

# Visual Insights: A Practical Guide to Making Sense of Data

**Katy Börner**

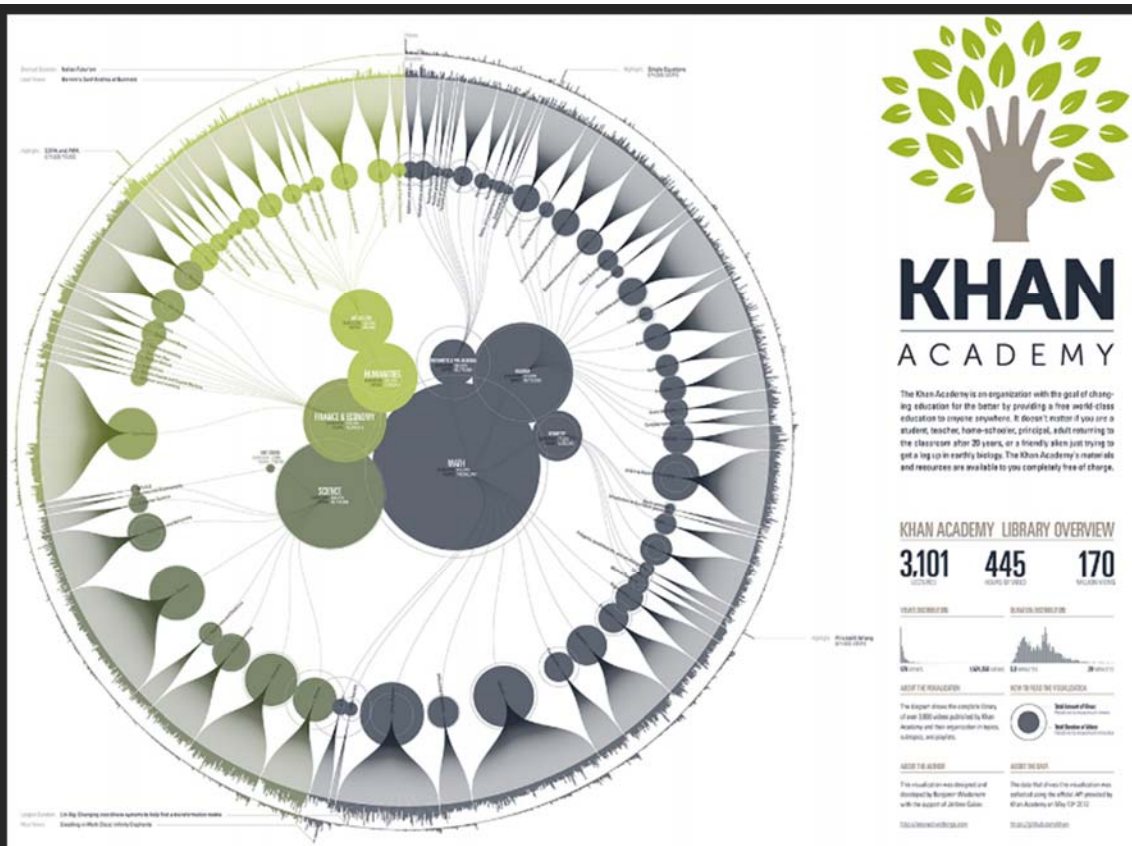
Cyberinfrastructure for Network Science Center, Director  
Information Visualization Laboratory, Director  
School of Library and Information Science  
Indiana University, Bloomington, IN  
[katy@indiana.edu](mailto:katy@indiana.edu)



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*

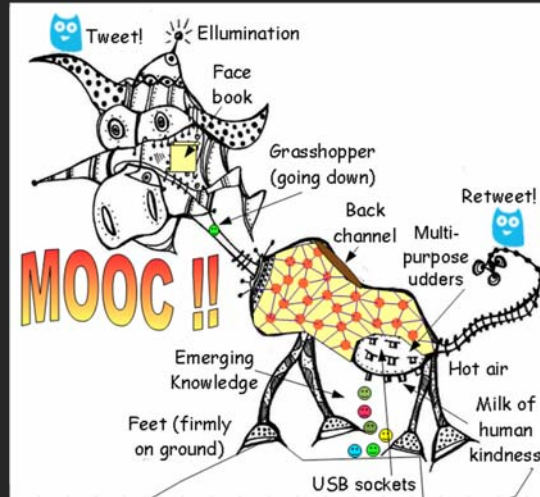


[http://scimaps.org/maps/map/khan\\_academy\\_library\\_147](http://scimaps.org/maps/map/khan_academy_library_147) by Benjamin Wiederkehr & Jérôme Cukier

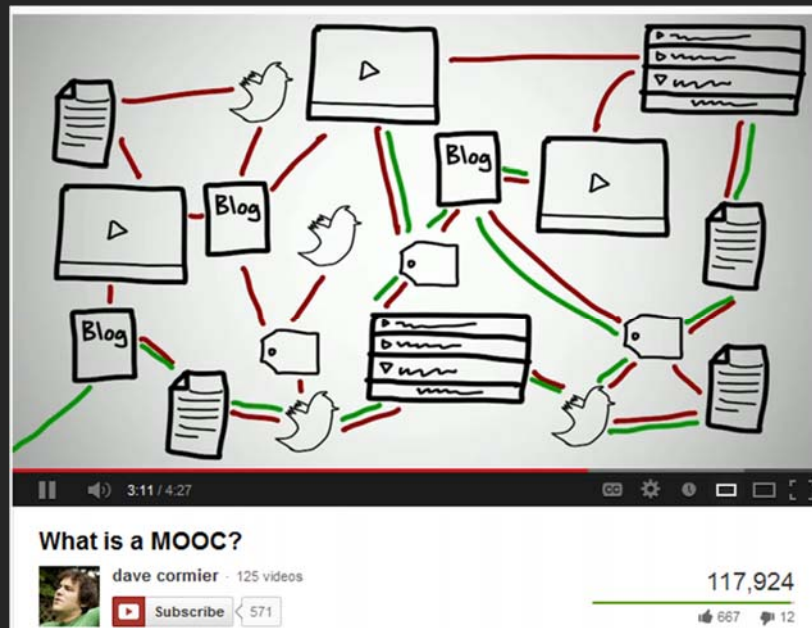
# MOOCs

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.



3



[http://www.youtube.com/watch?feature=player\\_embedded&v=eW3gMGqcZQc#at=128](http://www.youtube.com/watch?feature=player_embedded&v=eW3gMGqcZQc#at=128)

**Class Central**

search for courses

Class Central is a free online course aka MOOC aggregator from top universities like Stanford, MIT, Harvard, etc. offered via Coursera, Udacity, edX, Canvas Network, & others [Learn More](#)

Recently started or starting soon (55) Just Announced (31) Courses in Progress (117) Future courses (279)  
Self Paced (25) Finished courses (135)

<http://www.class-central.com>

**Information Visualization MOOC** INDIANA UNIVERSITY CNS

**Overview**

This course provides an overview about the state of the art in information visualization. It teaches the process of producing effective visualizations that take the needs of users into account.

Among other topics, the course covers:

- Data analysis algorithms that enable extraction of patterns and trends in data
- Major temporal, geospatial, topical, and network visualization techniques
- Discussions of systems that drive research and development.

Please watch the introduction video to get better acquainted with the course.

Everybody who registers gains free access to the Scholarly Database (26 million paper, patent, and grant records) and the Sci2 Tool (100+ algorithms and tools).

**Katy Börner, Ph.D.**  
Indiana University

IV MOOC

COMMUNICATIONS ACM

Plug-and-Play Macroeconomics

Börner, Katy. (March 2011). Plug-and-Play Macroeconomics. Communications of the ACM, 54(3), 6. <http://www.doi.org/10.1145/1934277>

**KATY BÖRNER**  
INDIANA UNIVERSITY

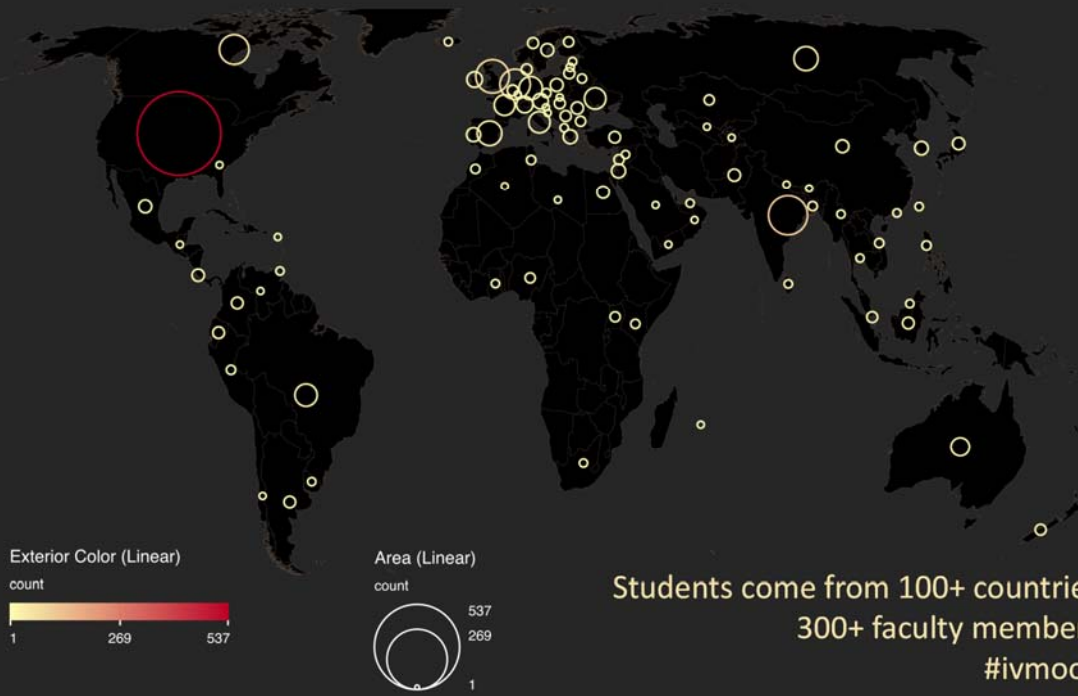
[Go To The Course](#)

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



7

## 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



8



# 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

9

# 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.

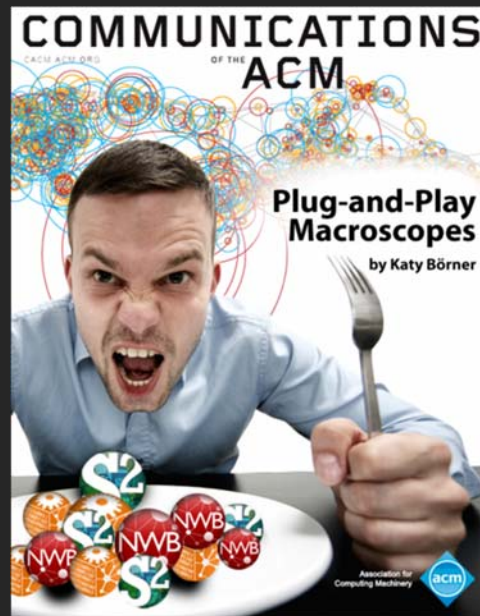
10

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.

