VIVO DATA ANALYSIS & VISUALIZATION

How to Program, Extend and Utilize

Instructors: Micah Linnemeier, Chintan Tank, Nianli Ma, and Katy Börner
Cyberinfrastructure for Network Science Center, Indiana University

Location: Break Out C **Time:** 8 AM - 12 PM

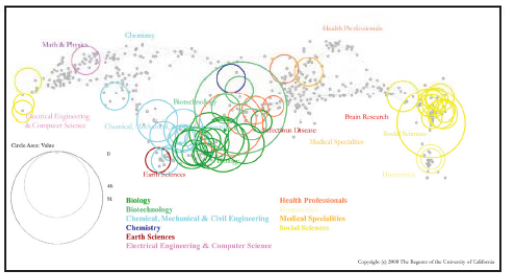
Through the VIVO project, high quality academic data from systems of record becomes available in a common format through Semantic Web technologies. Data that was previously difficult to access and combine becomes available to anyone, creating a unique opportunity for academic and industry stakeholders to utilize this data in conjunction with their own areas of expertise. The most compelling uses of VIVO data might come from 3rd-party developers creating analyses, applications, and services that meet their specific needs.

This hands-on workshop aims to empower participants to understand, access, and utilize VIVO data for administrative, commercial, or research purposes. It starts with a brief overview of techniques and workflows used to analyze and visualize temporal, geospatial, topical, and network datasets at a micro, meso, and macro level. Emphasis is on the design of insightful visualizations. Next, we will present the general VIVO architecture and explain and demonstrate different options to access and work with VIVO data and to use or extend VIVO code drawing on Indiana University's experience with VIVO service development. All services and applications are documented at a level of detail that makes it easy for others to replicate and extend them.

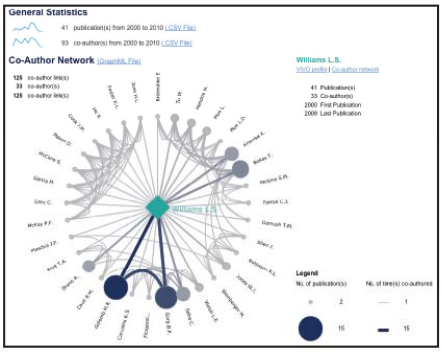
VIVO Team: Cornell University: Dean Krafft (Cornell PI), Manolo Bevia, Jim Blake, Nick Cappadona, Brian Caruso, Jon Corson-Rikert, Elly Cramer, Medha Devare, John Fereira, Brian Lowe, Stella Mitchell, Holly Mistlebauer, Anup Sawant, Christopher Westling, Rebecca Younes. **University of Florida:** Mike Conlon (VIVO and UF PI), Cecilia Botero, Kerry Britt, Erin Brooks, Amy Buhler, Ellie Bushhousen, Chris Case, Valrie Davis, Nita Ferree, Chris Haines, Rae Jesano, Margeaux Johnson, Sara Kreinest, Yang Li, Paula Markes, Sara Russell Gonzalez, Alexander Rockwell, Nancy Schaefer, Michele R. Tennant, George Hack, Chris Barnes, Narayan Raum, Brenda Stevens, Alicia Turner, Stephen Williams. **Indiana University**: Katy Borner (IU PI), William Barnett, Shanshan Chen, Ying Ding, Russell Duhon, Jon Dunn, Micah Linnemeier, Nianli Ma, Robert McDonald, Barbara Ann O'Leary, Mark Price, Yuyin Sun, Alan Walsh, Brian Wheeler, Angela Zoss. **Ponce School of Medicine:** Richard Noel (Ponce PI), Ricardo Espada, Damaris Torres. **The Scripps Research Institute:** Gerald Joyce (Scripps PI), Greg Dunlap, Catherine Dunn, Brant Kelley, Paula King, Angela Murrell, Barbara Noble, Cary Thomas, Michaeleen Trimarchi. **Washington University, St. Louis**: Rakesh Nagarajan (WUSTL PI), Kristi L. Holmes, Sunita B. Koul, Leslie D. McIntosh. **Weill Cornell Medical College:** Curtis Cole (Weill PI), Paul Albert, Victor Brodsky, Adam Cheriff, Oscar Cruz, Dan Dickinson, Chris Huang, Itay Klaz, Peter Michelini, Grace Migliorisi, John Ruffing, Jason Specland, Tru Tran, Jesse Turner, Vinay Varughese.

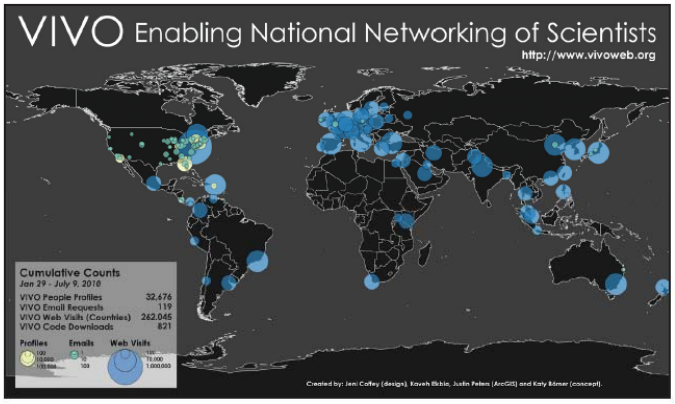
Last but not least, we will showcase different data analyses and visualizations of VIVO data at the individual, institution, and national level such as:

- **Individual level.** Statistics and ego-centric scholarly networks on VIVO Profile pages.
- **Institutional level.** Analyses and visualizations of funding intake and publication output for departments and centers accessible via the VIVO Index page. Download of relevant data in tabular and network formats for further analysis using the Network Workbench tool.
- **National level.** Visualization of VIVO installations and their profile holdings together with web page access and general VIVO information requests. Plus, services that use VIVO URIs to access data across different VIVO instances.



The workshop concludes with a general question and answer session.





Workshop Attendees

Registered by Aug. 4, 2010

	A	B	C	D
1	**NAME**	**TITLE**	**INSTITUTION / AFFILIATION**	**EMAIL**
2	Bertuzzi, Stefano	Health Science Policy Analyst	NIH	bertuzzs@nih.gov
3	De Vine, Lance Neil	Research Support Specialist	Queensland University of Technology	l.devine@qut.edu.au
4	Drumm, Sharon D.	Staff Officer	USDA	sharon.drumm@ars.usda.gov
5	Gomes, Edward D.	Sr. Associate Dean	Duke University	edward.gomes@duke.edu
6	Henderson, Courtney E.	Administrative Director - CTSI	Brown University	courtney_henderson@brown.edu
7	Lougee, Robin	IBM Research	IBM	robinlh@us.ibm.com
8	Mons, Barend	Scientific Director Support and External Relations	NBIC	barendmons@gmail.com
9	Porter, Simon	Information Manager	University of Melbourne	simon.porter@unimelb.edu.au
10	Pye, John M.	Program Manager, eResearch	US Forest Service	jpye@fs.fed.us
11	Ratner, Howard	CTO, Executive Vice President	Nature Publishing Group	h.ratner@us.nature.com
12	Schneggenburger, Mark N.		SUNY Buffalo	mschnegg@buffalo.edu
13	Spadaro, John	Director, Technical Architecture & Outreach	Brown University	john_spadaro@brown.edu

Workshop Schedule – Part I

Social Network Visualizations

- Visualization Types and Levels
- Exemplary User Needs
- Proposed VIVO Visualizations

15 min break

Accessing and Using VIVO Data

- VIVO Architecture
- Accessing VIVO data
- Developing for VIVO
- How we use VIVO data: The Visualization Pipeline

Workshop Schedule – Part II

Visualization Details

- Co-author visualization
- Sparklines
- Data download and analysis

Hands-On

- Creating and Executing SPARQL queries
- Creating visualizations based on SPARQL query results

Outlook

- Institution Level Visualizations Under Development
- National Level Visualizations

Q&A

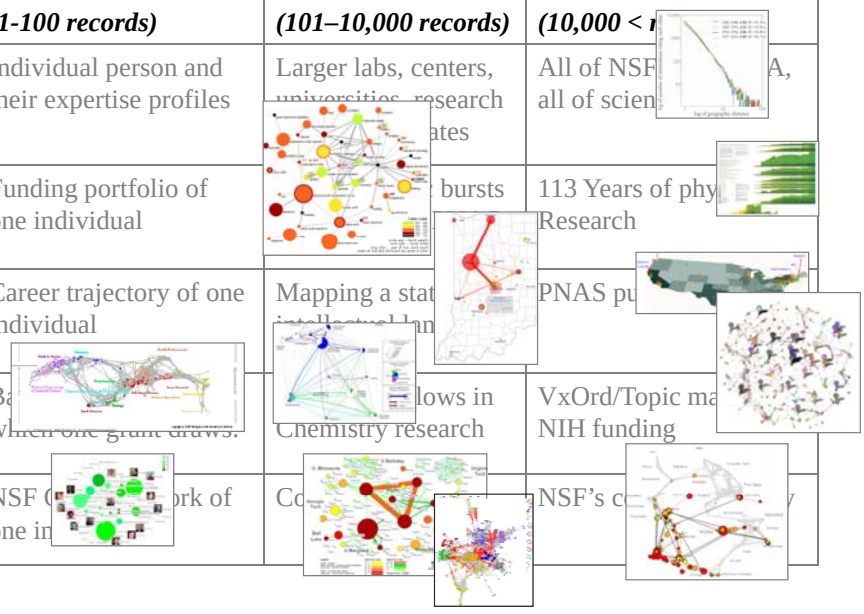
Social Network Visualizations

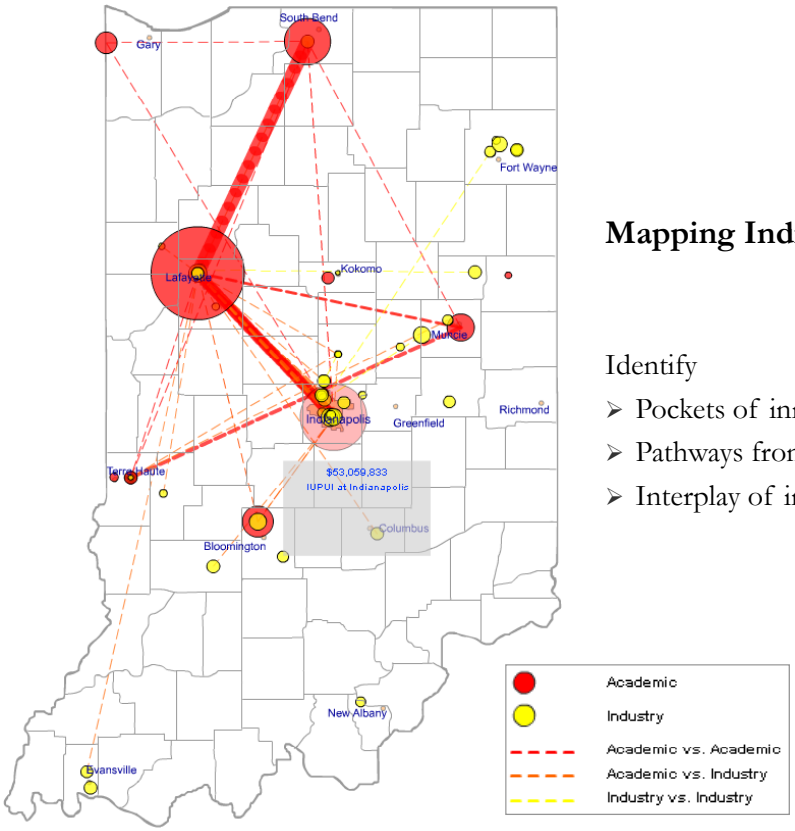
- Visualization Types and Levels
- Exemplary User Needs
- Proposed VIVO Visualizations

