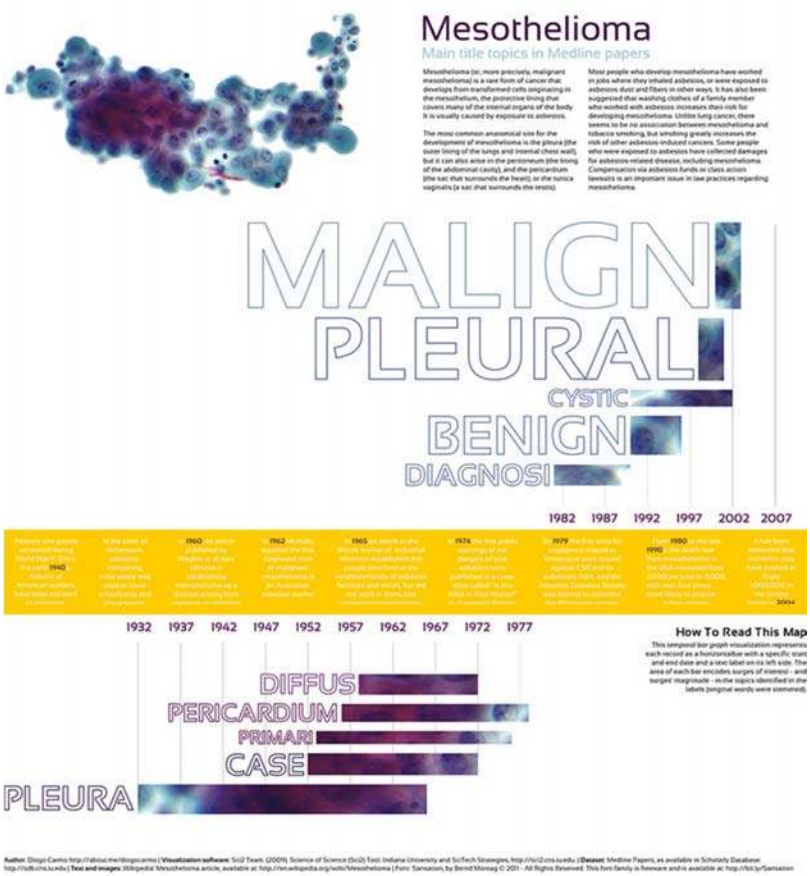
List of Clients

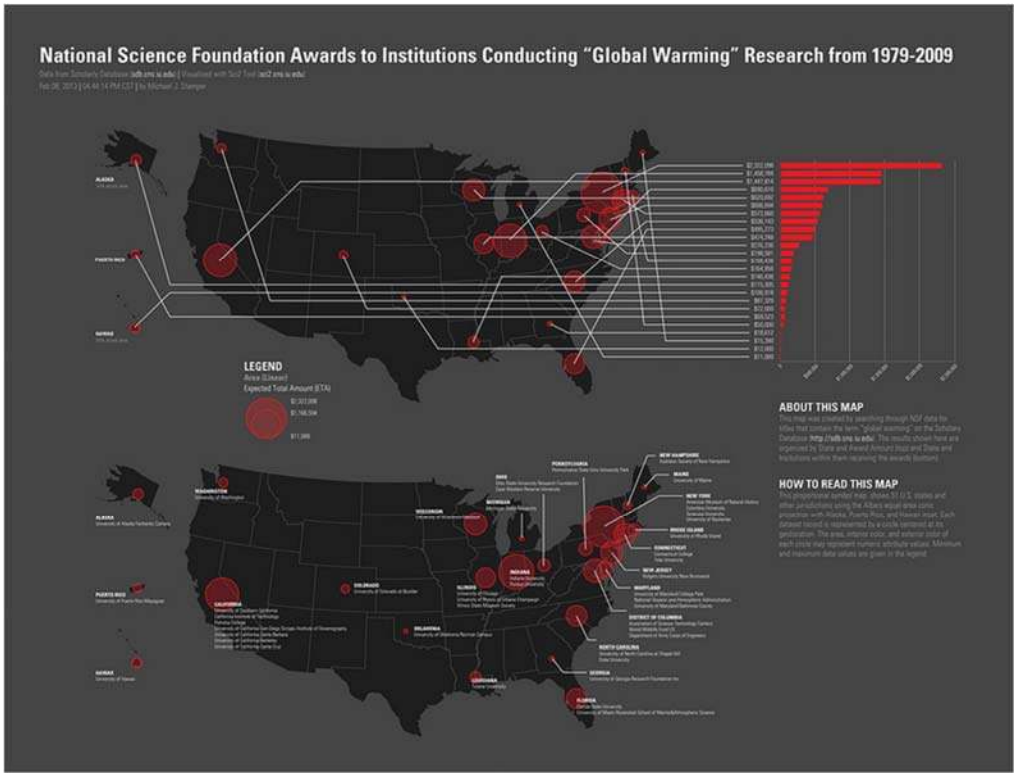
Project Title: Isis: 100 Years
Client Name: Jay Malone
Project goal/scientific or practical value: A visual representation Isis' contributors and locales over the past 100 years. Isis is the journal of the History of Science Society. This representation will provide a dynamic picture of how scholarship in the history of science has shifted over the past century.
Information on dataset(s) to be used: Citation information, author locale, and issue number for Isis publications.
Relevant publications, websites, etc: http://www.press.uchicago.edu/ucp/journals/journal/isis.html
Conditions under which students can publish results and/or add project results to their resume: Client would like to approve results.

Project Title: e-Xploration
Client Name: Luiy
Project goal/scientific or practical value: e-Xploration is an agent-based model for the ethnographic observation and the registry, analysis, and interpretation of social practices in virtual communities for intervention in the development of collaboration and cooperation. This project will analyze the interactions between subjects and objects in a platform collaborative community called OYCIB, a project based on e-Xploration (e-crick.net).
Information on dataset(s) to be used: I can provide a data base in .graphml format for the students. The file .graphml contains the interactions between subjects and objects in a platform collaborative community called OYCIB. In the level of practice, it is not necessary that students know agent-based models for using the database. But, in another level, for example: the collaborate level for the OYCIB development, it is necessary to have basic knowledge in AMS or MAS and another competences like PHP and MySQL.
Relevant publications, websites, etc: http://www.e-crick.net/logs
Conditions under which students can publish results and/or add project results to their resume: If any person or institution use my dataset or another info about eXploration (e-crick.net, oycib.net), I need to approve the results and appear as co-author.