The image shows two screenshots of the Scholarly Database website. The left screenshot displays the login page with options for 'IU User' and 'Non-IU User'. The right screenshot shows the search interface with fields for search terms, filters (creators, title, abstract, full text), and date ranges. It also includes a 'Search' button and a sidebar with search tips.

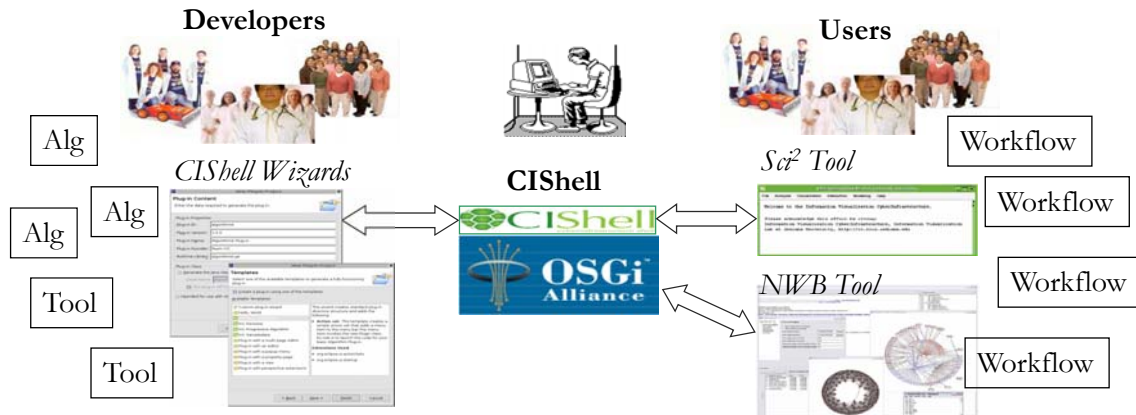
Register for free access at <http://sdb.cns.iu.edu>

Plug-and-Play Macroscopes  
cishell.org



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

- 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



13

## 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)

14

# 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

15

## Different Question Types



Terabytes of data

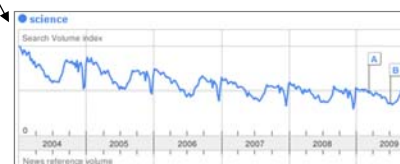
Descriptive &  
Predictive  
Models



Find your way



Find collaborators, friends



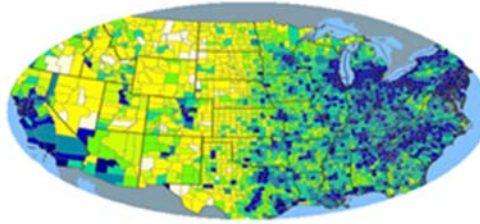
Identify trends

16



## Different Levels of Abstraction/Analysis

Macro/Global  
Population Level



Meso/Local  
Group Level



Micro  
Individual Level



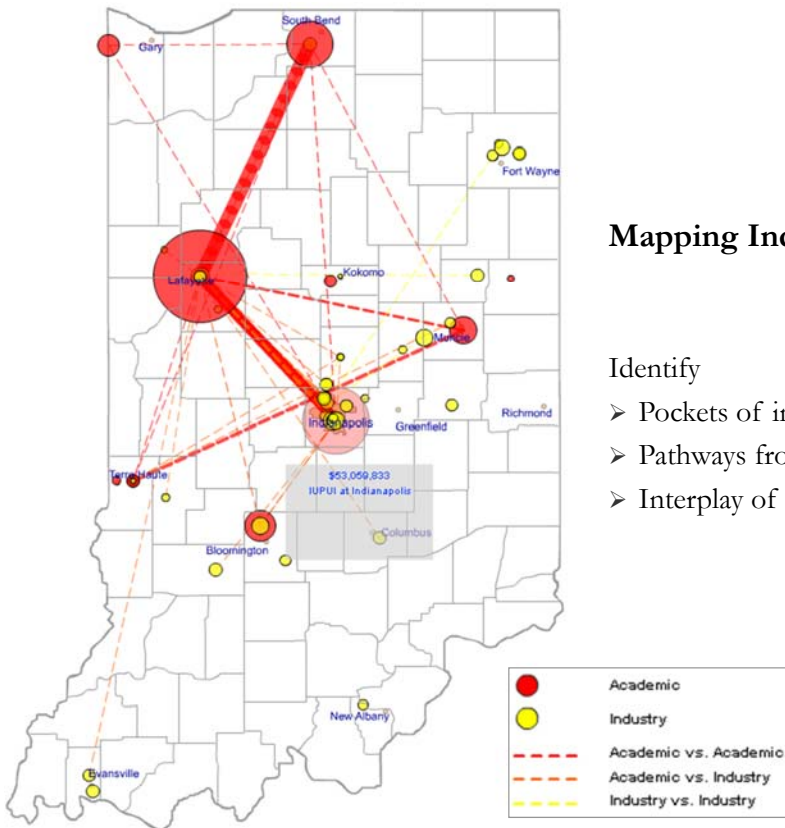
17

## Type of Analysis vs. Level of Analysis

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

18





## Mapping Indiana's Intellectual Space

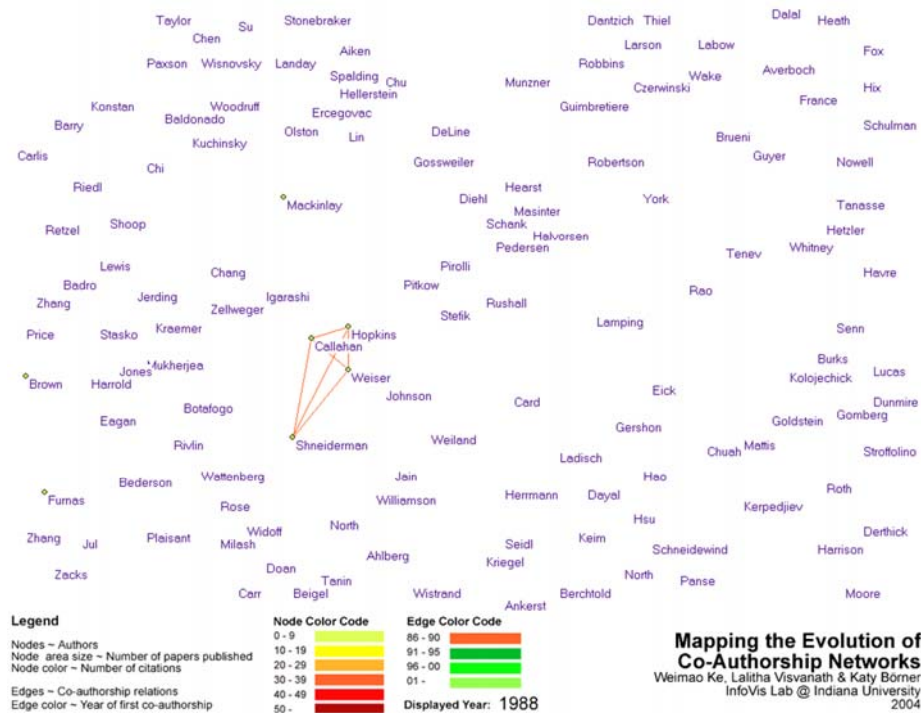
Identify

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

21

## Mapping the Evolution of Co-Authorship Networks

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

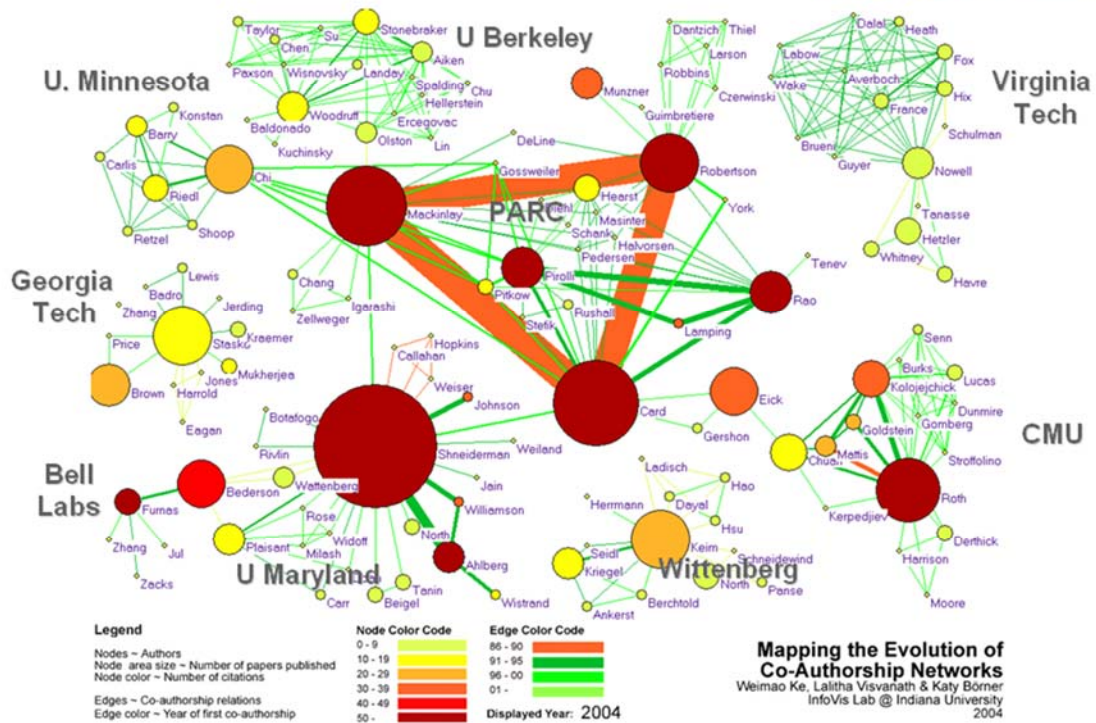


22



# Mapping the Evolution of Co-Authorship Networks

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



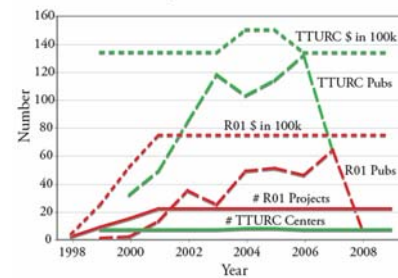
## Mapping Transdisciplinary Tobacco Use Research Centers Publications