Type of Analysis vs. Scale of 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 states 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	NSF's core competency

Type of Analysis vs. Scale of Level of Analysis

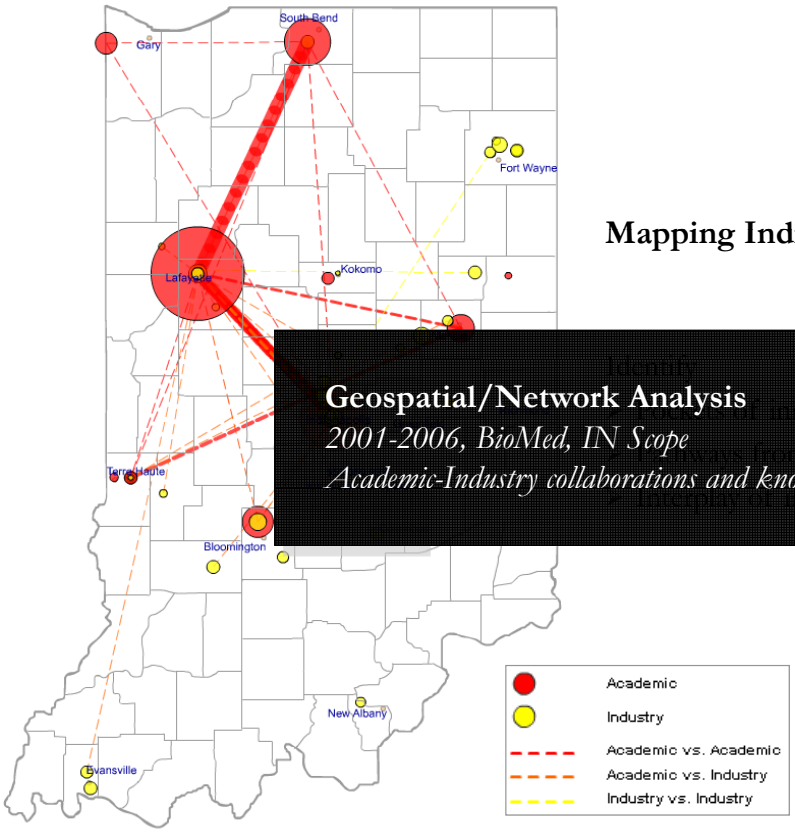




Mapping Indiana's Intellectual Space

Identify

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



Mapping Indiana's Intellectual Space

Geospatial/Network Analysis

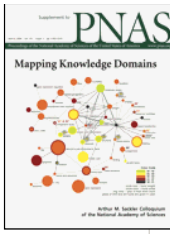
2001-2006, BioMed, IN Scope

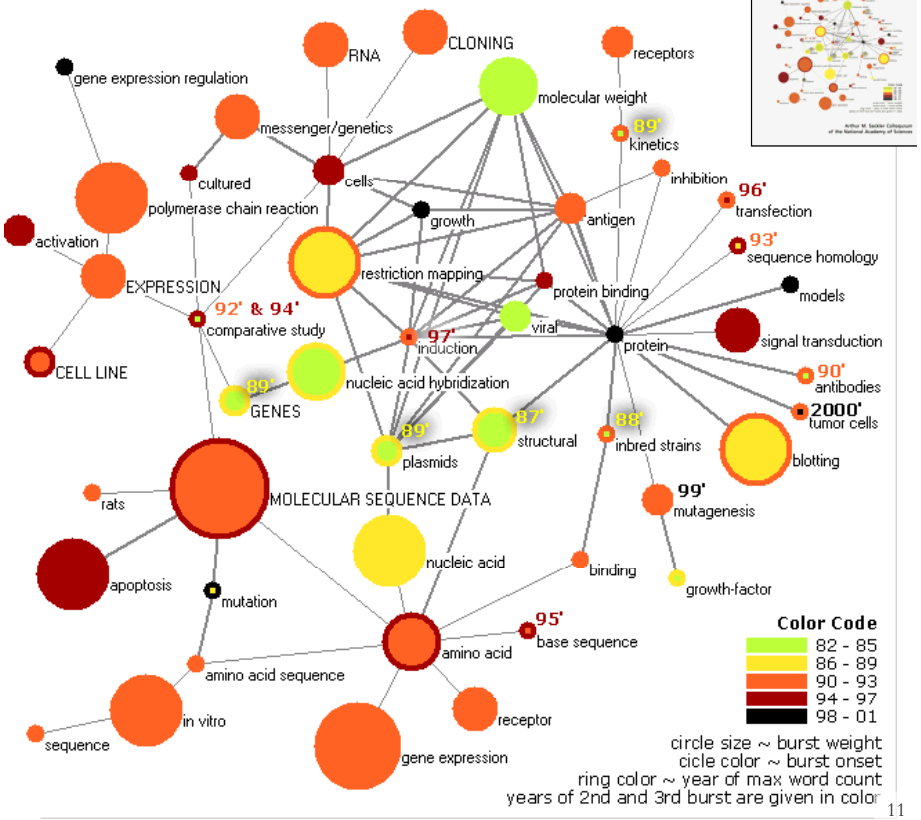
Academic-Industry collaborations and knowledge diffusion

Mapping Topic Bursts

Co-word space of the top 50 highly frequent and bursty words used in the top 10% most highly cited PNAS publications in 1982-2001.

Mane & Börner. (2004) PNAS, 101(Suppl. 1): 5287-5290.

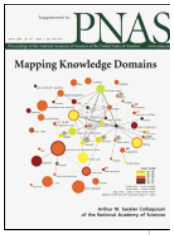


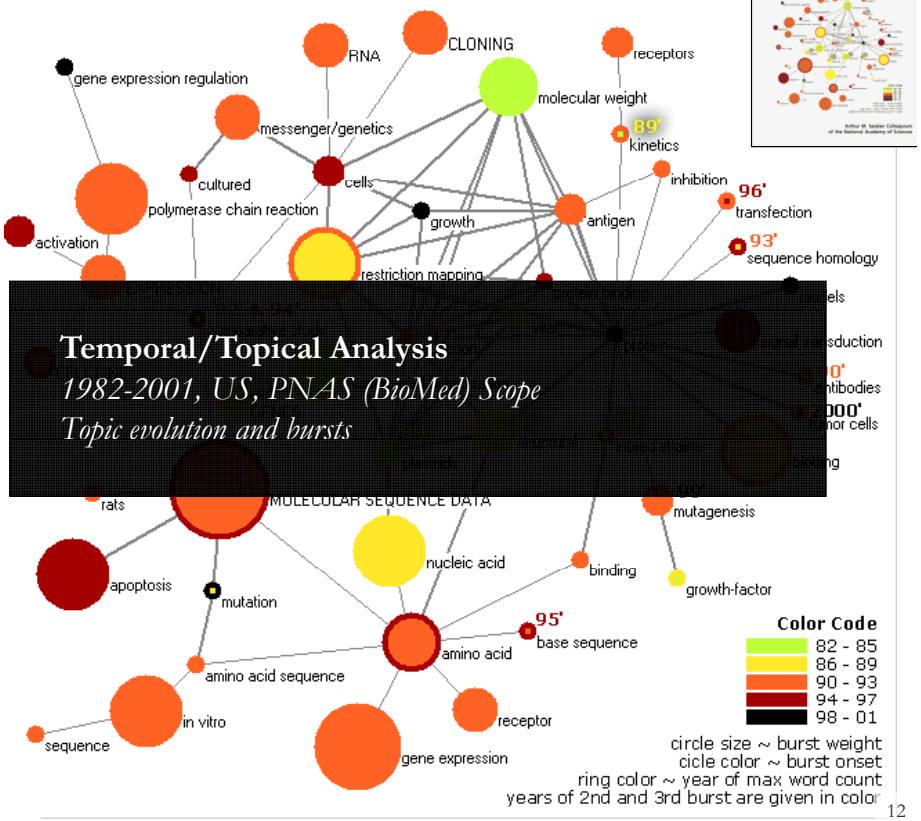


Mapping Topic Bursts

Co-word space of the top 50 highly frequent and bursty words used in the top 10% most highly cited PNAS publications in 1982-2001.

Mane & Börner. (2004) PNAS, 101(Suppl. 1): 5287-5290.



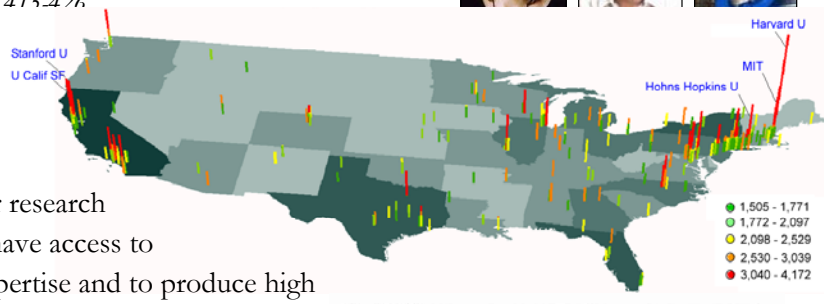


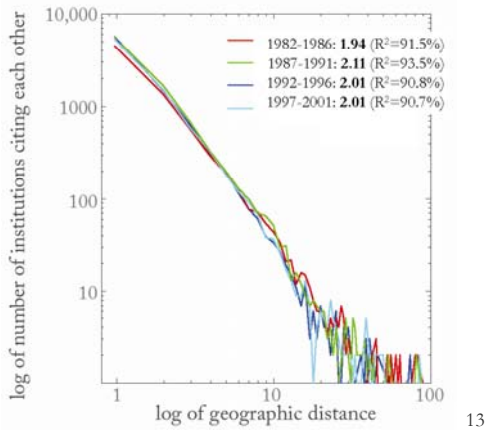
Spatio-Temporal Information Production and Consumption of Major U.S. Research Institutions

Börner, Katy, Penumarthy, Shashikant, Meiss, Mark and Ke, Weimao. (2006) Mapping the Diffusion of Scholarly Knowledge Among Major U.S. Research Institutions. Scientometrics. 68(3), pp. 415-426.