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

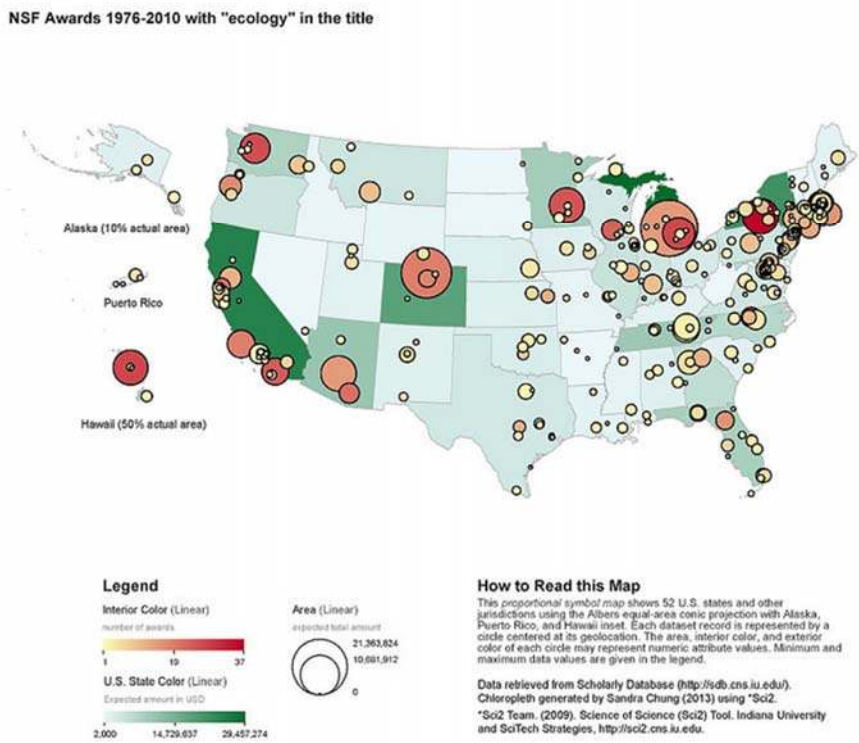
Diogo Carmo

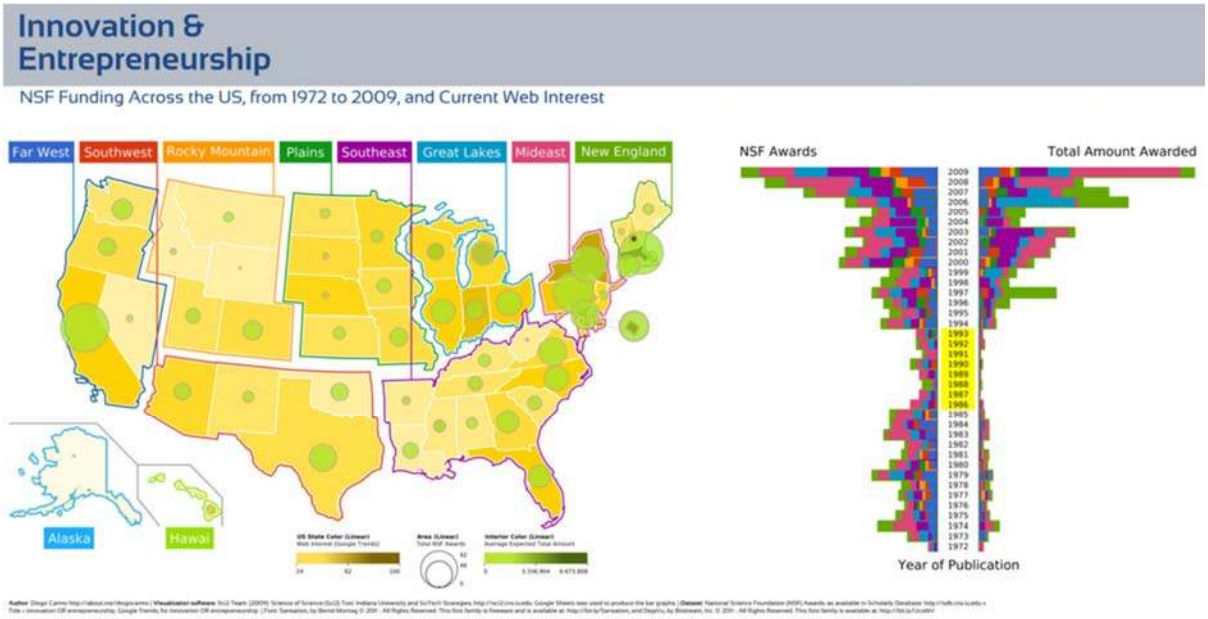




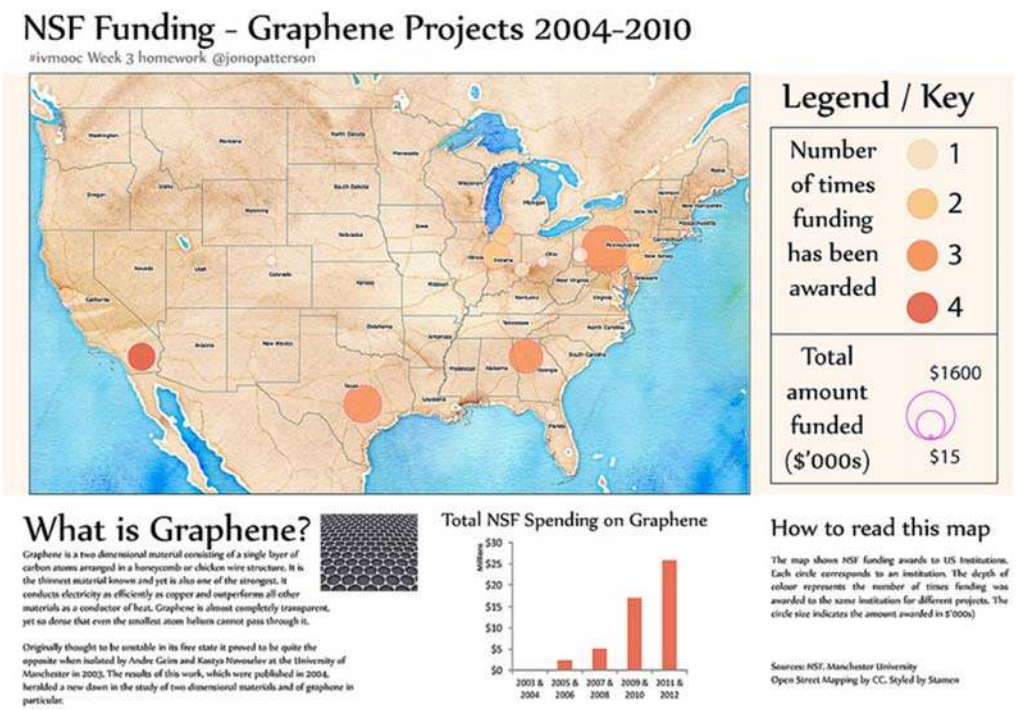
mjstamper_ivmooc

Sandra M. Chung

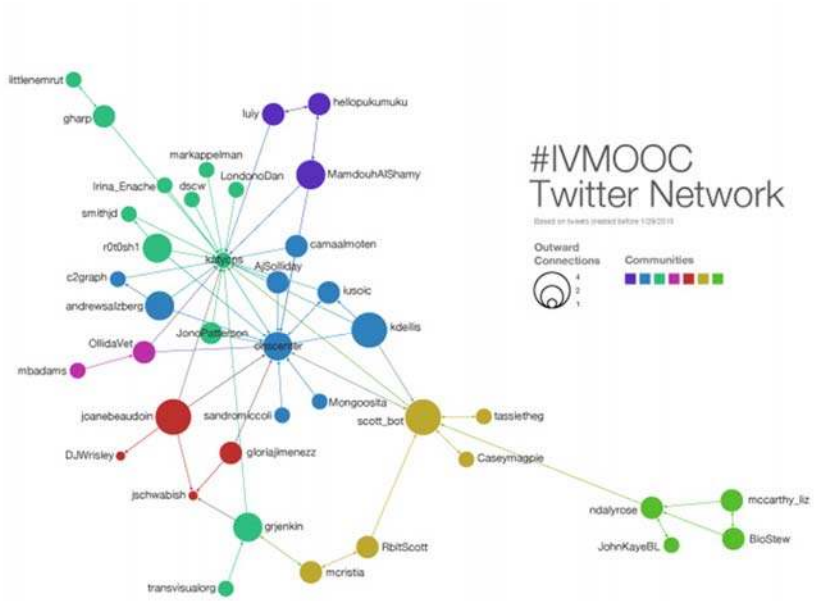




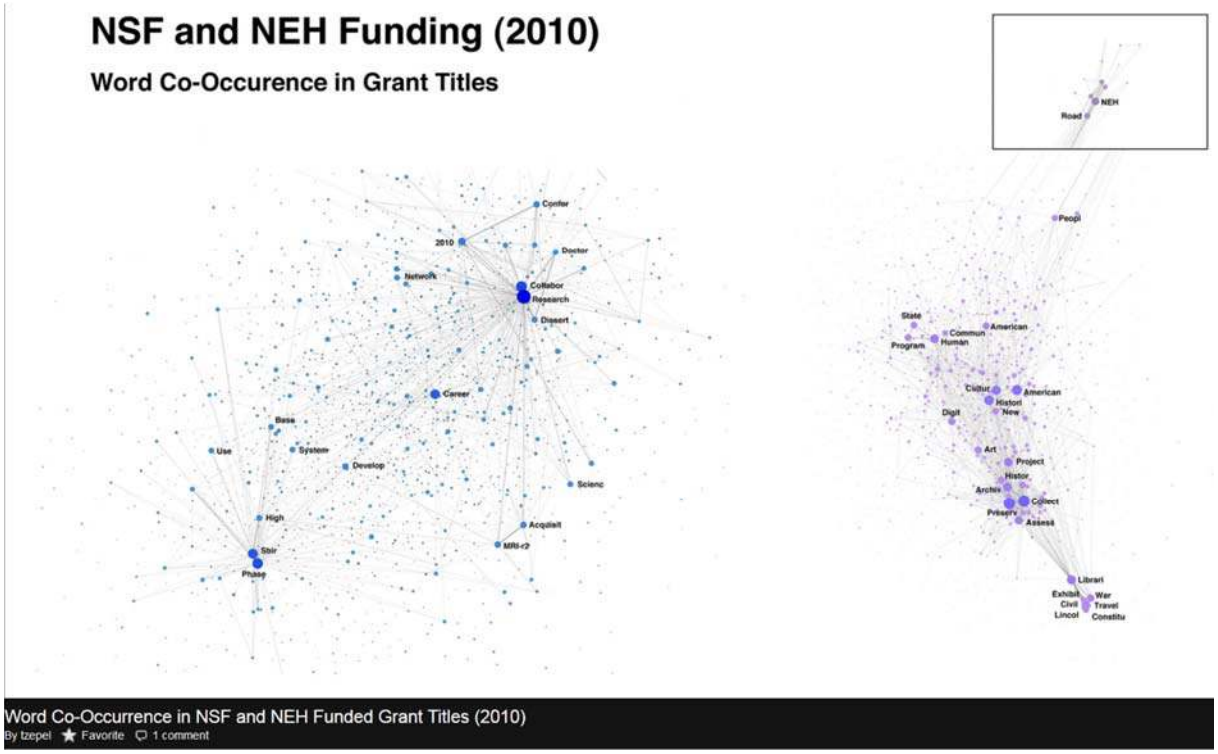
Diogo Carmo



JonoPatterson


#IVMOOC
Twitter Network
Outward Connections
Communities
littlenemrut
gharp
luiy
hellopukumuku
markappelman
LondonoDan
MamdouhAlShamy
Irina_Enache
dscw
smithjd
r0t0sh1
camaalmoten
AjSolitidy
c2graph
lusoic
andrewsalzberg
kdellis
OllidaVet
mbadams
joanebeaudoin
sandromiccoli
Mongoosita
scott_bot
tassietheg
Caseymagpie
DJWrisley
gloriajimenezz
jschwabish
grjenkin
RbitScott
mcristia
transvisualorg
ndalyrose
mccarthy_liz
JohnKayeBL
BlcStew

camaal


NSF and NEH Funding (2010)
Word Co-Occurence in Grant Titles
Word Co-Occurrence in NSF and NEH Funded Grant Titles (2010)
By tzepel ★ Favorite 1 comment