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

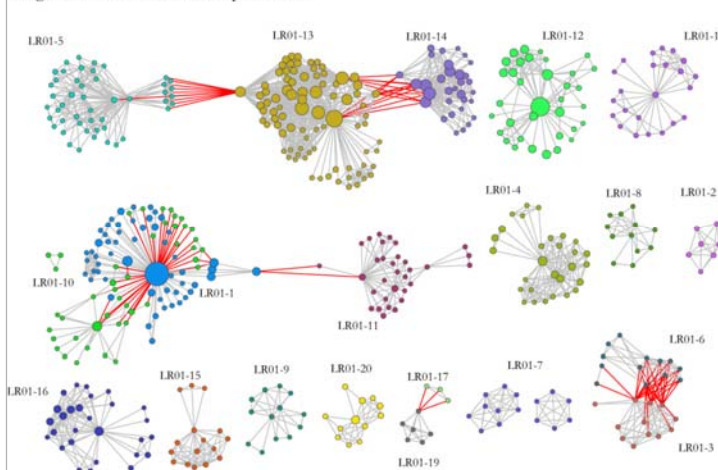
Zoss & Börner, forthcoming.

Supported by NIH/NCI Contract HHSN261200800812

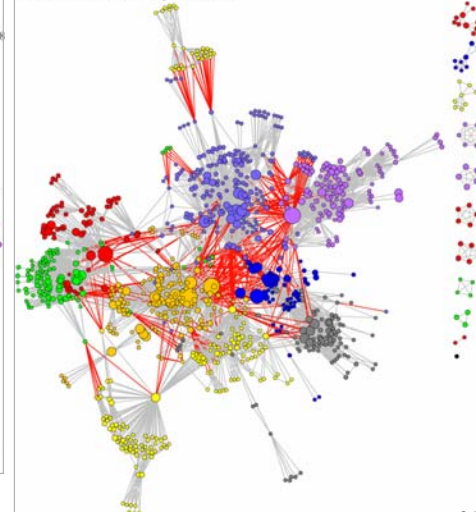
R01 & TTURC Project Information



Longitudinal R01 Co-Authorship Network



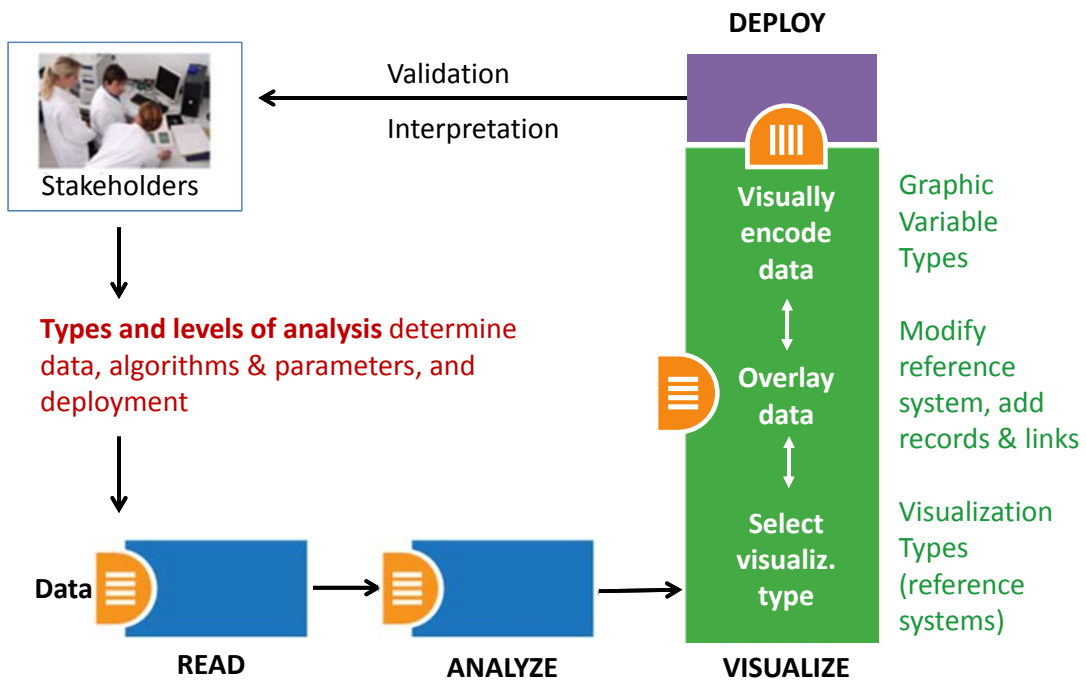
TTURC Co-Authorship Network





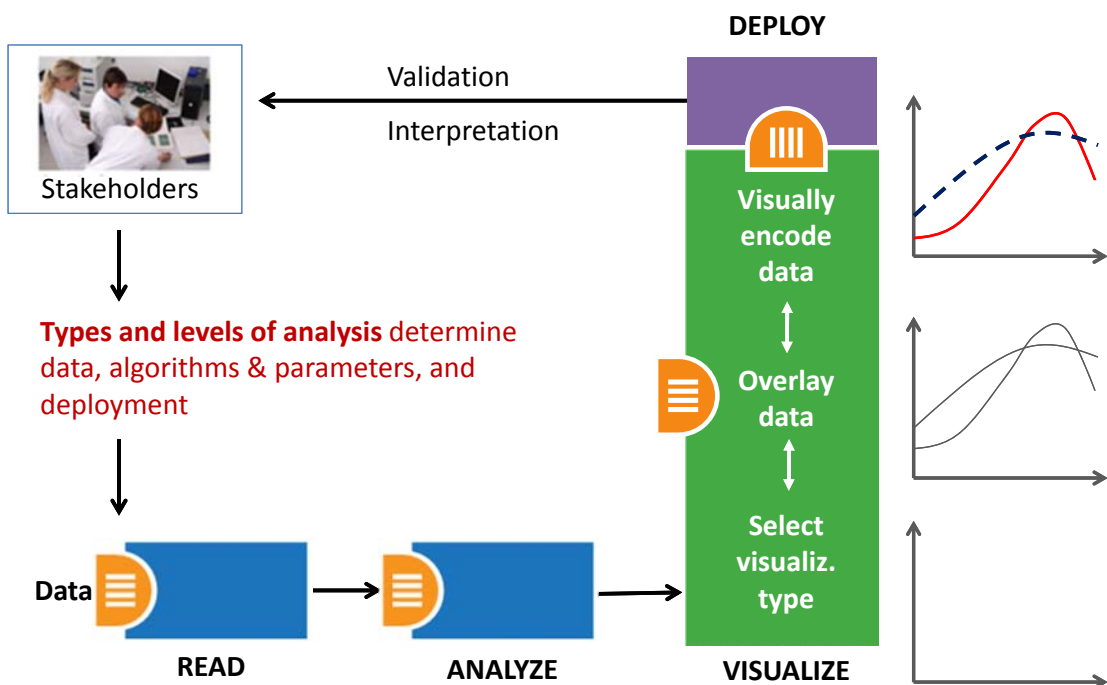


# Needs-Driven Workflow Design



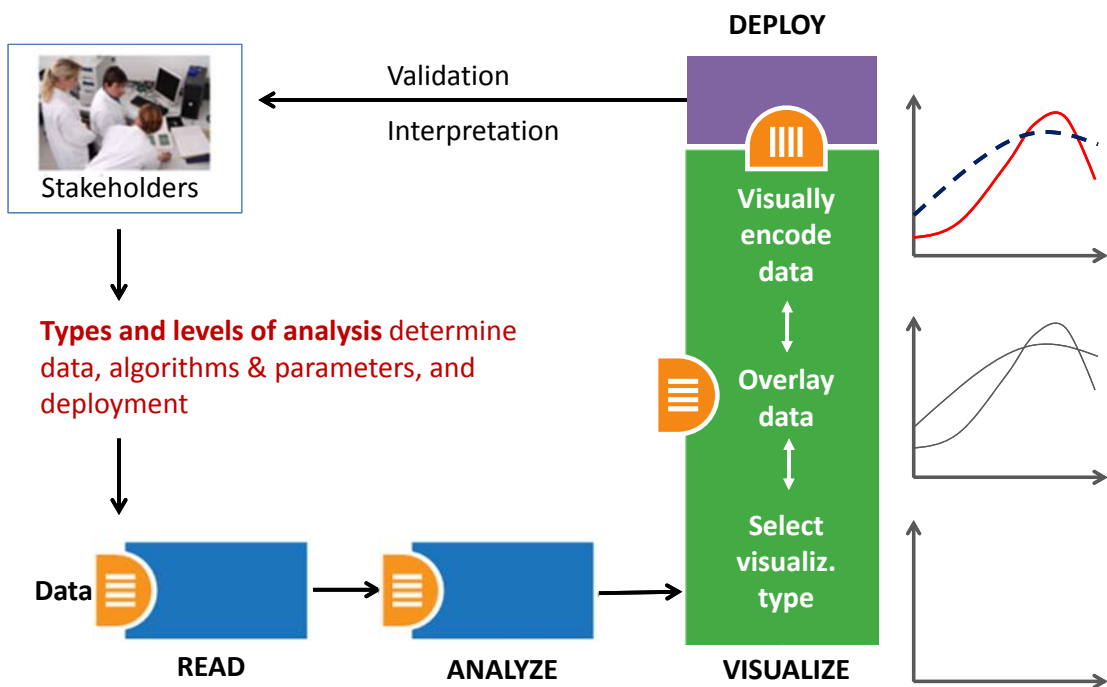
27

# Needs-Driven Workflow Design



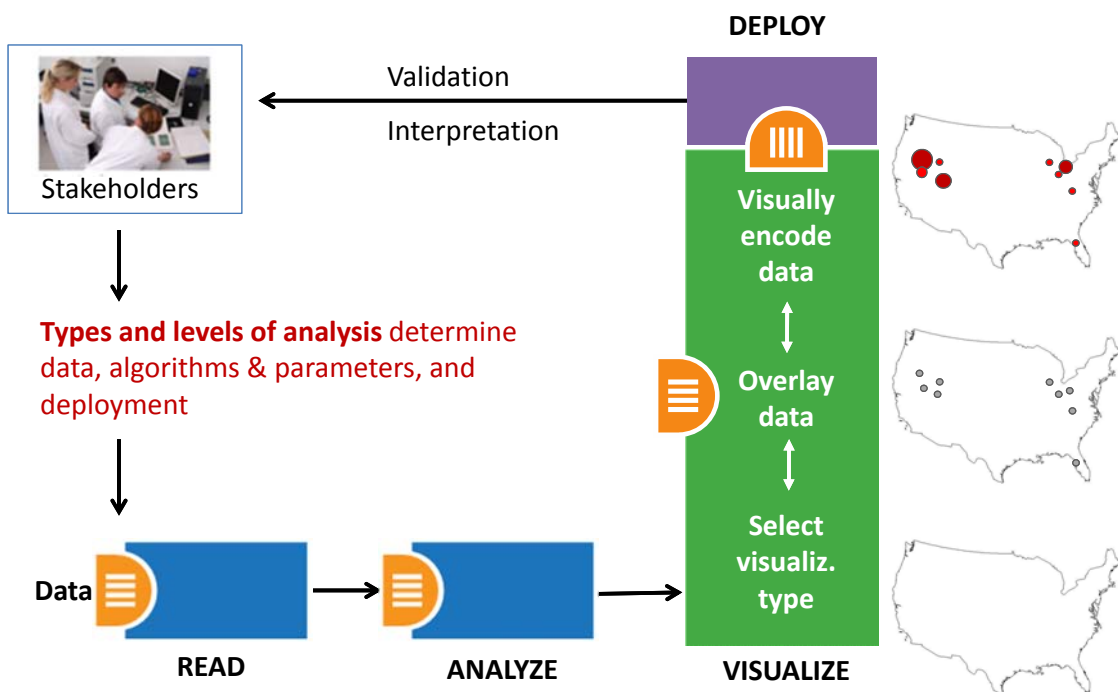
28

# Needs-Driven Workflow Design






29

# Needs-Driven Workflow Design



30




## Visualization Types vs. Data Overlays

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

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

31

## Visualization Types vs. Data Overlays

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

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

32



# 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

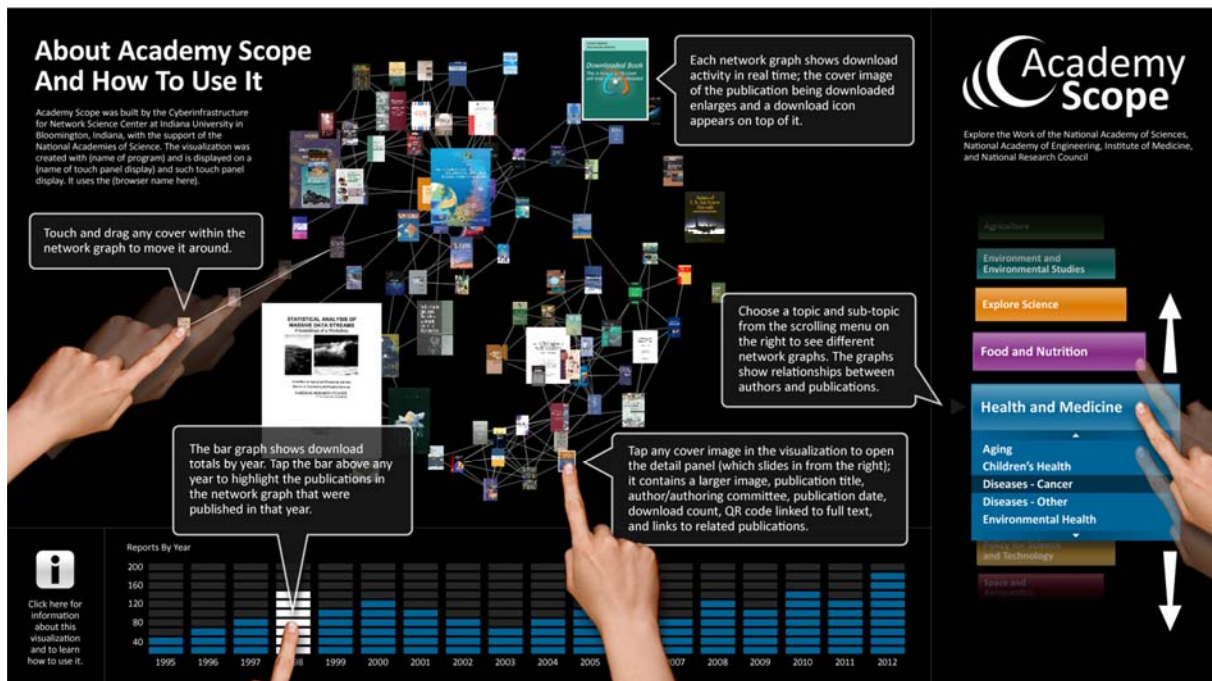
## Students work in teams with clients.

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

## Final Exam

33

# Dynamic Visualizations and Deployment



[http://www.youtube.com/watch?feature=player\\_embedded&v=m\\_TwZXnZrkg](http://www.youtube.com/watch?feature=player_embedded&v=m_TwZXnZrkg)

34

# Clients

Information Visualization MOOC INDIANA UNIVERSITY CNS Twitter Facebook

## List of Clients

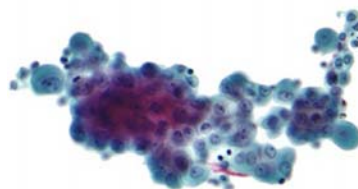