Research questions:

1. Does space still matter in the Internet age?
2. Does one still have to study and work at major research institutions in order to have access to high quality data and expertise and to produce high quality research?
3. Does the Internet lead to more global citation patterns, i.e., more citation links between papers produced at geographically distant research instructions?





Contributions:

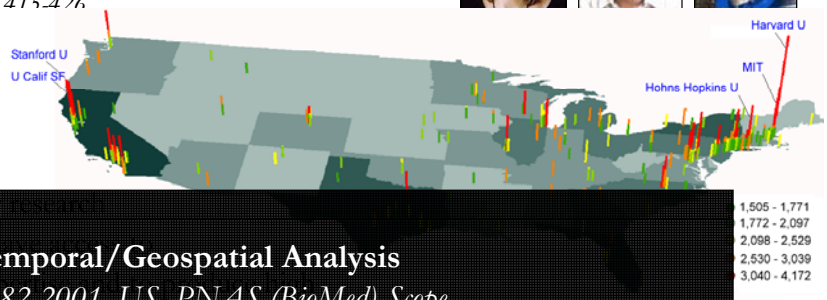
- Answer to Qs 1 + 2 is YES.
- Answer to Qs 3 is NO.
- Novel approach to analyzing the dual role of institutions as information producers and consumers and to study and visualize the diffusion of information among them.

Spatio-Temporal Information Production and Consumption of Major U.S. Research Institutions

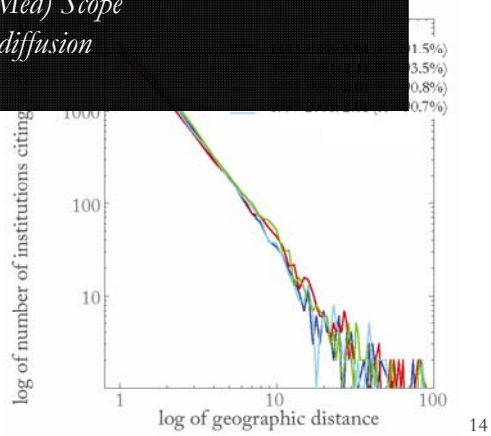
Börner, Katy, Penumarthy, Shashikant, Meiss, Mark and Ke, Weimao. (2006) Mapping the Diffusion of Scholarly Knowledge Among Major U.S. Research Institutions. Scientometrics. 68(3), pp. 415-426.

Research questions:

1. Does space still matter in the Internet age?
2. Does one still have to study and work at major research institutions in order to have access to high quality data and expertise and to produce high quality research?
3. Does the Internet lead to more global citation patterns, i.e., more citation links between papers produced at geographically distant research instructions?



Temporal/Geospatial Analysis
1982-2001, US, PNAS (BioMed) Scope
Citation impact and knowledge diffusion



Contributions:

- Answer to Qs 1 + 2 is YES.
- Answer to Qs 3 is NO.
- Novel approach to analyzing the dual role of institutions as information producers and consumers and to study and visualize the diffusion of information among them.

Research Collaborations by the Chinese Academy of Sciences

By Weixia (Bonnie) Huang, Russell J. Duhon, Elisha F. Hardy, Katy Börner, Indiana University, USA

This map highlights the research co-authorship collaborations of the Chinese Academy of Sciences with locations in China and countries around the world. The large geographic map shows the research collaborations of all CAS institutes. Each smaller geographic map shows the research collaborations by the CAS researchers in one province-level administrative division. Collaborations between CAS researchers are not included in the data. On each map, locations are colored on a logarithmic scale by the number of collaborations from red to yellow. The darkest red is 3,395 collaborations by all of CAS with researchers in Beijing. Also, flow lines are drawn from the location of focus to all locations collaborated with. The width of the flow line is linearly proportional to the number of collaborations with the locations it goes to, with the smallest flow lines representing one collaboration and the largest representing differing amounts on each geographic map.

Research Collaborations by the Chinese Academy of Sciences

By Weixia (Bonnie) Huang, Russell J. Duhon, Elisha F. Hardy, Katy Börner, Indiana University, USA

Geospatial Analysis

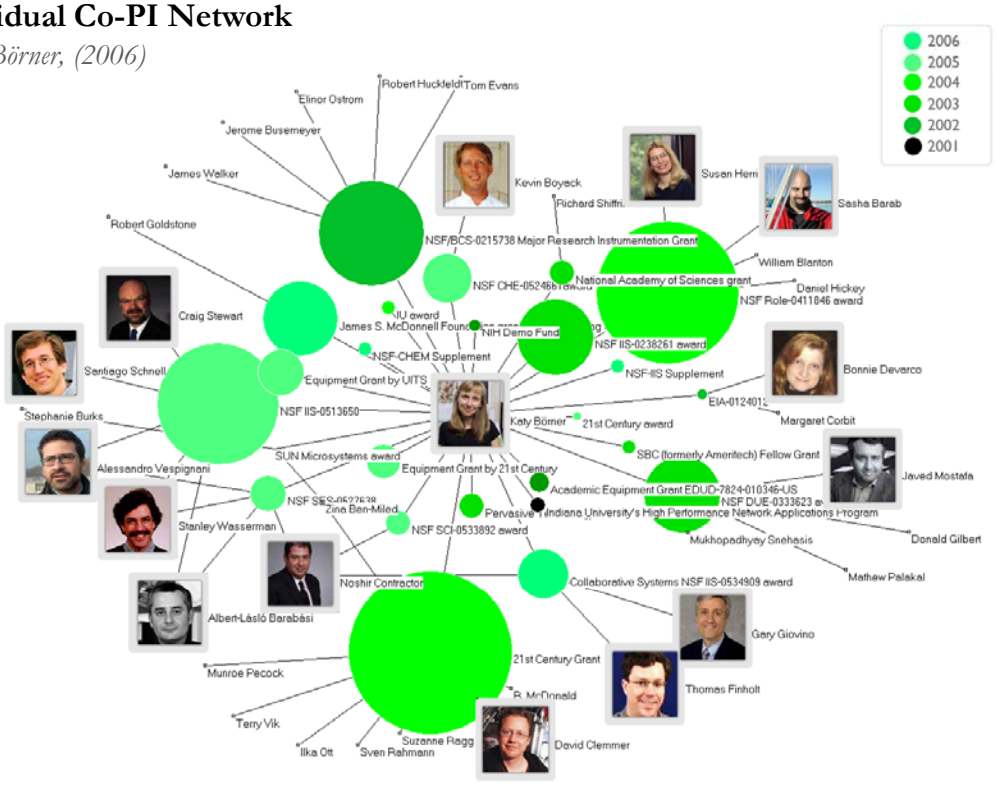
World, Chinese Academy of Science

Collaboration and knowledge diffusion via co-author networks

This map highlights the research co-authorship collaborations of the Chinese Academy of Sciences with locations in China and countries around the world. The large geographic map shows the research collaborations of all CAS institutes. Each smaller geographic map shows the research collaborations by the CAS researchers in one province-level administrative division. Collaborations between CAS researchers are not included in the data. On each map, locations are colored on a logarithmic scale by the number of collaborations from red to yellow. The darkest red is 3,395 collaborations by all of CAS with researchers in Beijing. Also, flow lines are drawn from the location of focus to all locations collaborated with. The width of the flow line is linearly proportional to the number of collaborations with the locations it goes to, with the smallest flow lines representing one collaboration and the largest representing differing amounts on each geographic map.

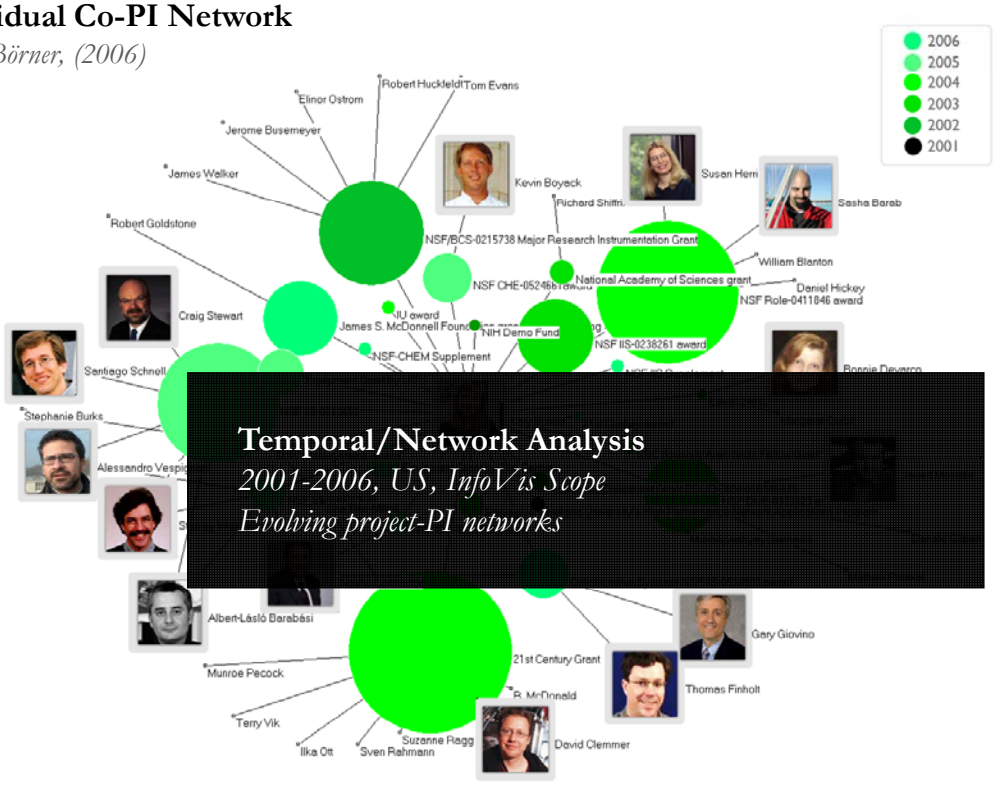
Individual Co-PI Network

Ke & Börner, (2006)



Individual Co-PI Network

Ke & Börner, (2006)