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

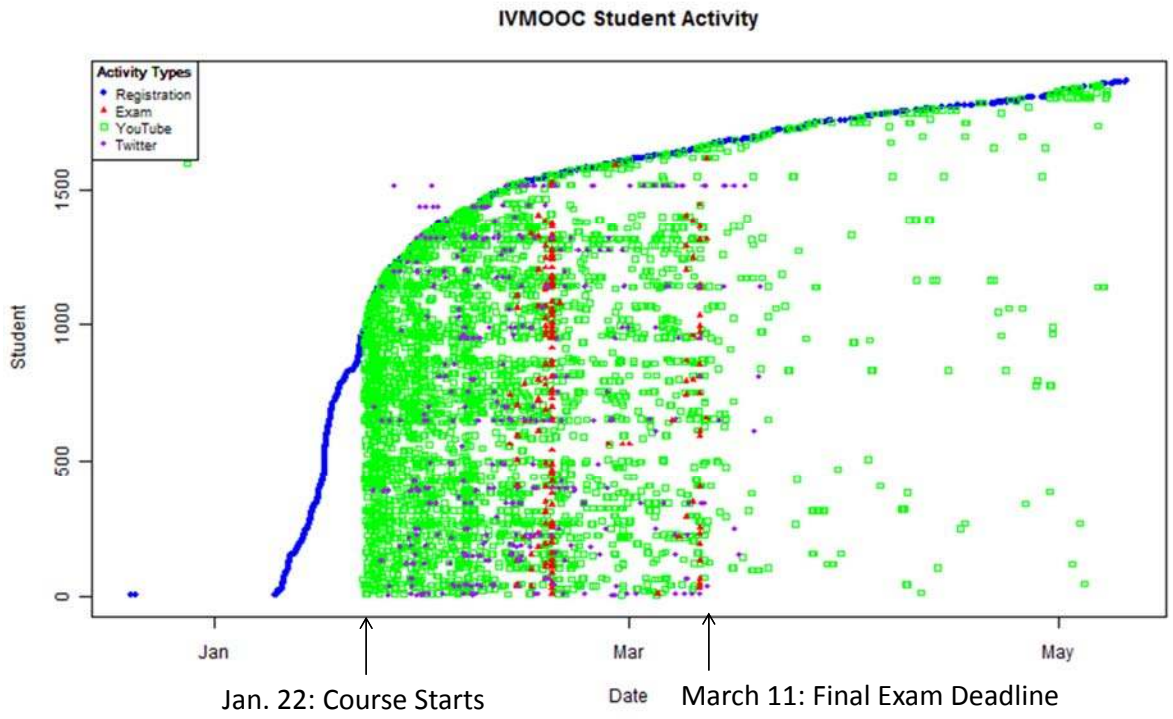
- Empowering **teachers**: How to make sense of the activities of thousands of students? How to guide them?
- Empowering **administrators**: What courses have the highest success rates are most profitable, etc.?
- Supporting **students**: How to navigate learning materials and develop successful learning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

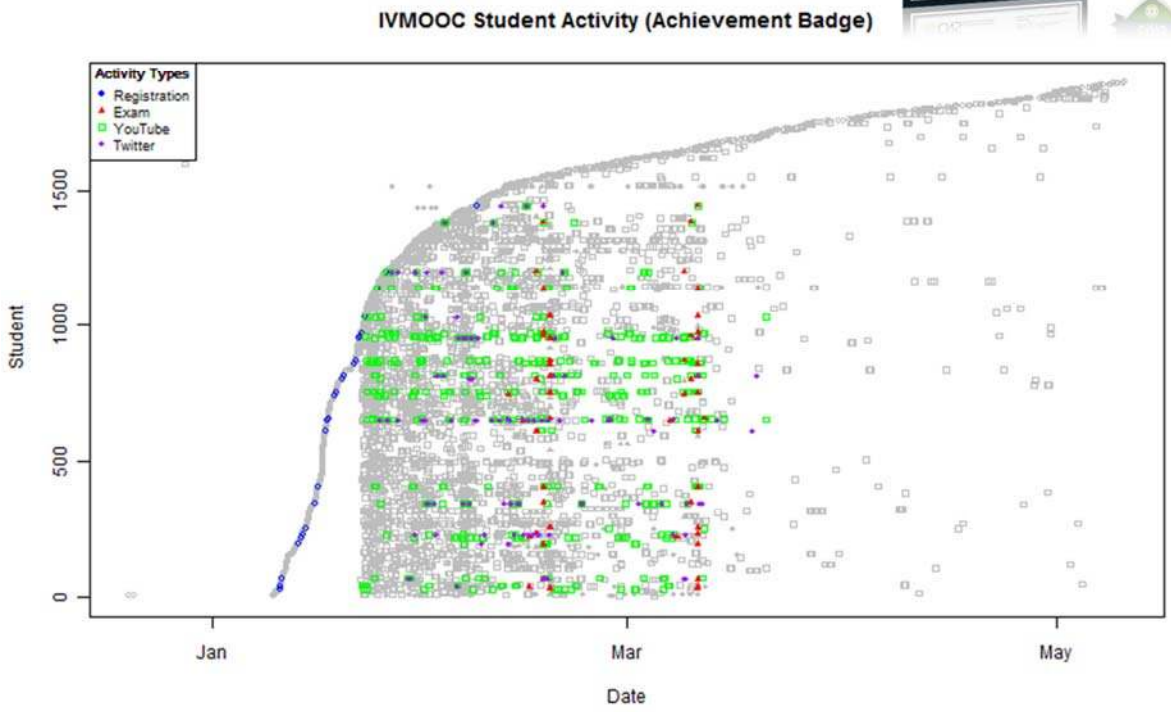
Visualizing IVMOOC Data

Data was collected from different sources:

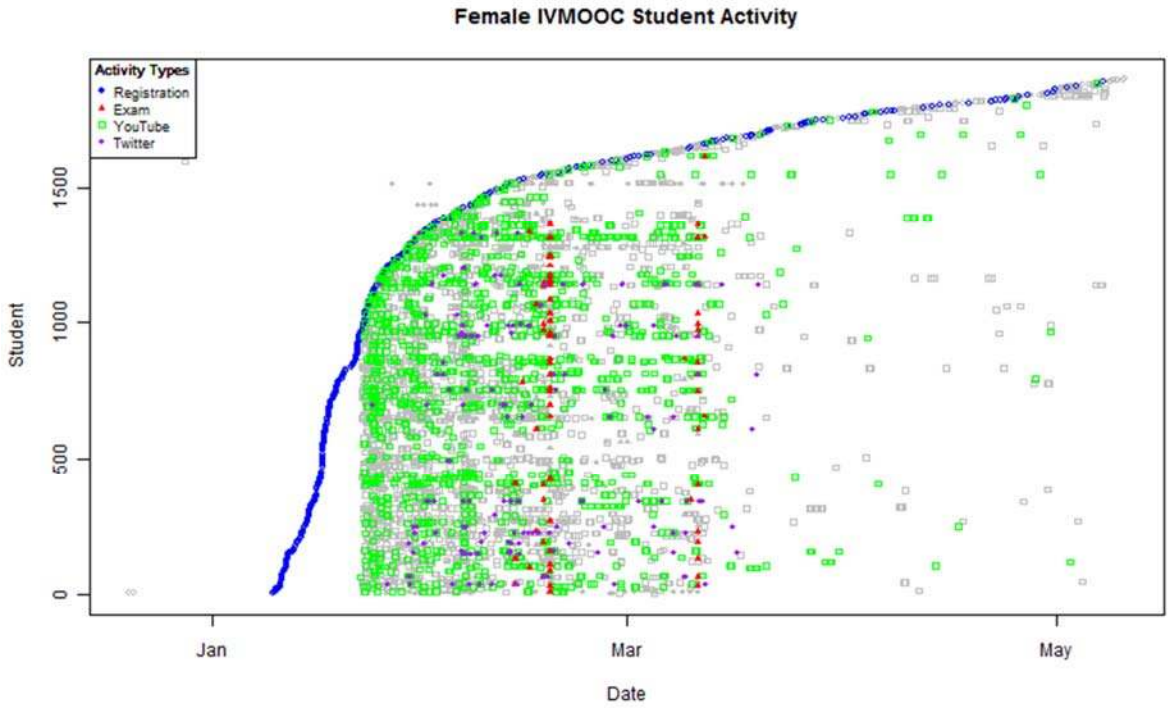
- 1,901 students registered via GCB (1215 male/557 female)
- 52,557 slide downloads from our server
- 18,893 video views via YouTube
- 193 accounts made 730 tweets
- 134 students took 183 exams in GCB
- 674 remarks on 215 different forum threads in Drupal
- 64 students submitted projects via Drupal