**ISIS** **Project Title:** Isis: 100 Years  
**Client Name:** Jay Malone  
**Project goal/scientific or practical value:** A visual representation Isis' contributors and locales over the past 100 years. Isis is the journal of the History of Science Society. This representation will provide a dynamic 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.

**oycib** **Project Title:** e-Xploration  
**Client Name:** Luyi  
**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](http://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](http://e-crick.net), [oycib.net](http://oycib.net)), I need to approve the results and appear as co-author.

[http://ivmooc.cns.iu.edu/ivmooc\\_clientprojects.html](http://ivmooc.cns.iu.edu/ivmooc_clientprojects.html)

35

[Diogo Carmo](#)



## Mesothelioma

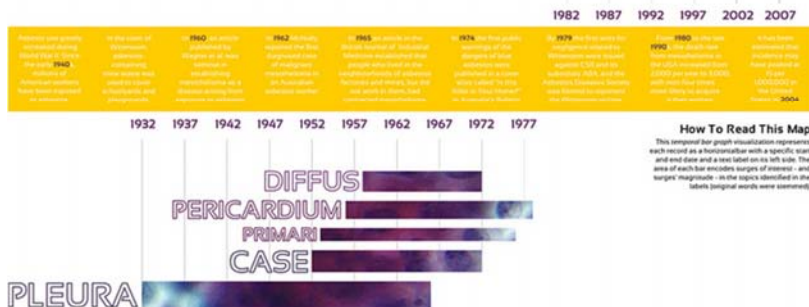
Main title topics in Medline papers

Mesothelioma (or, more precisely, malignant mesothelioma) is a rare form of cancer that develops from transformed cells originating in the mesothelium, the protective lining that covers many of the internal organs of the body. It is usually caused by exposure to asbestos.

The most common anatomical site for the development of mesothelioma is the pleura (the outer lining of the lung and internal chest wall), but it can also arise in the peritoneum (the lining of the abdominal cavity) and the pericardium (the sac that surrounds the heart) or the tunica vaginalis (a sac that surrounds the testis).