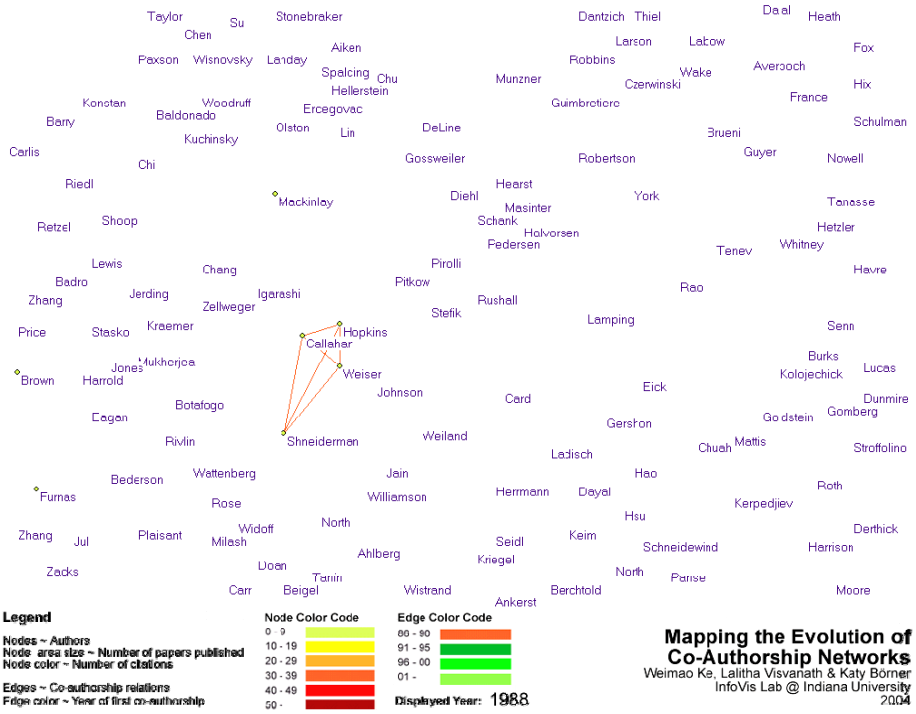
Temporal/Network Analysis

2001-2006, US, InfoVis Scope

Evolving project-PI networks

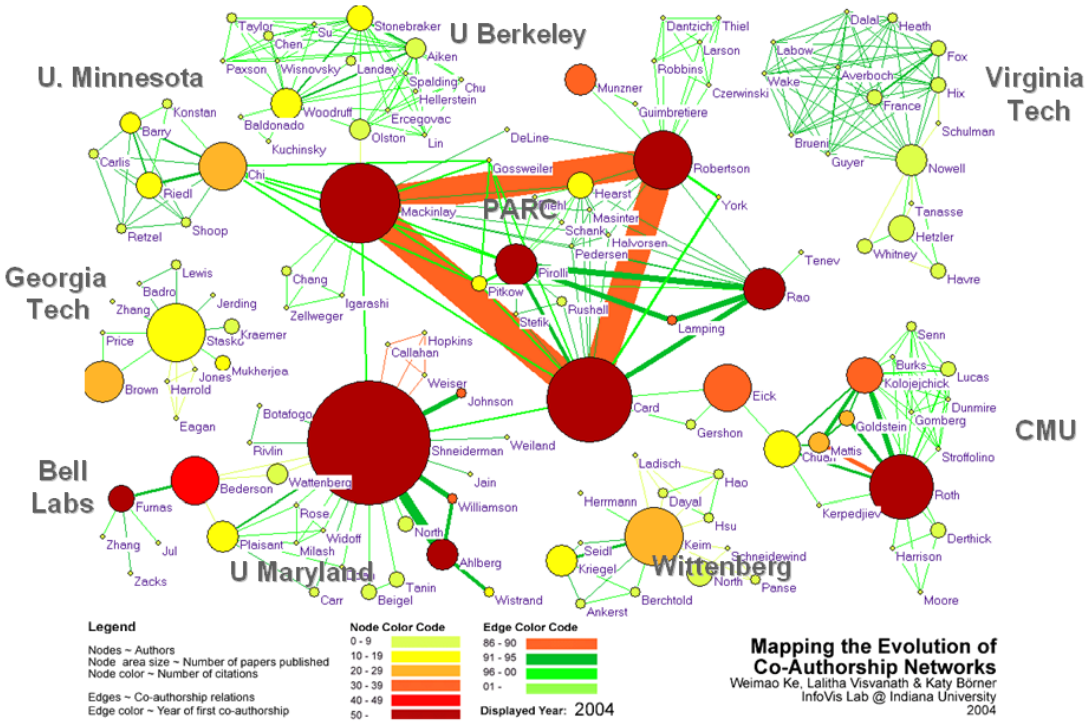
Mapping the Evolution of Co-Authorship Networks

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



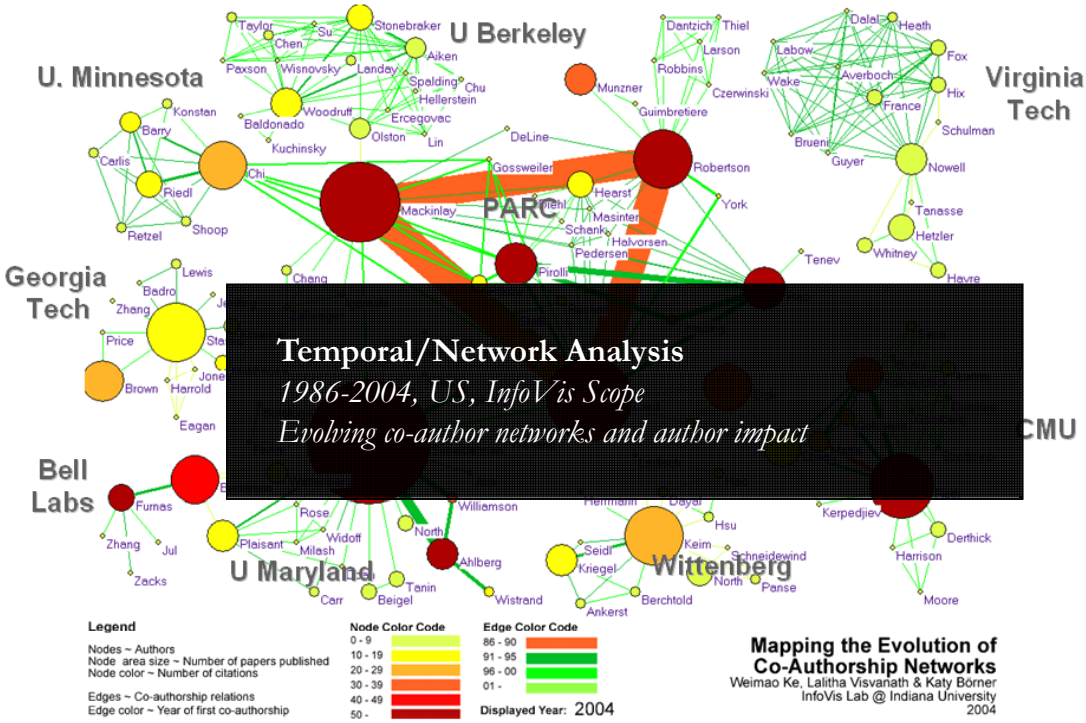
Mapping the Evolution of Co-Authorship Networks

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



Mapping the Evolution of Co-Authorship Networks

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



Studying the Emerging Global Brain: Analyzing and Visualizing the Impact of Co-Authorship Teams

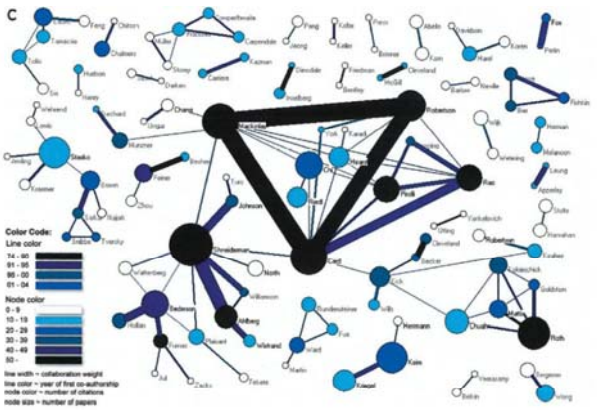
Börner, Dall'Asta, Ke & Vespignani (2005) Complexity, 10(4):58-67.

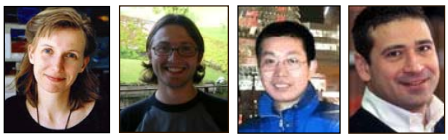
Research question:

- Is science driven by prolific single experts or by high-impact co-authorship teams?

Contributions:

- New approach to allocate citational credit.
- Novel weighted graph representation.
- Visualization of the growth of weighted co-author network.
- Centrality measures to identify author impact.
- Global statistical analysis of paper production and citations in correlation with co-authorship team size over time.
- Local, author-centered entropy measure.





Studying the Emerging Global Brain: Analyzing and Visualizing the Impact of Co-Authorship Teams

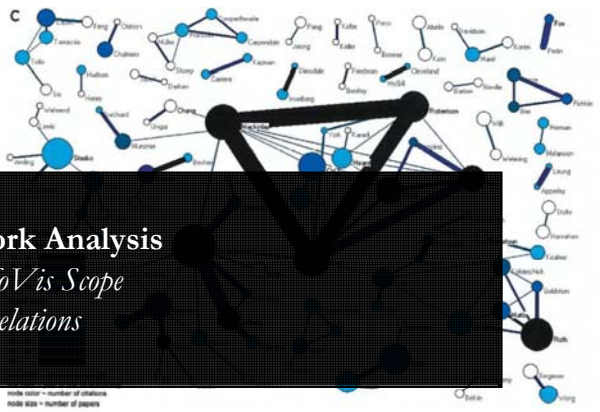
Börner, Dall'Asta, Ke & Vespignani (2005) Complexity, 10(4):58-67.

Research question:

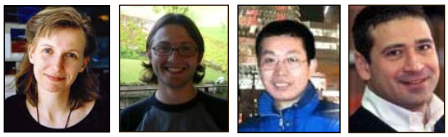
- Is science driven by prolific single experts or by high-impact co-authorship teams?

Contributions:

- New approach to allocate citational credit.
- Novel weighted gr
- Visualization of th co-author network
- Centrality measure impact.
- Global statistical analysis of paper production and citations in correlation with co-authorship team size over time.
- Local, author-centered entropy measure.