Visualizing IVMOOC Data

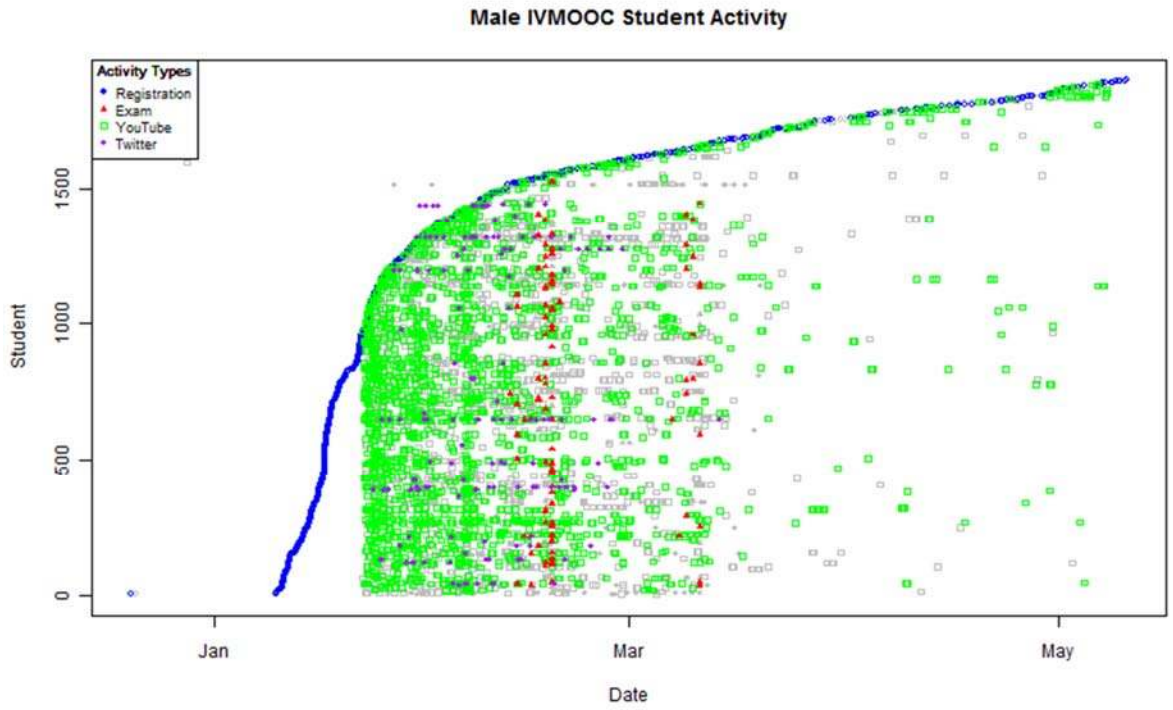


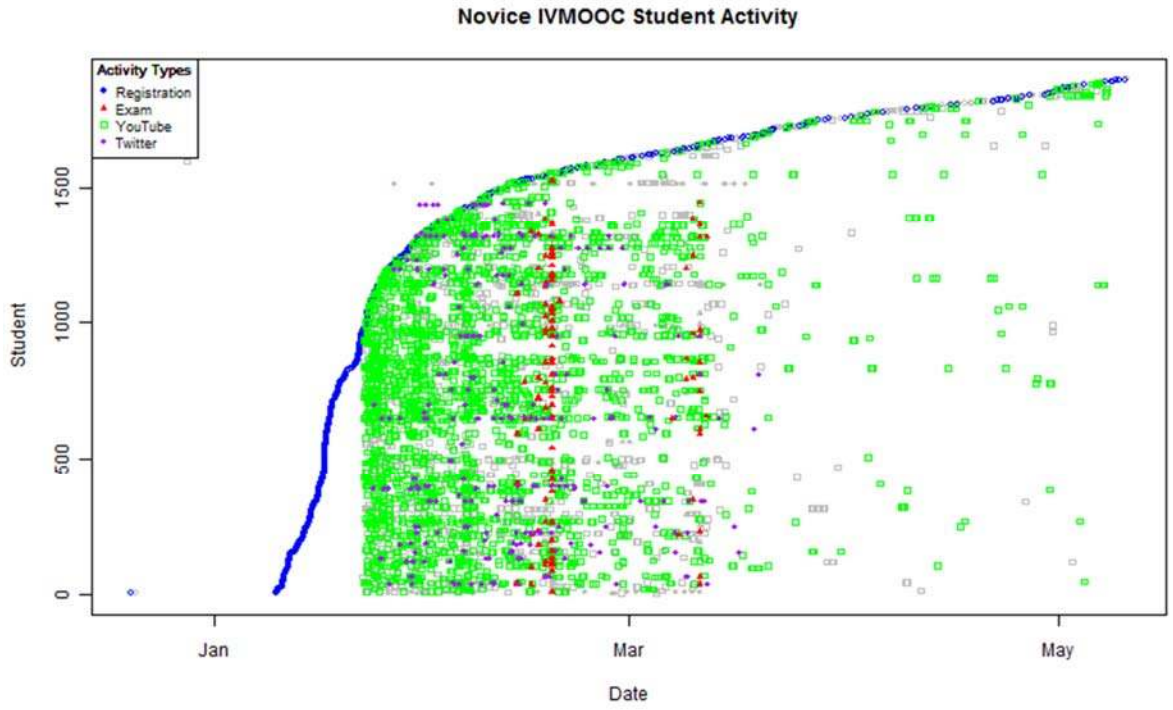


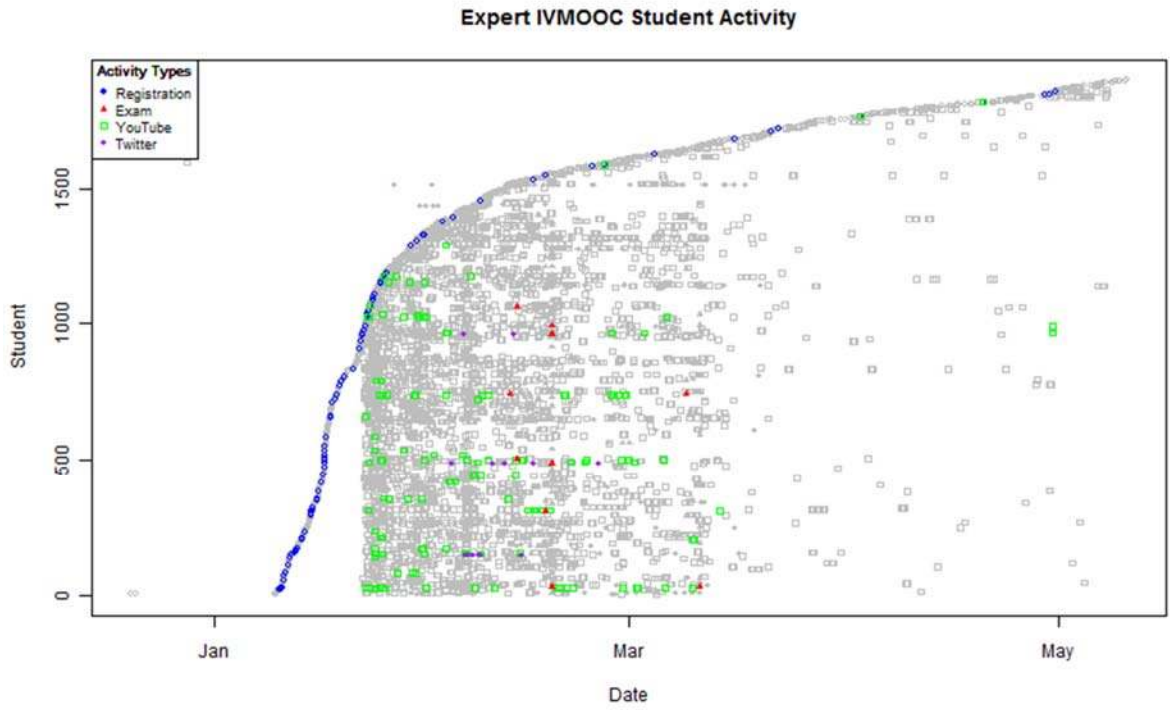
1215 male students
557 female students