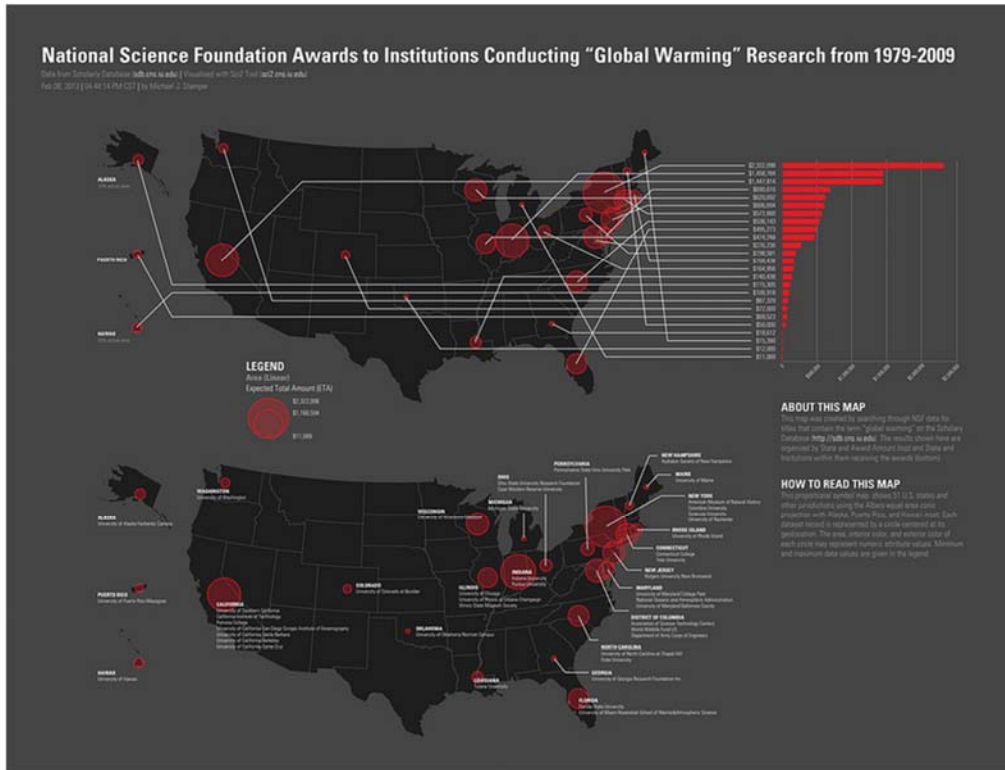
Most people who develop mesothelioma have worked in jobs where they inhaled asbestos, or were exposed to asbestos dust and fibers in other ways. It has also been suggested that smoking, children of a family member who worked with asbestos increases their risk for developing mesothelioma. Unlike lung cancer, there seems to be no association between mesothelioma and tobacco smoking, but smoking greatly increases the risk of other asbestos-induced cancers. Some people who were exposed to asbestos have collected damages for asbestos-related disease, including mesothelioma. Compensation via asbestos funds or class action lawsuits is an important issue in law practices regarding mesothelioma.

MALIGNANT  
PLEURAL  
CYSTIC  
BENIGN  
DIAGNOSIS



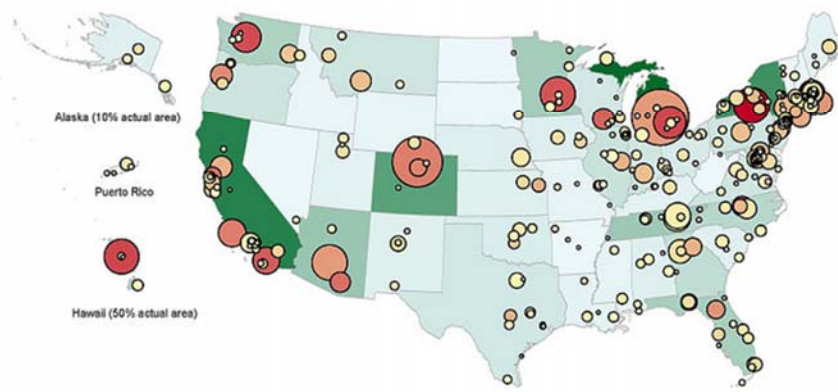
Author Diogo Carmo <http://fabrizio.med@rog.com> Visualization software: Sci2 Tools (2009) Science of Science (Sci2) Team, Indiana University and SciTech Strategies, <http://vis2.cns.iu.edu/> (Disease: Medline Papers, as available in SciSearch Database <http://pubs.ncbi.nlm.nih.gov>) Data and images: Histopathological Mesothelioma article, available at <http://www.histopathology.com/Mesothelioma> (Photo: Scanlon, by Bernd Böhning © 2007 - All Rights Reserved. This font family is licensed and available at <http://bit.ly/8y4m2m>)

36



[mjstamper\\_ivmoo](#)

### NSF Awards 1976-2010 with "ecology" in the title



[Sandra M. Chung](#)



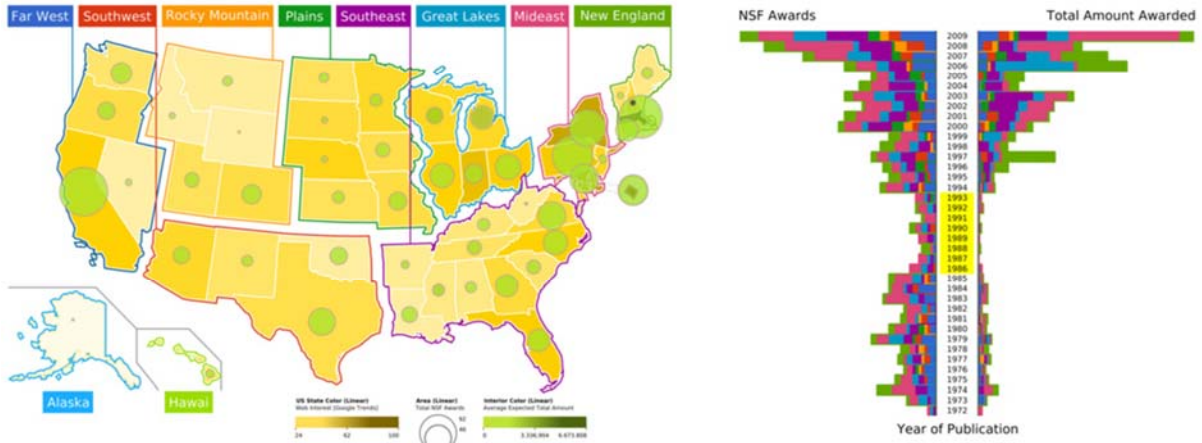
**How to Read this Map**  
 This proportional symbol map shows 52 U.S. states and other jurisdictions using the Albers equal-area conic projection with Alaska, Puerto Rico, and Hawaii inset. Each dataset record is represented by a circle centered at its geolocation. The area, interior color, and exterior color of each circle may represent numeric attribute values. Minimum and maximum data values are given in the legend.

Data retrieved from Scholarly Database (<http://isdb.cns.iu.edu/>).  
 Choropleth generated by Sandra Chung (2013) using "So2".  
 "So2 Team. (2009). Science of Science (So2) Tool. Indiana University and So2Tech Strategies, <http://so2.cns.iu.edu/>.



# Innovation & Entrepreneurship

NSF Funding Across the US, from 1972 to 2009, and Current Web Interest

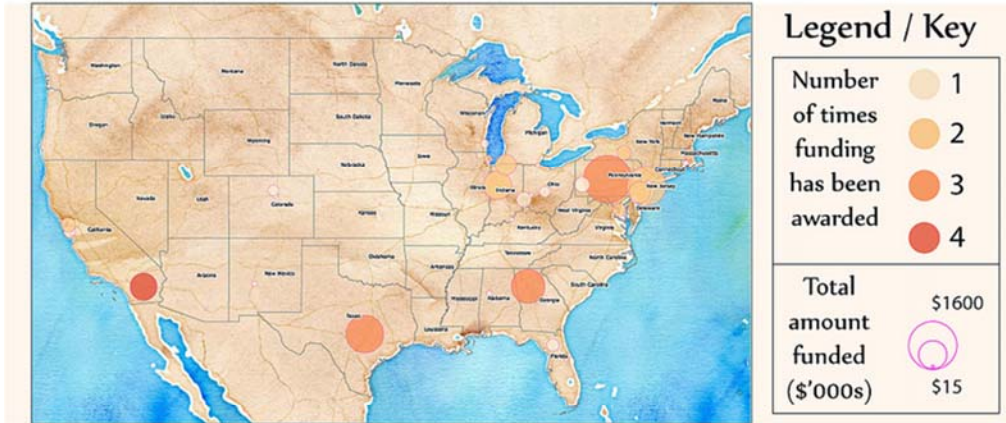


Author: Diogo Carmo <http://diogo.carmo@imim.upenn.edu> | Visualization software: Gird Team (2009), Science of Science (2012) Tool: Indiana University and TuTech Strategies, <http://tu.tech.com/links>; Google Sheets was used to produce the bar graphs | Dataset: National Science Foundation (NSF) Awards, as available in Scholarly Database <http://dx.doi.org/10.1016/j.resonance.2012.08.001>; Time + innovation OR entrepreneurship; Google Trends, for innovation OR entrepreneurship | Photo: Tomaric, by Bernd Montag (2010) | All Rights Reserved. This font family is licensed and is available at <http://fontbundles.net> and, Dribbble, by Blommons, Inc. © 2014. All Rights Reserved. This font family is available at <http://fontbundles.net>

[Diogo Carmo](#)

## NSF Funding - Graphene Projects 2004-2010

#ivmooc Week 3 homework @jonopatterson

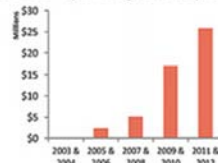


### What is Graphene?

Graphene is a two dimensional material consisting of a single layer of carbon atoms arranged in a honeycomb or chicken wire structure. It is the thinnest material known and yet is also one of the strongest. It conducts electricity as efficiently as copper and outperforms all other materials as a conductor of heat. Graphene is almost completely transparent, yet so dense that even the smallest atom helium cannot pass through it.

Originally thought to be unstable in its free state it proved to be quite the opposite when isolated by Andre Geim and Konstantin Novoselov at the University of Manchester in 2003. The results of this work, which were published in 2004, heralded a new dawn in the study of two dimensional materials and of graphene in particular.

### Total NSF Spending on Graphene



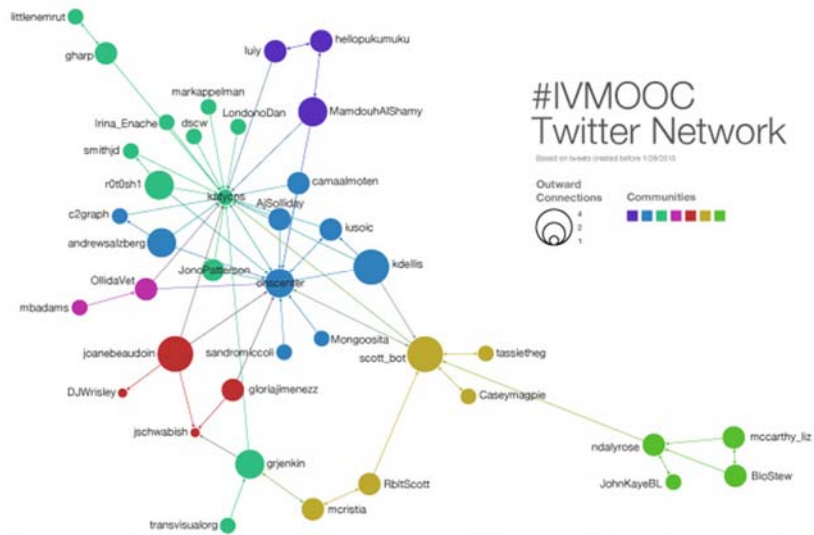
### How to read this map

The map shows NSF funding awards to US institutions. Each circle corresponds to an institution. The depth of colour represents the number of times funding was awarded to the same institution for different projects. The circle size indicates the amount awarded in \$'000s.