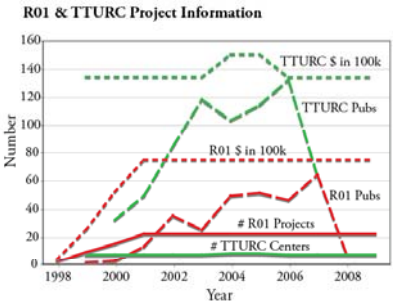
Temporal/Network Analysis
1986-2004, US, InfoVis Scope
Impact of co-author relations

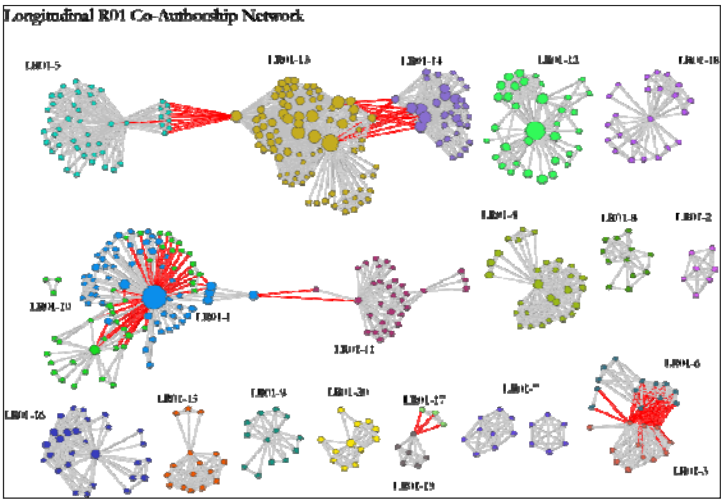


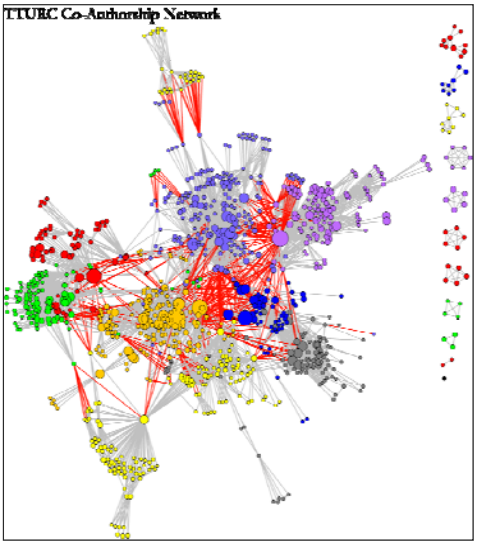
Mapping Transdisciplinary Tobacco Use Research Centers Publications

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

Zoss & Börner, forthcoming.



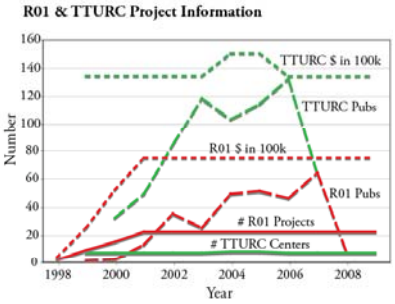


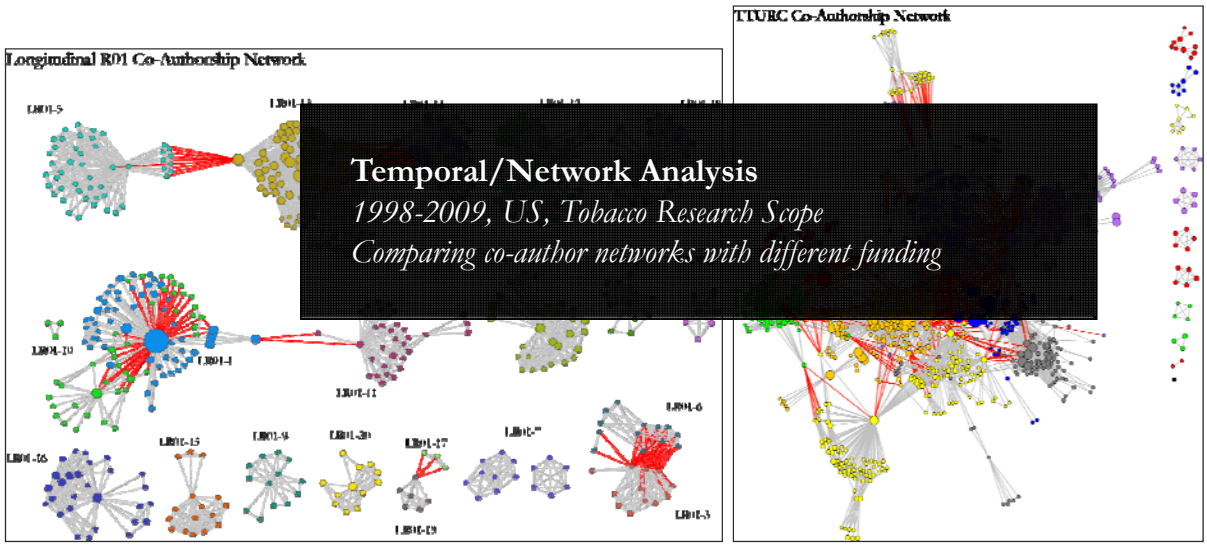


Mapping Transdisciplinary Tobacco Use Research Centers Publications

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

Zoss & Börner, forthcoming.





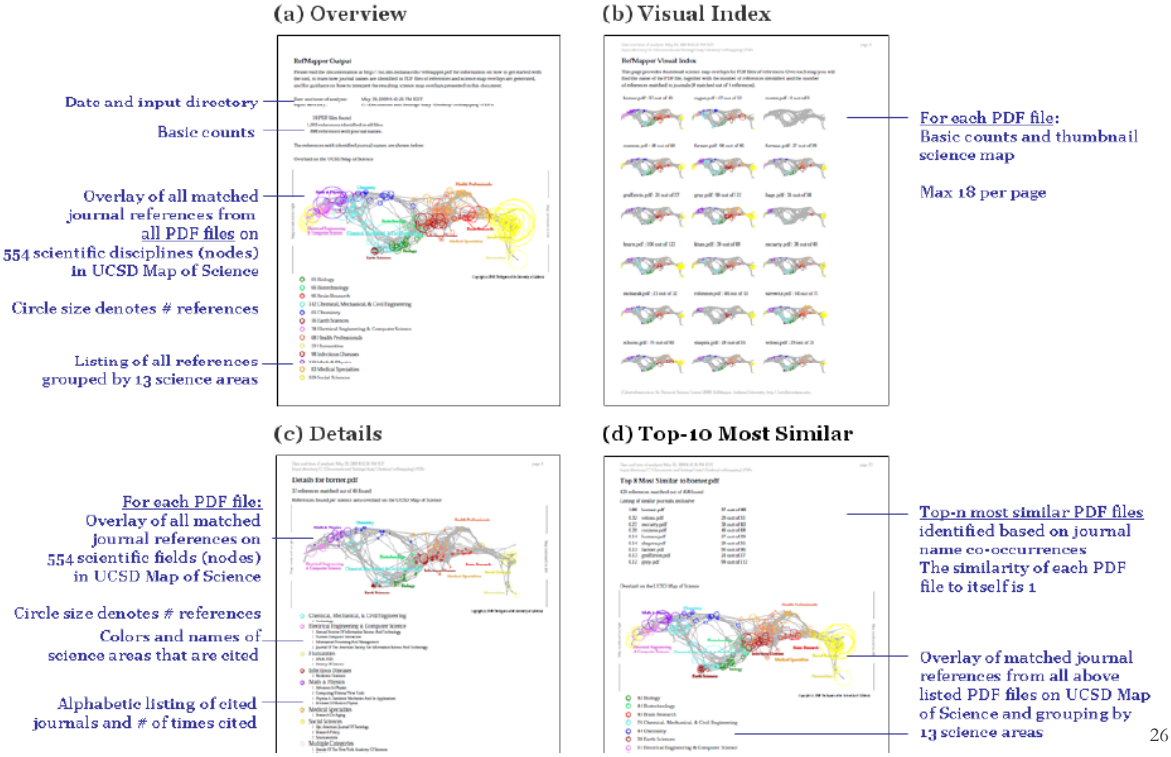
Temporal/Network Analysis

1998-2009, US, Tobacco Research Scope

Comparing co-author networks with different funding

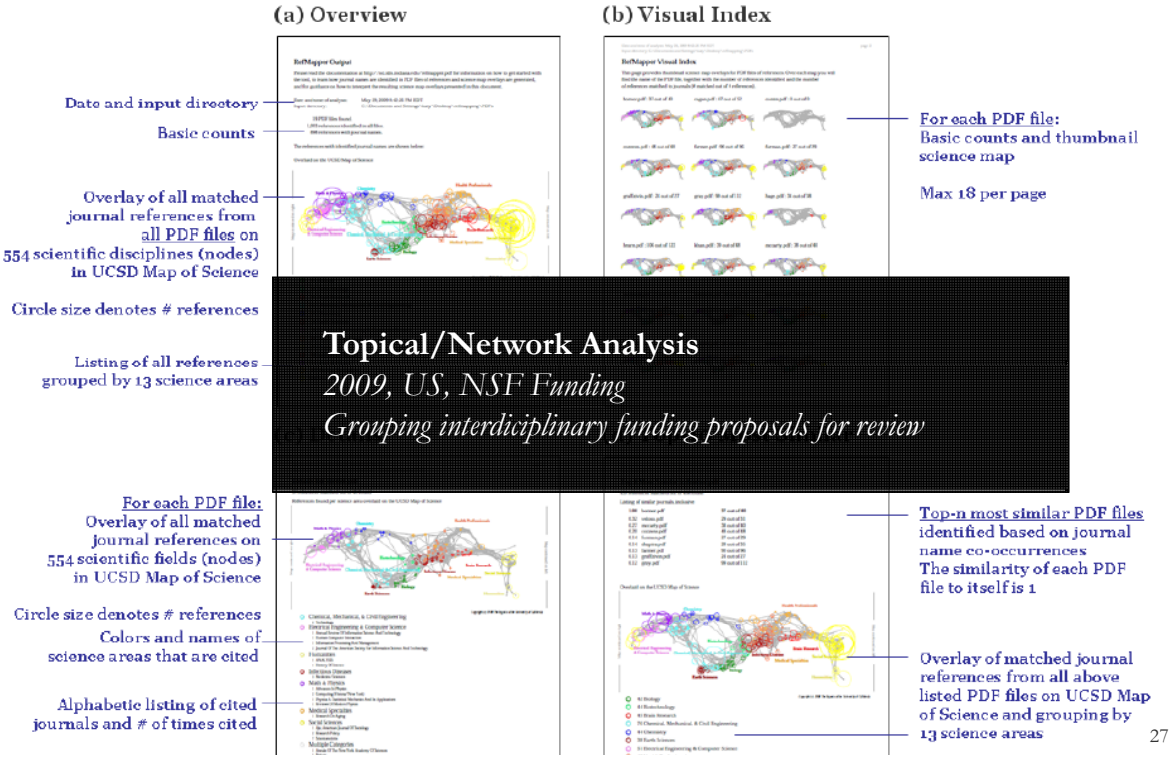
Reference Mapper

Duhon & Börner, forthcoming.