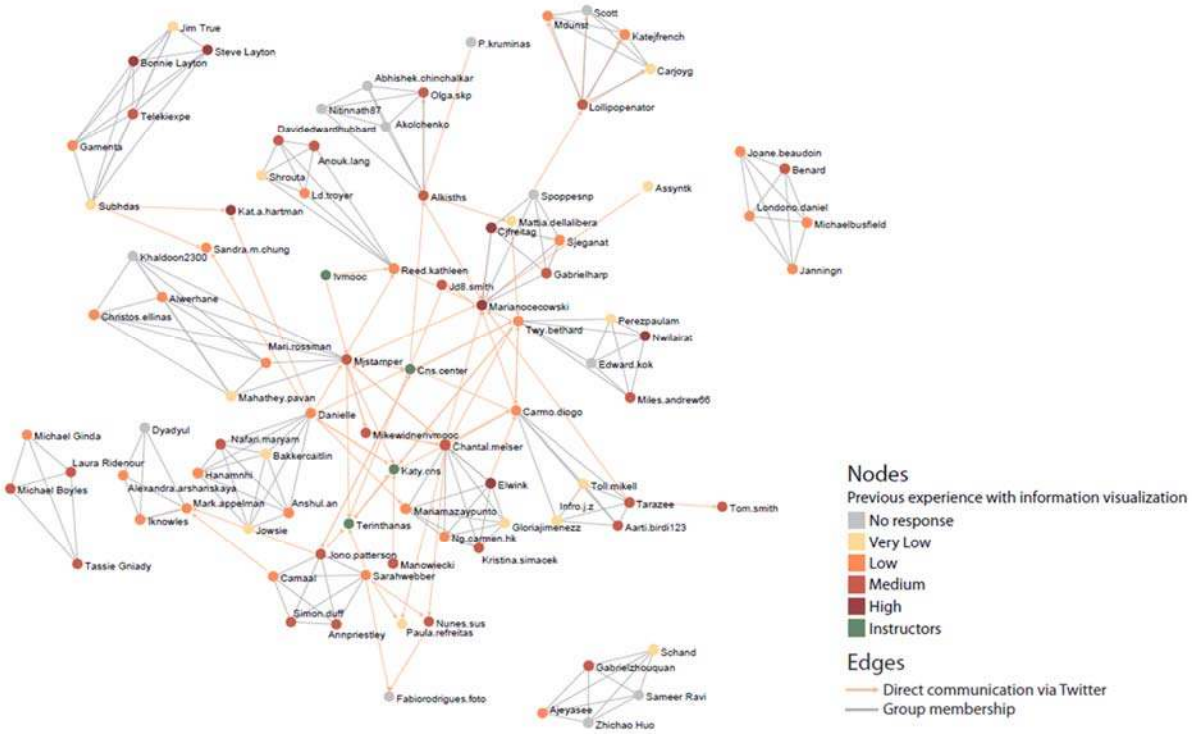
1215 male students
557 female students



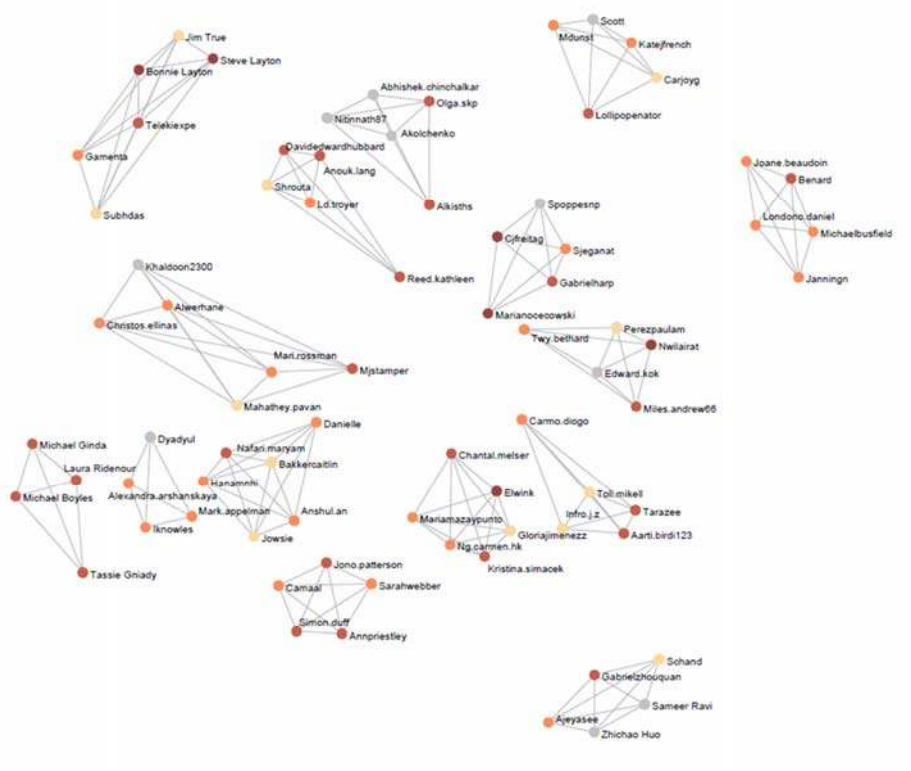




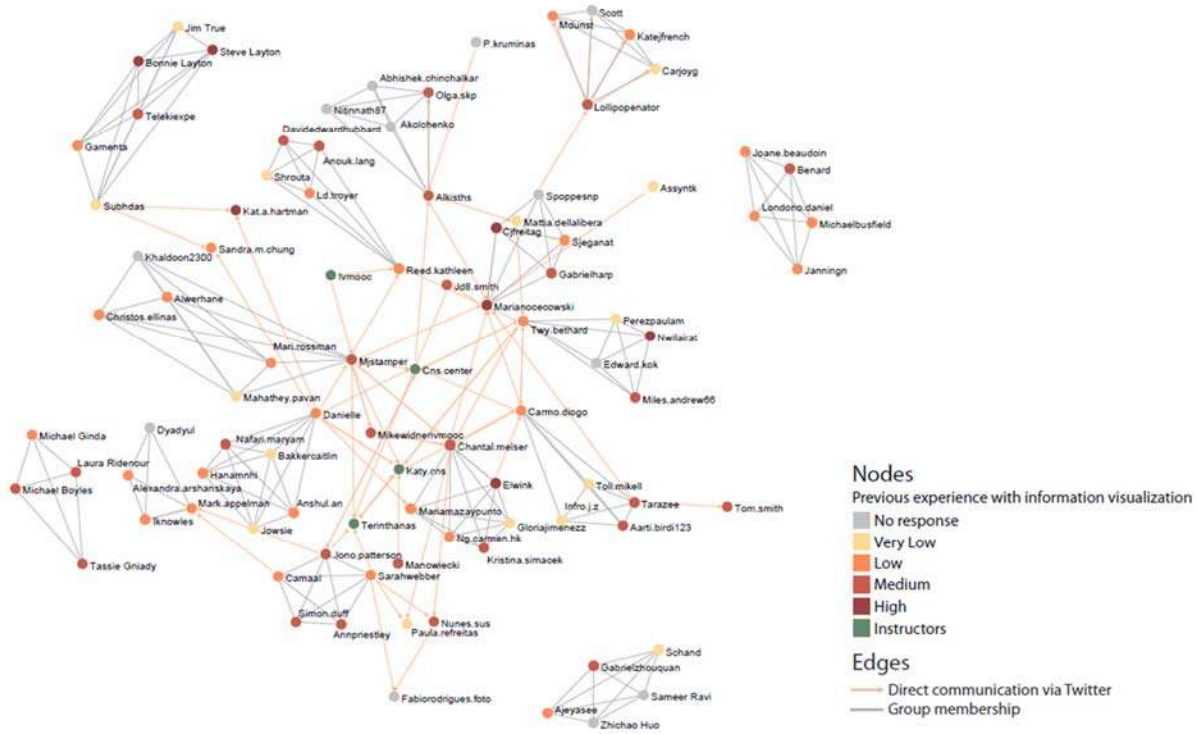
Student Client Projects: All Interactions