Sources: NSF; Manchester University  
Open Street Mapping by CC. Styled by Stamen

[JonoPatterson](#)

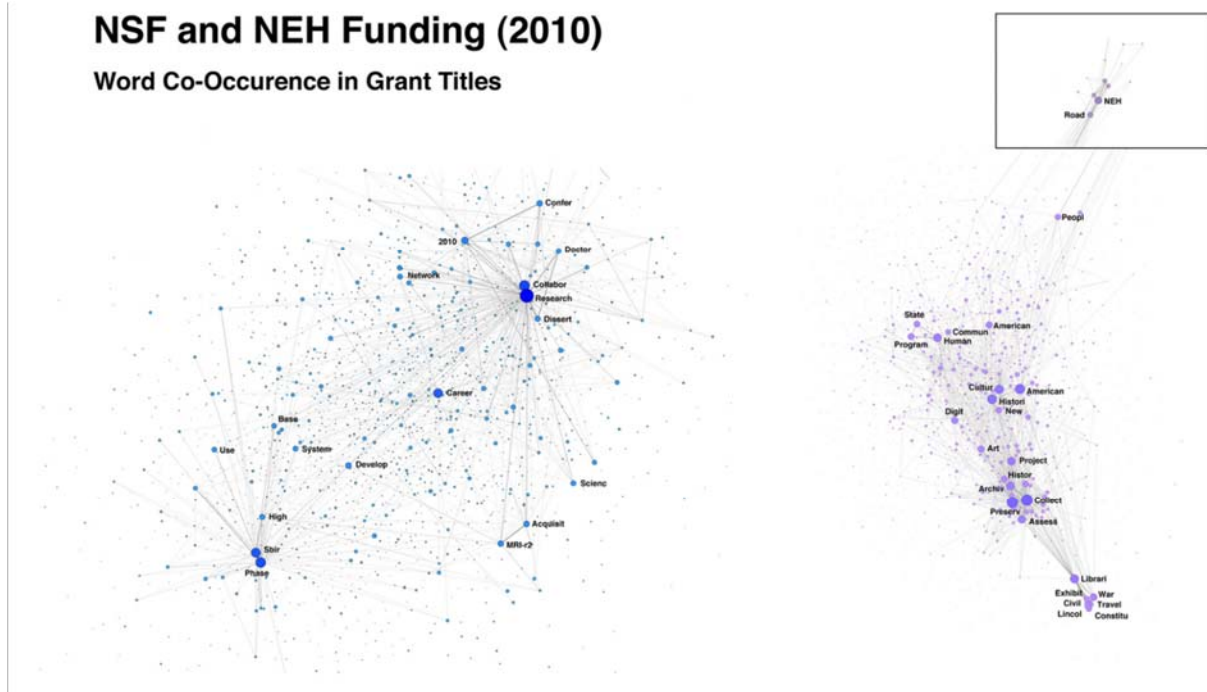




[camaal](#)

## NSF and NEH Funding (2010)

Word Co-Occurrence in Grant Titles



Word Co-Occurrence in NSF and NEH Funded Grant Titles (2010)  
By tzepe | ★ 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

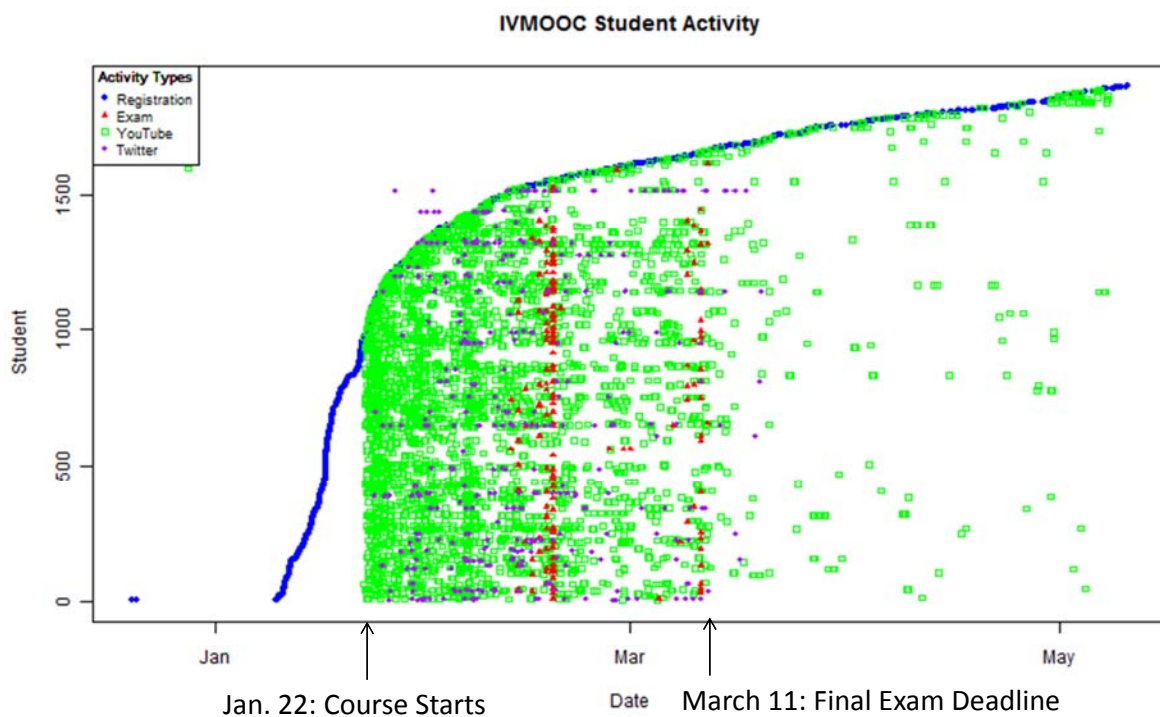
- Empowering **teachers**: How to make sense of the activities of thousands of students? How to guide them?
- Empowering **administrators**: What courses have the highest success rates are most profitable, etc.?
- Supporting **students**: How to navigate learning materials and develop successful learning 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:

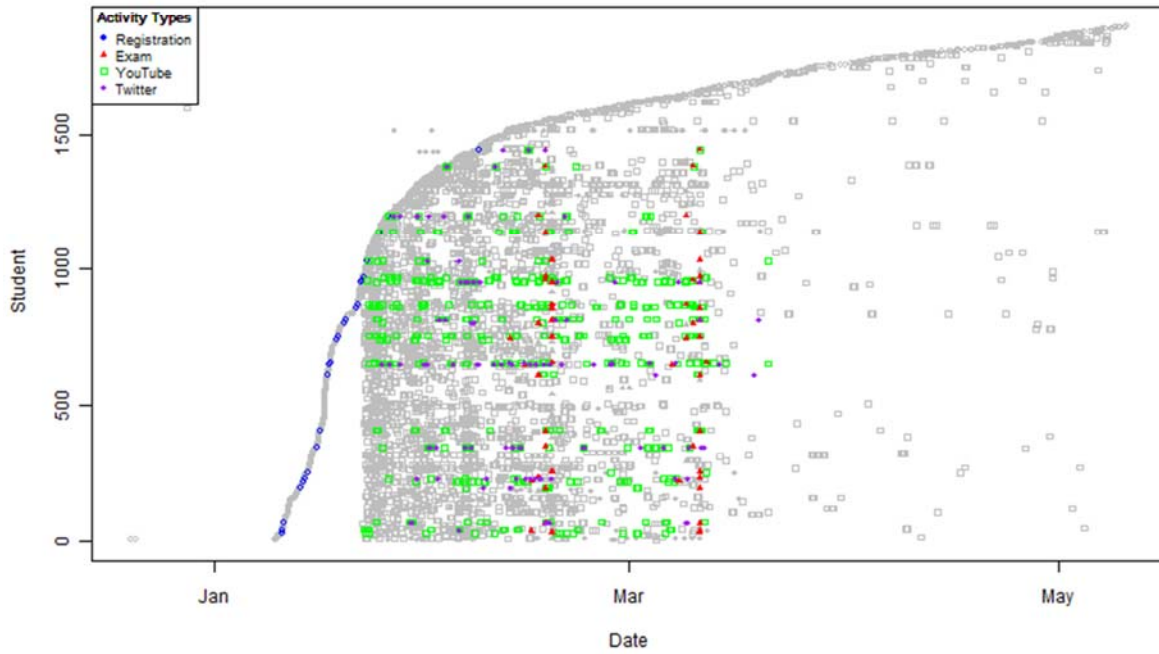
- 1,901 students registered via GCB (1215 male/557 female)
- 52,557 slide downloads from our server
- 18,893 video views via YouTube
- 193 accounts made 730 tweets
- 134 students took 183 exams in GCB
- 674 remarks on 215 different forum threads in Drupal
- 64 students submitted projects via Drupal