Reference Mapper

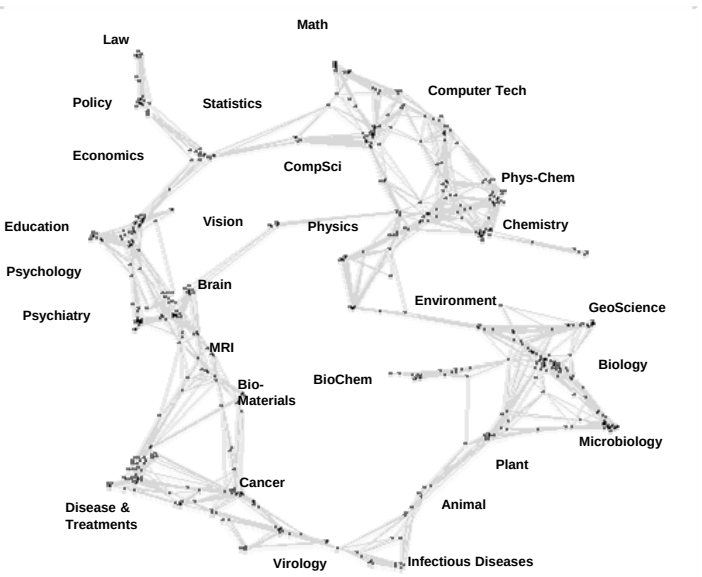
Duhon & Börner, forthcoming.



Latest 'Base Map' of Science

Kevin W. Boyack, Katy Börner, & Richard Klavans (2007). Mapping the Structure and Evolution of Chemistry Research. 11th International Conference on Scientometrics and Informetrics. pp. 112-123.

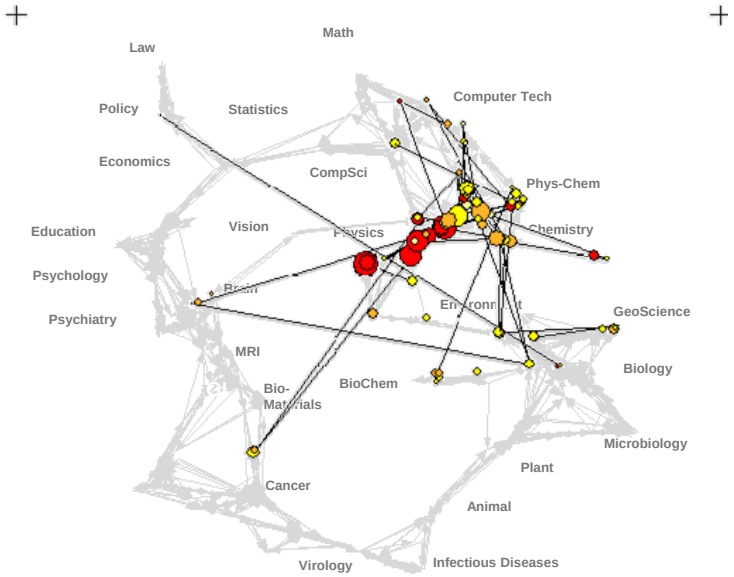
- Uses combined SCI/SSCI from 2002
  - 1.07M papers, 24.5M references, 7,300 journals
  - Bibliographic coupling of papers, aggregated to journals
- Initial ordination and clustering of journals gave 671 clusters
- Coupling counts were reaggregated at the journal cluster level to calculate the
  - (x,y) positions for each journal cluster
  - by association, (x,y) positions for each journal



Science map applications: Identifying core competency

Kevin W. Boyack, Katy Börner, & Richard Klavans (2007).

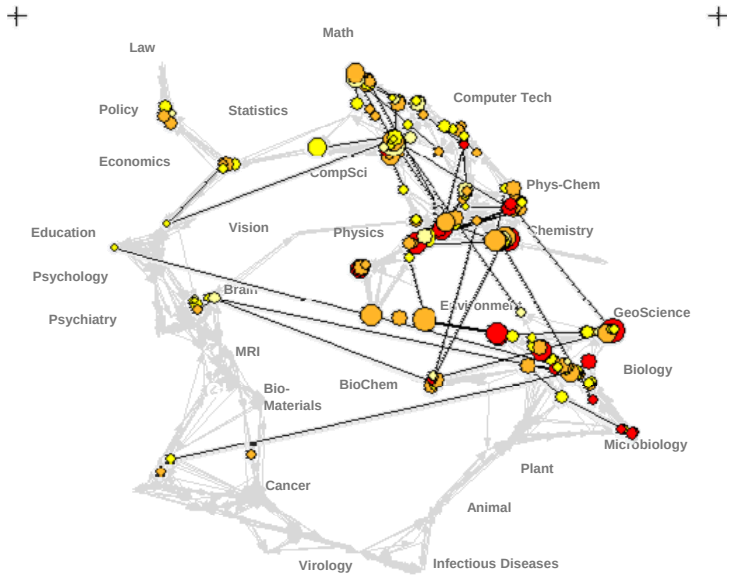
Funding patterns of the US Department of Energy (DOE)



Science map applications: Identifying core competency

Kevin W. Boyack, Katy Börner, & Richard Klavans (2007).

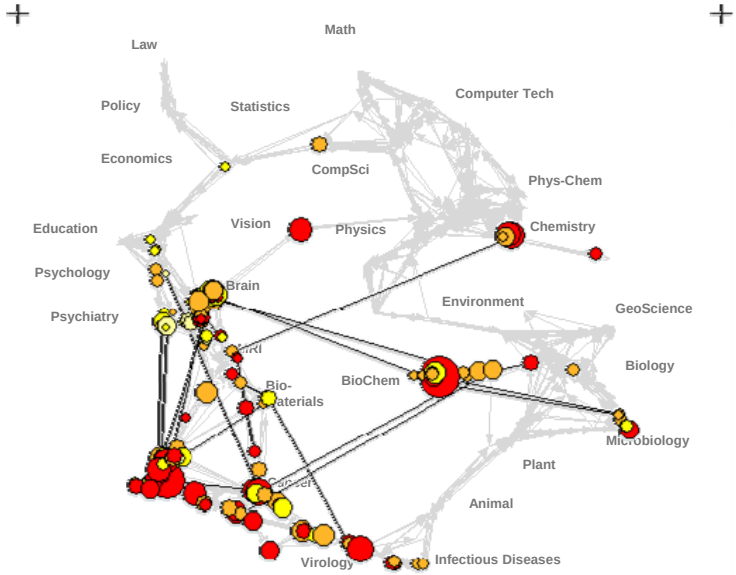
Funding Patterns of the National Science Foundation (NSF)



Science map applications: Identifying core competency

Kevin W. Boyack, Katy Börner, & Richard Klavans (2007).

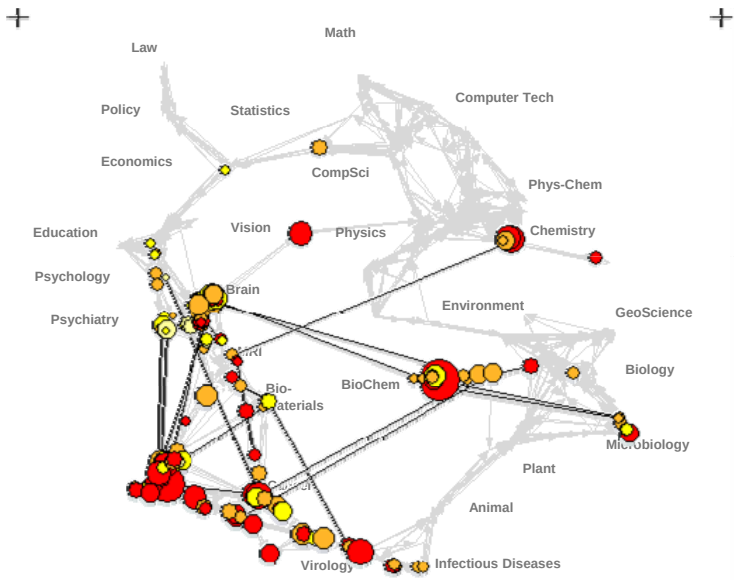
Funding Patterns of the National Institutes of Health (NIH)



Science map applications: Identifying core competency

Kevin W. Boyack, Katy Börner, & Richard Klavans (2007).

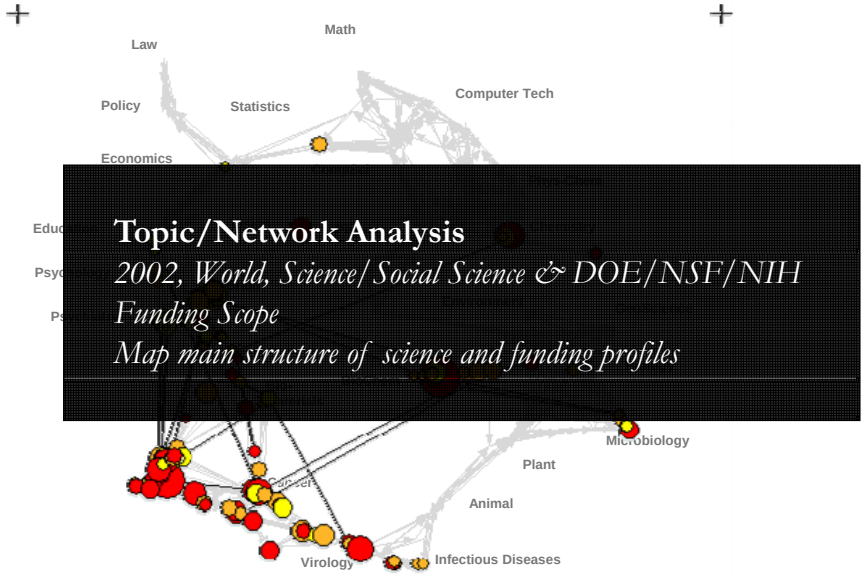
Funding Patterns of the National Institutes of Health (NIH)