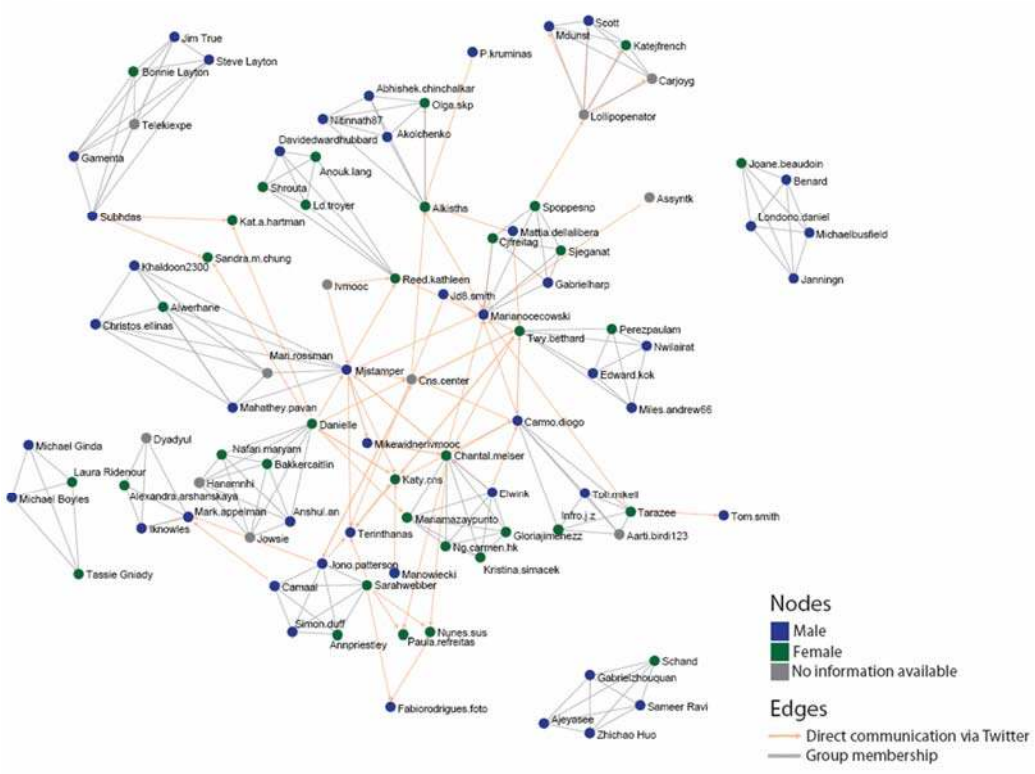
Student Client Projects: Group Memberships



Student Client Projects: All Interactions



Student Client Projects: Gender



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

Börner, Katy, Chen, Chaomei, and Boyack, Kevin. (2003). **Visualizing Knowledge Domains.** In Blaise Cronin (Ed.), *ARIST*, Medford, NJ: Information Today, Volume 37, Chapter 5, pp. 179-255. http://ivl.slis.indiana.edu/km/pub/2003-borner-arist.pdf

Shiffrin, Richard M. and Börner, Katy (Eds.) (2004). **Mapping Knowledge Domains**. *Proceedings of the National Academy of Sciences of the United States of America*, 101(Suppl_1). http://www.pnas.org/content/vol101/suppl_1/

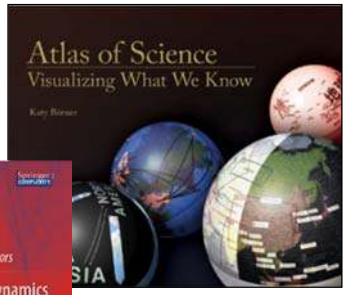
Börner, Katy, Sanyal, Soma and Vespignani, Alessandro (2007). **Network Science.** In Blaise Cronin (Ed.), *ARIST*, Information Today, Inc., Volume 41, Chapter 12, pp. 537-607.
http://ivl.slis.indiana.edu/km/pub/2007-borner-arist.pdf

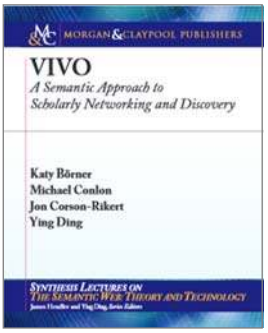
Börner, Katy (2010) **Atlas of Science**. MIT Press. http://scimaps.org/atlas

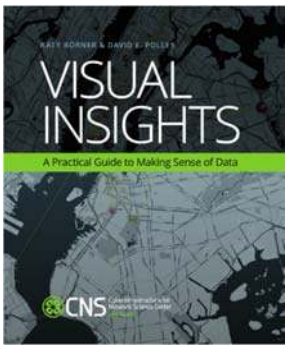
Scharnhorst, Andrea, Börner, Katy, van den Besselaar, Peter (2012) **Models of Science Dynamics**. Springer Verlag.

Katy Börner, Michael Conlon, Jon Corson-Rikert, Cornell, Ying Ding (2012) **VIVO: A Semantic Approach to Scholarly Networking and Discovery**. Morgan & Claypool.

Katy Börner and David E Polley (2014) **Visual Insights: A Practical Guide to Making Sense of Data**. MIT Press.







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