# Visualizing IVMOOC Data



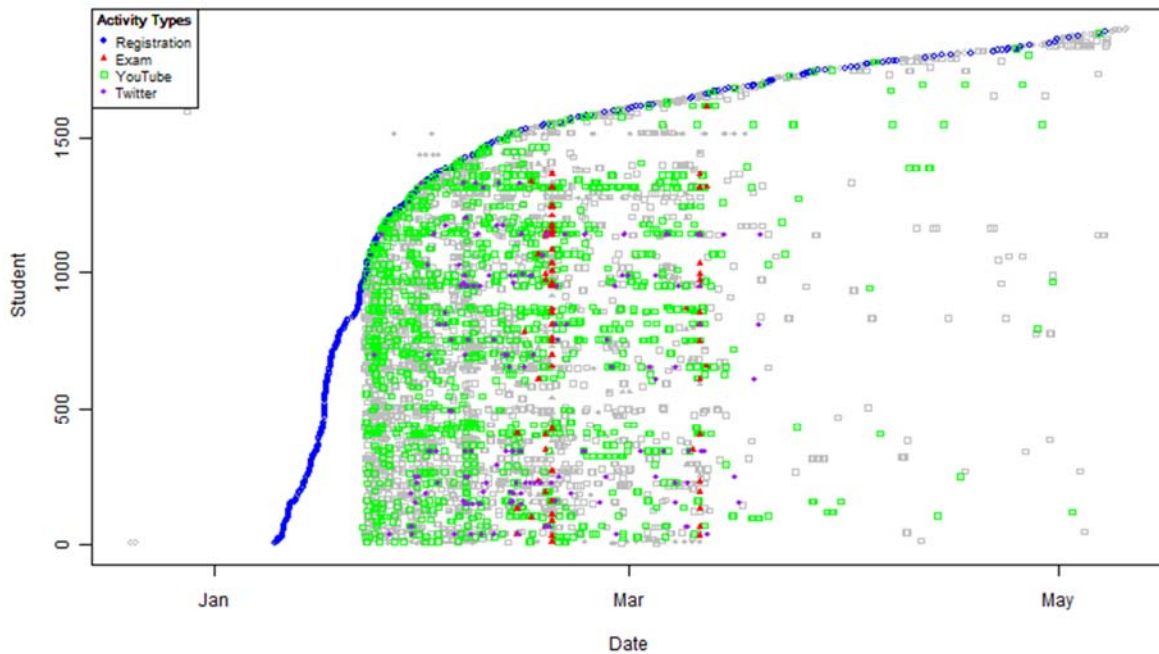


### IVMOOC Student Activity (Achievement Badge)



1215 male students  
557 female students

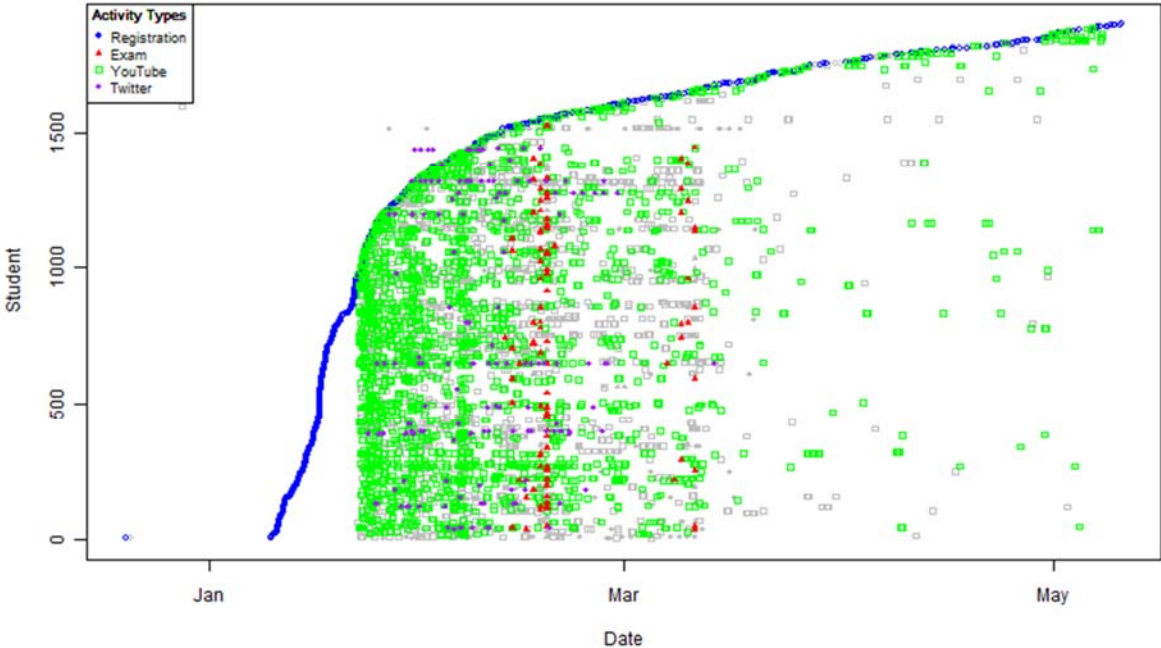
### Female IVMOOC Student Activity



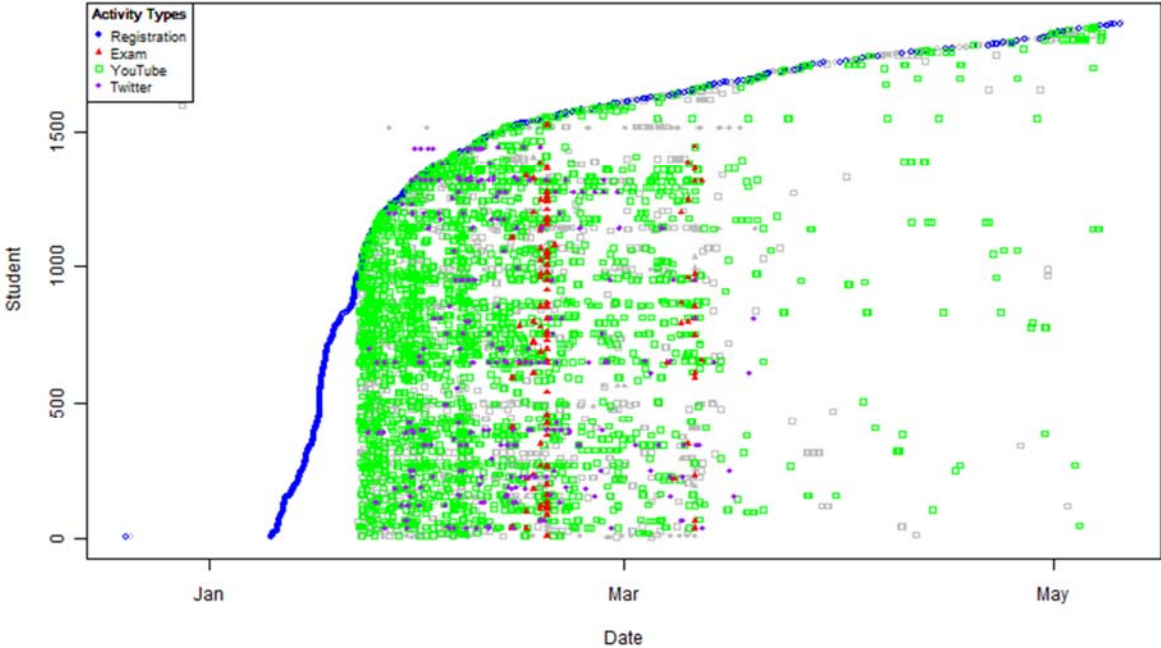


1215 male students  
557 female students

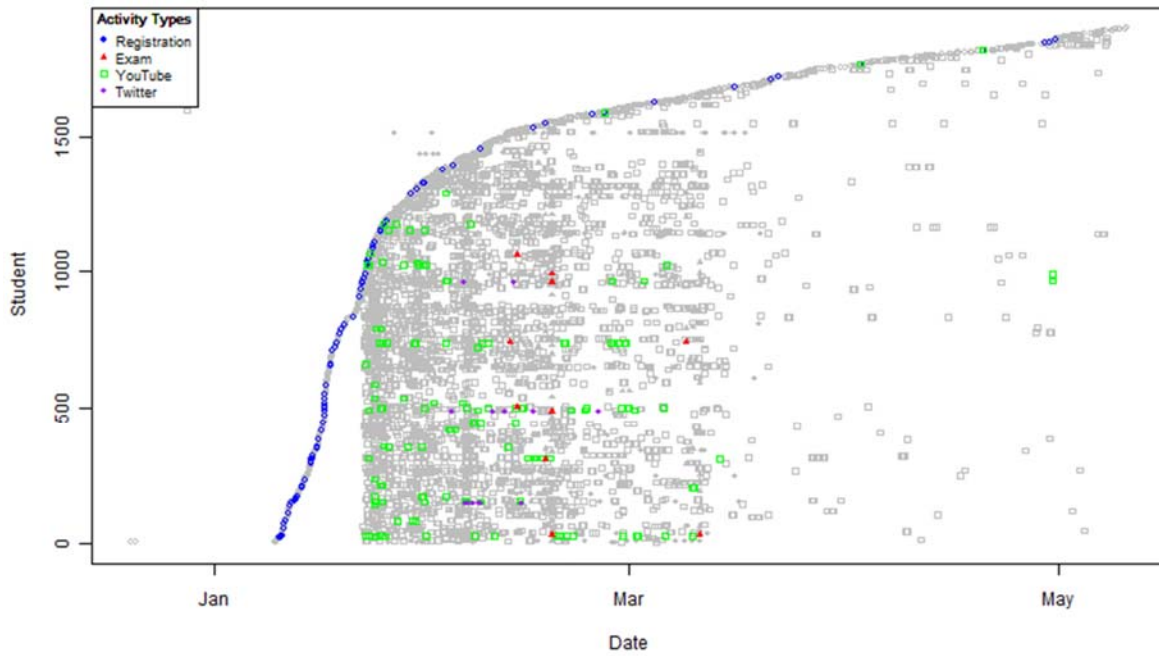
Male IVMOOC Student Activity



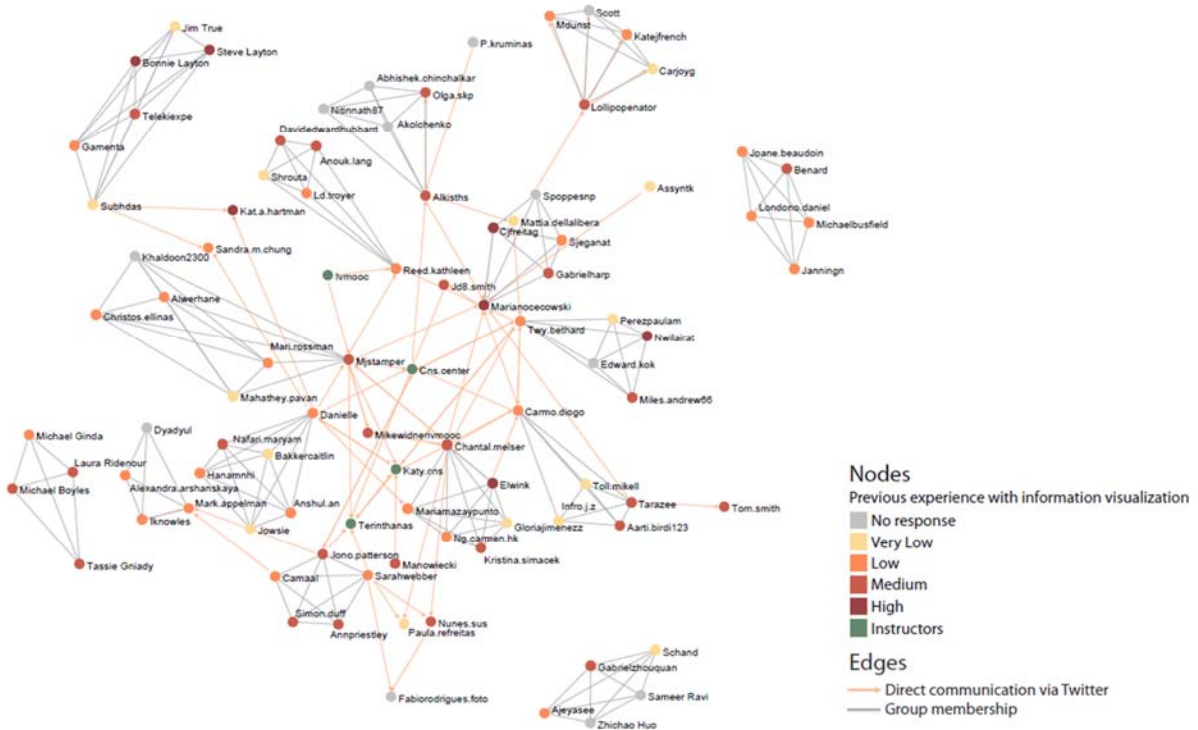
Novice IVMOOC Student Activity



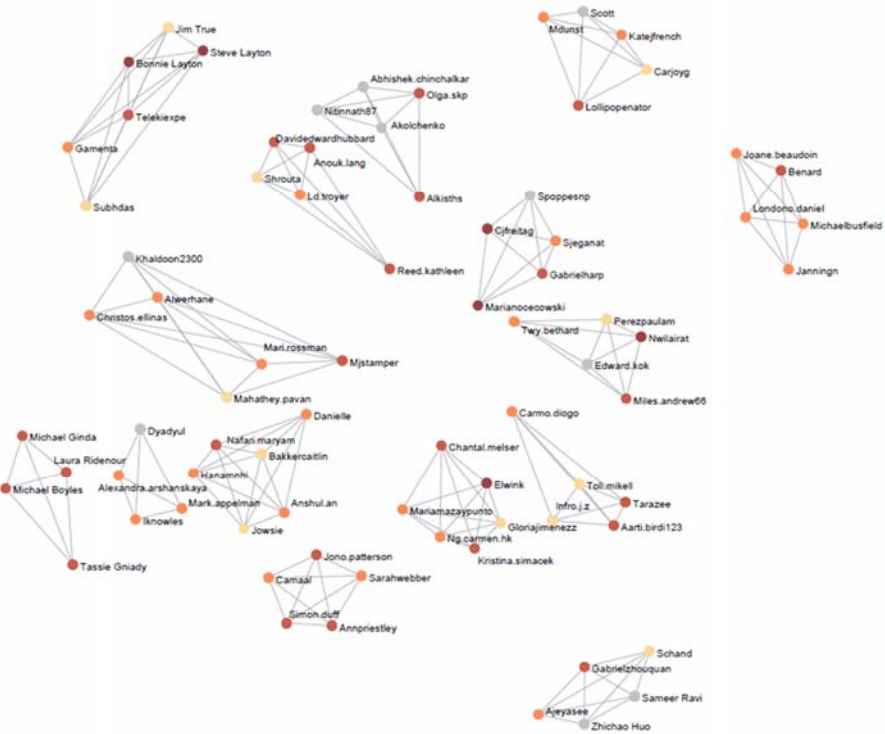
### Expert IVMOOC Student Activity



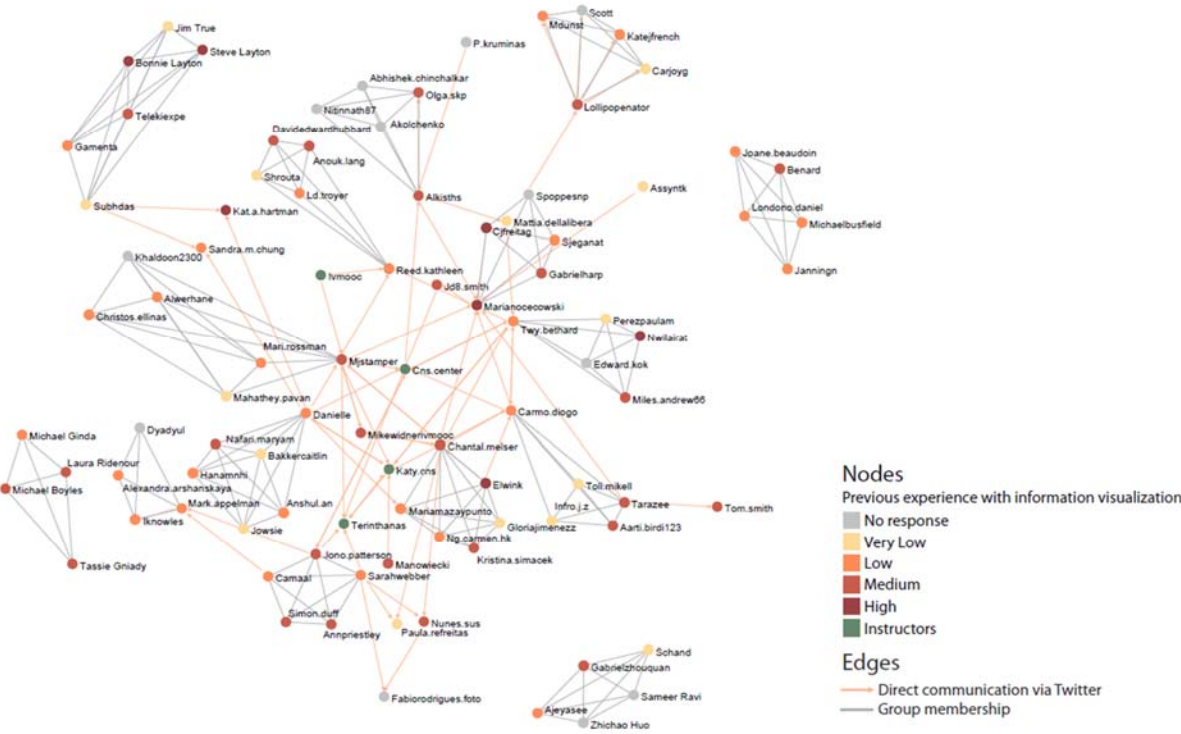