Science map applications: Identifying core competency

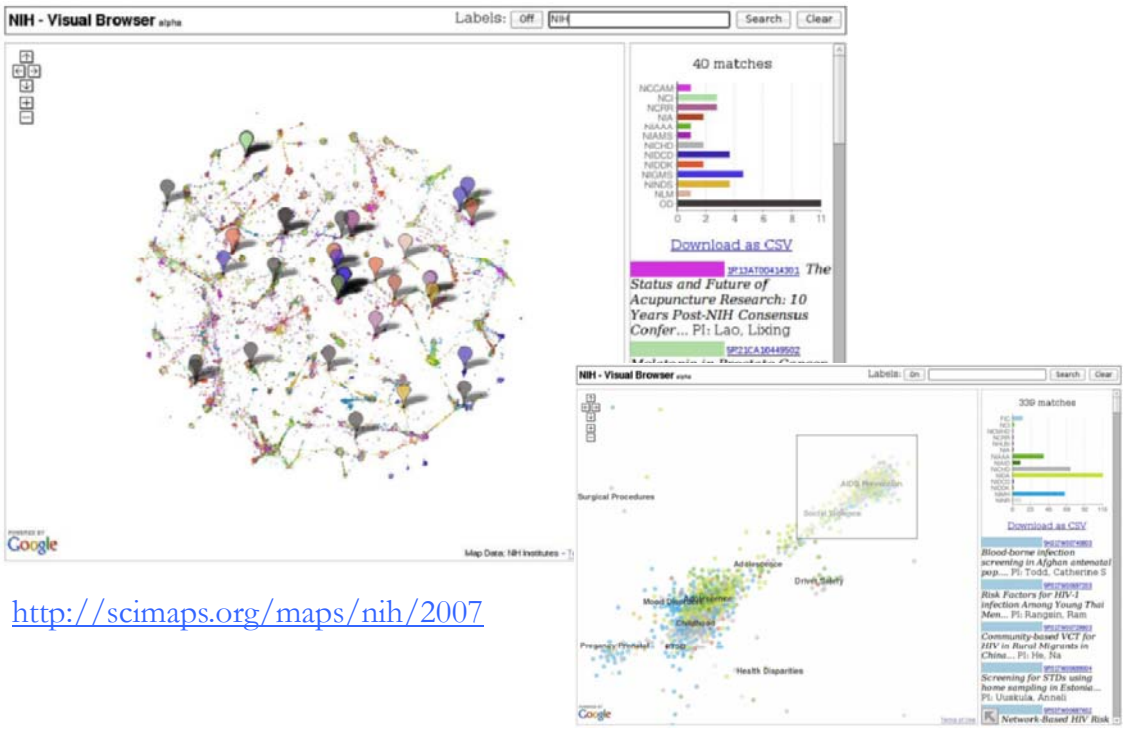
Kevin W. Boyack, Katy Börner, & Richard Klavans (2007).

Funding Patterns of the National Institutes of Health (NIH)



Interactive Science Map of NIH Funding

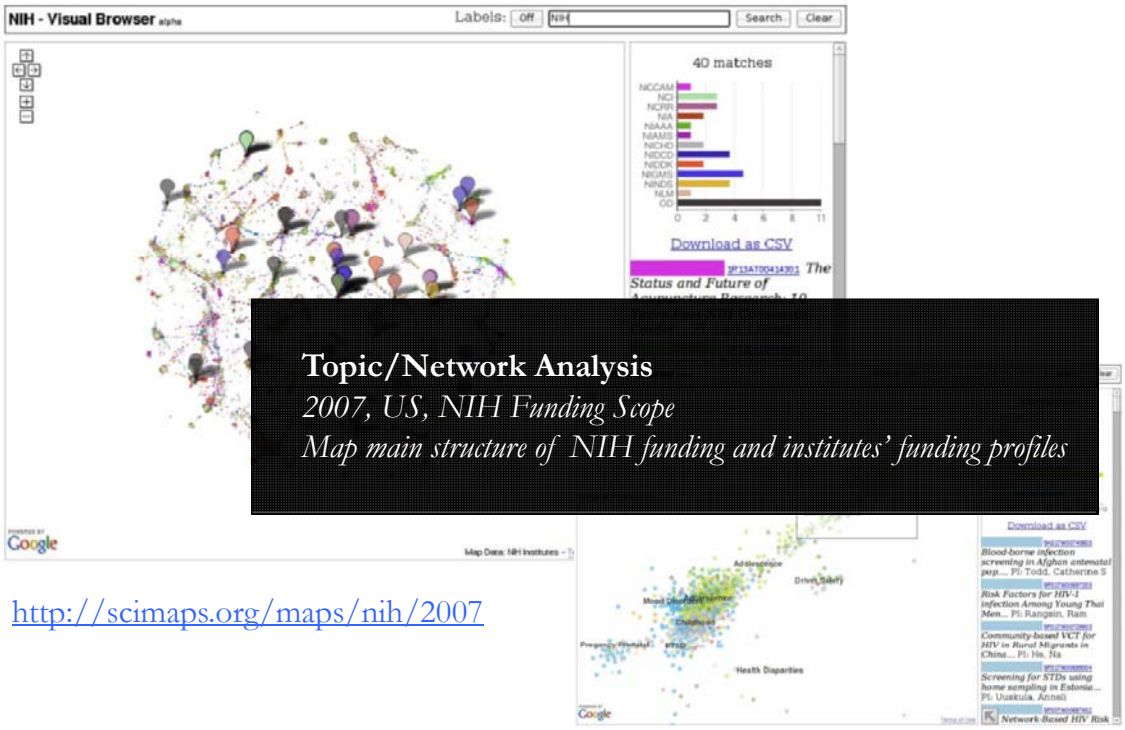
Herr II, Bruce W., Talley, Edmund M, Burns, Gully APC, Newman, David & La Rowe, Gavin. (2009).



http://scimaps.org/maps/nih/2007

Interactive Science Map of NIH Funding

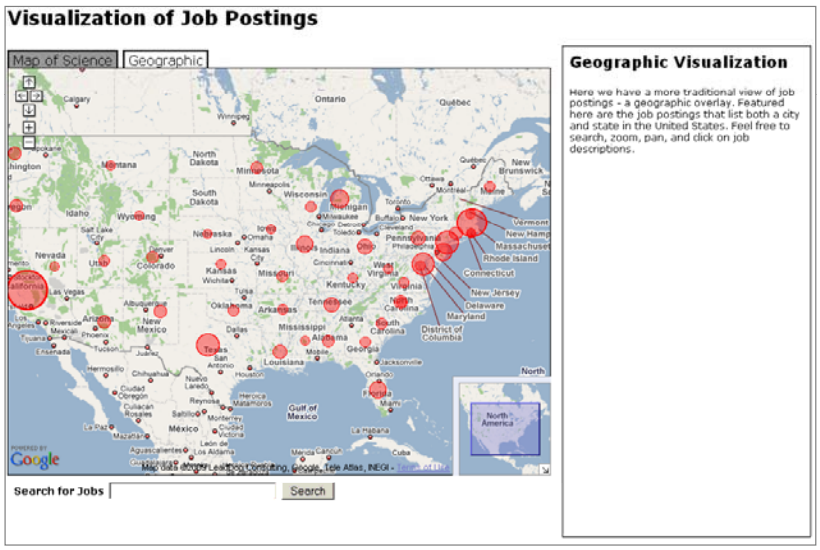
Herr II, Bruce W., Talley, Edmund M, Burns, Gully APC, Newman, David & La Rowe, Gavin. (2009).



http://scimaps.org/maps/nih/2007

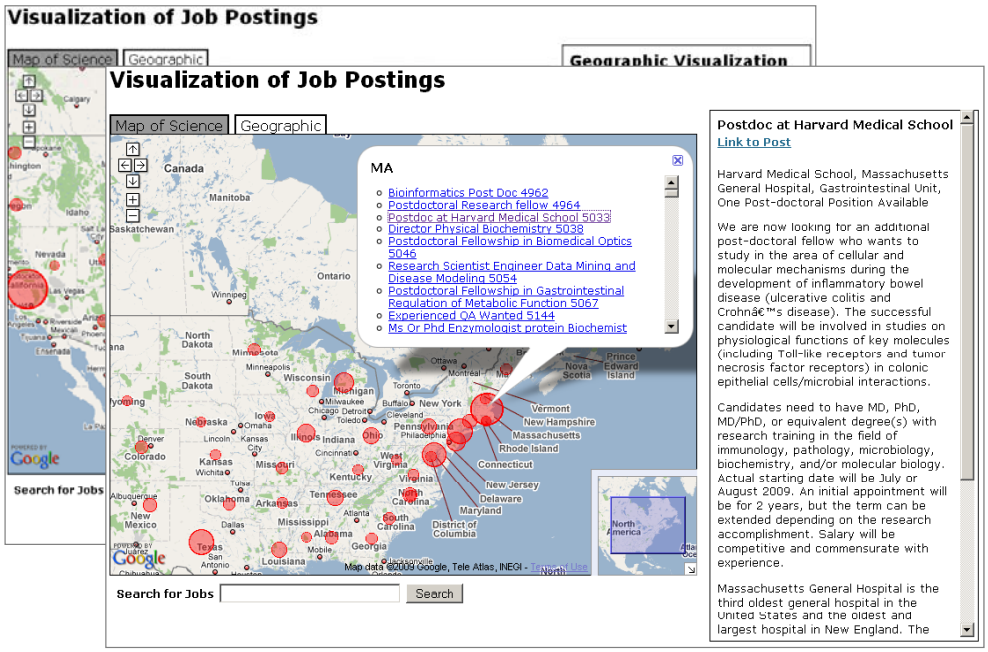
Interactive World and Science Map of S&T Jobs

Angela Zoss, Michael Connover, Katy Börner (2010).