### Student Client Projects: All Interactions



# Student Client Projects: Group Memberships



# Student Client Projects: All Interactions



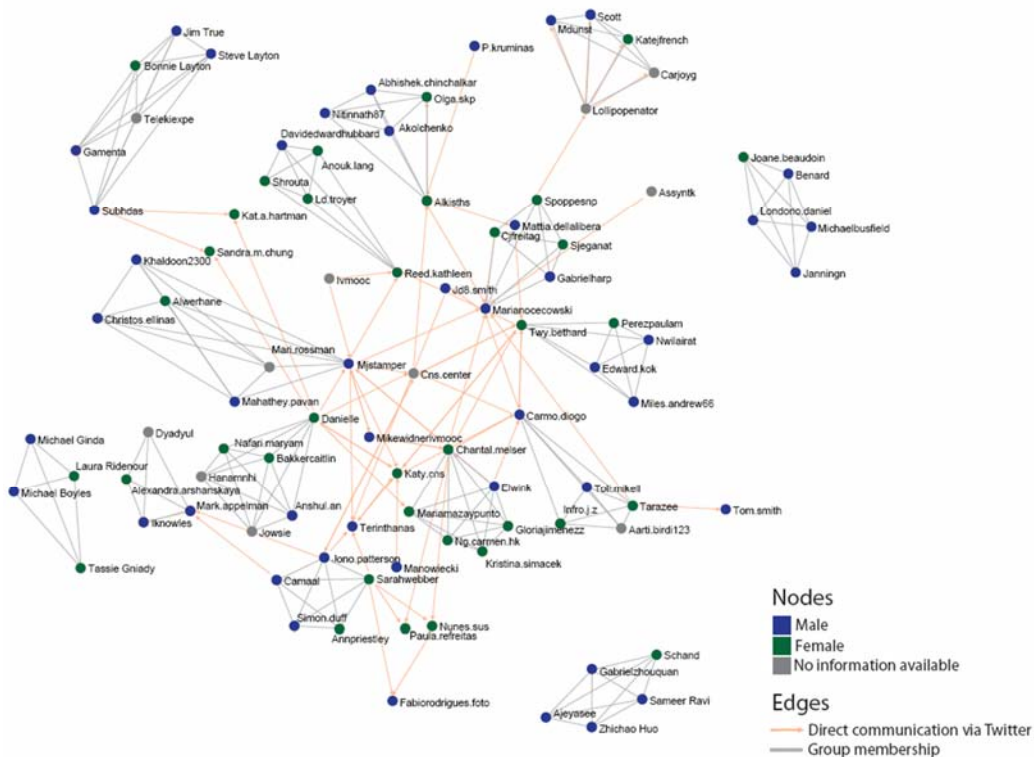
**Nodes**  
 Previous experience with information visualization

- Grey circle: No response
- Yellow circle: Very Low
- Orange circle: Low
- Red circle: Medium
- Dark red circle: High
- Green circle: Instructors

**Edges**

- Orange line: Direct communication via Twitter
- Grey line: Group membership

## Student Client Projects: Gender



55

## 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?

56



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/](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>