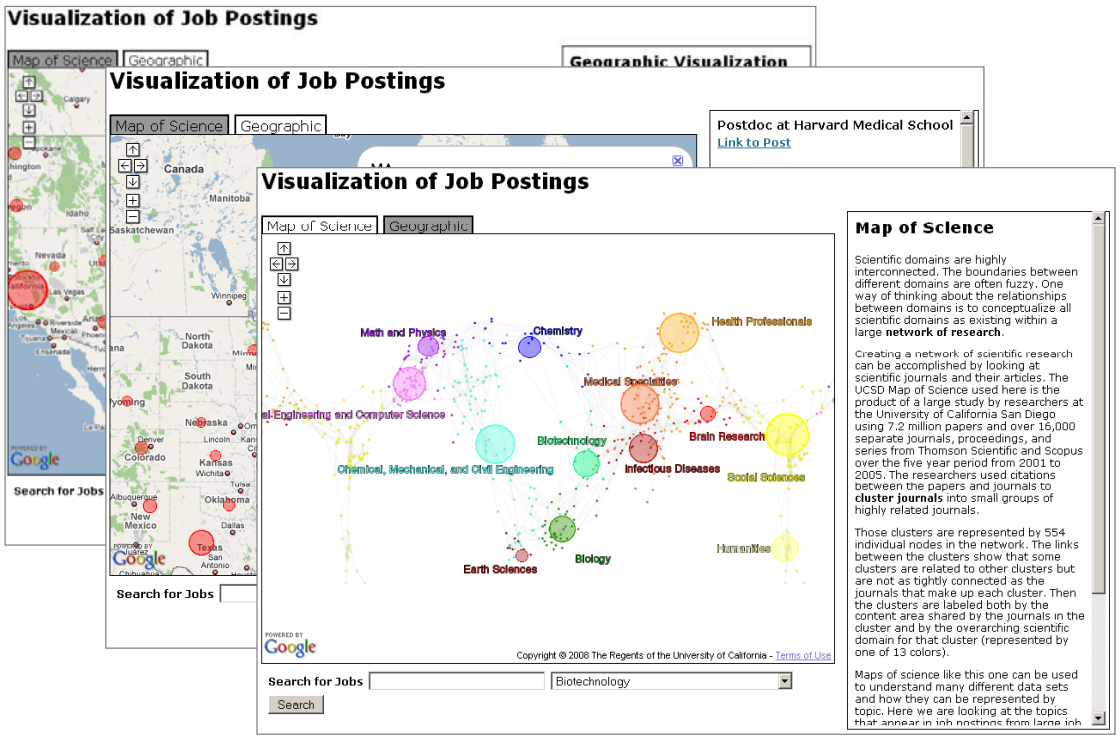
Interactive World and Science Map of S&T Jobs

Angela Zoss, Michael Connover, Katy Börner (2010).



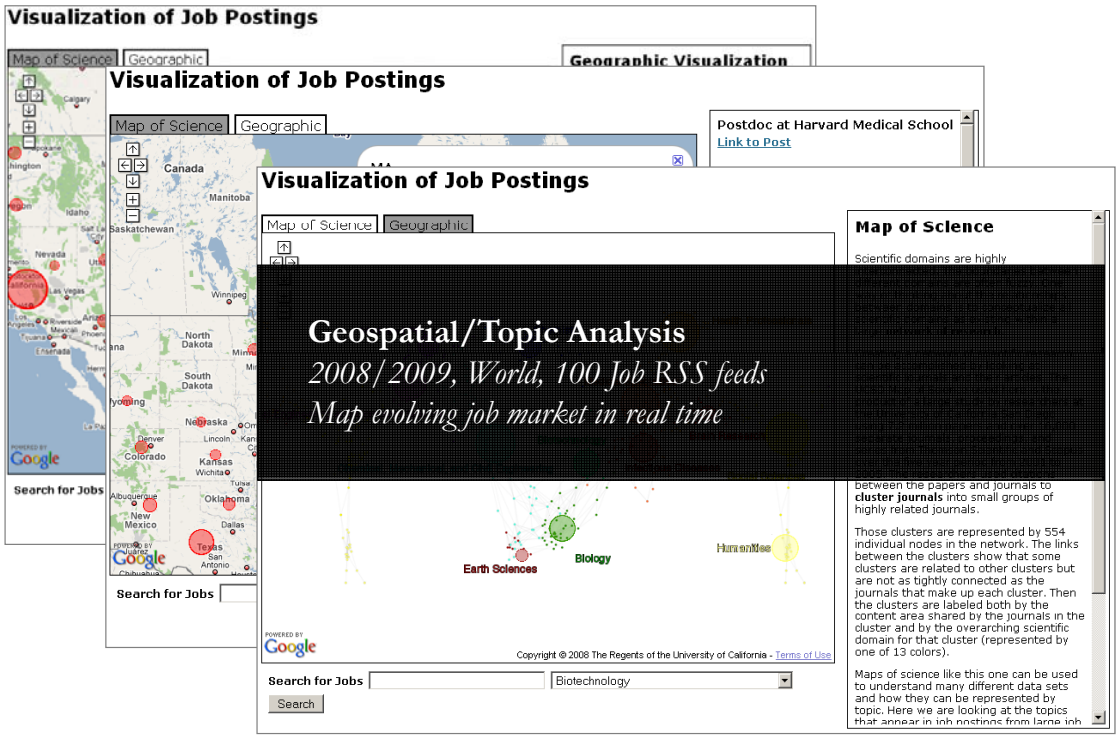
Interactive World and Science Map of S&T Jobs

Angela Zoss, Michael Connover, Katy Börner (2010).



Interactive World and Science Map of S&T Jobs

Angela Zoss, Michael Connover, Katy Börner (2010).



Social Network Visualizations

- Visualization Types and Levels
- Exemplary User Needs
- Proposed VIVO Visualizations

Exemplary User Needs

- **Individual level.** Researchers would like to enter data once and then use it to print cv's, annual summary reports, find team members & mentors, render web pages to "become effortlessly visible" in support of collaboration and research.
- **Institutional level.** Campus level officials need to pool (expertise) resources for major grant applications, understand research strengths and trends of different units as part of competitive landscape analysis, advertise their institution to recruit and retain students and faculty.
- **National level.** Funding agencies and others need to understand who is working on what topic(s), what research areas/expertise centers are emerging, or who is funding/supporting a certain topic/expert team.

Social Network Visualizations

- Visualization Types and Levels
- Exemplary User Needs
- Proposed VIVO Visualizations

Proposed VIVO Visualizations

- **Individual level.** Statistics and ego-centric scholarly networks on VIVO Profile pages.
- **Institutional level.** Analyses and visualizations of funding intake and publication output for departments and centers accessible via the VIVO Index page. Download of relevant data in tabular and network formats for further analysis using the Network Workbench tool.
- **National level.** Visualization of VIVO installations and their profile holdings together with web page access and general VIVO information requests.

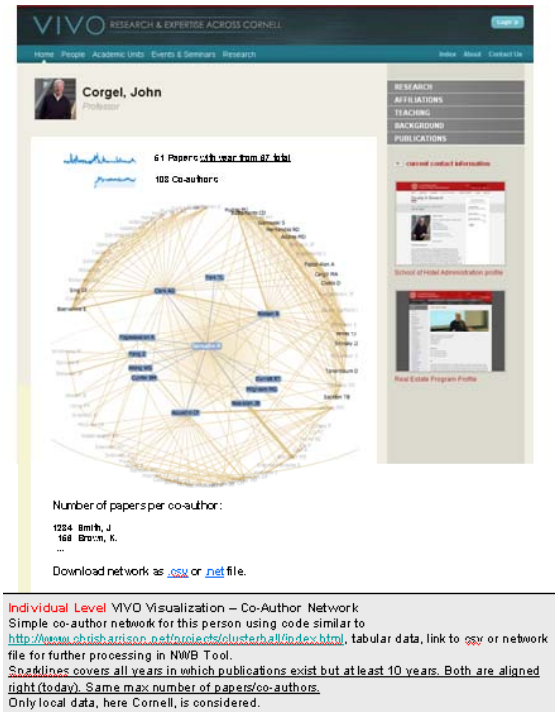
- **Graphic Design Visualizations** that show VIVO team, VIVO Conference Attendees, and VIVO activity.

Proposed VIVO Visualizations

- **Individual level.** Statistics and ego-centric scholarly networks on VIVO Profile pages.
- **Institutional level.** Analyses and visualizations of funding intake and publication output for departments and centers accessible via the VIVO Index page. Download of relevant data in tabular and network formats for further analysis using the Network Workbench tool.
- **National level.** Visualization of VIVO installations and their profile holdings together with web page access and general VIVO information requests.

- **Graphic Design Visualizations** that show VIVO team, VIVO Conference Attendees, and VIVO activity.

VIVO Individual Level Visualizations



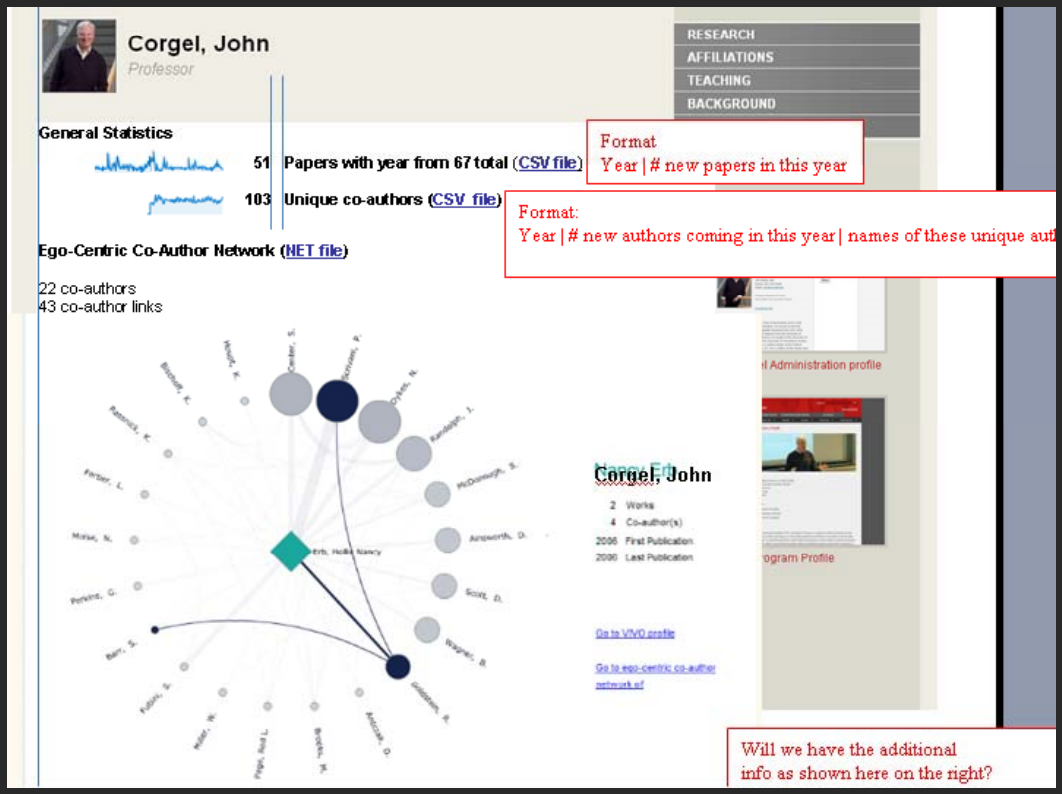
Ego-centric visualizations show the
- number of papers,
- number of authors,
- co-author network, and
- tables with counts per year.

Networks can be traversed, i.e., users can travel from people profile node to the next.

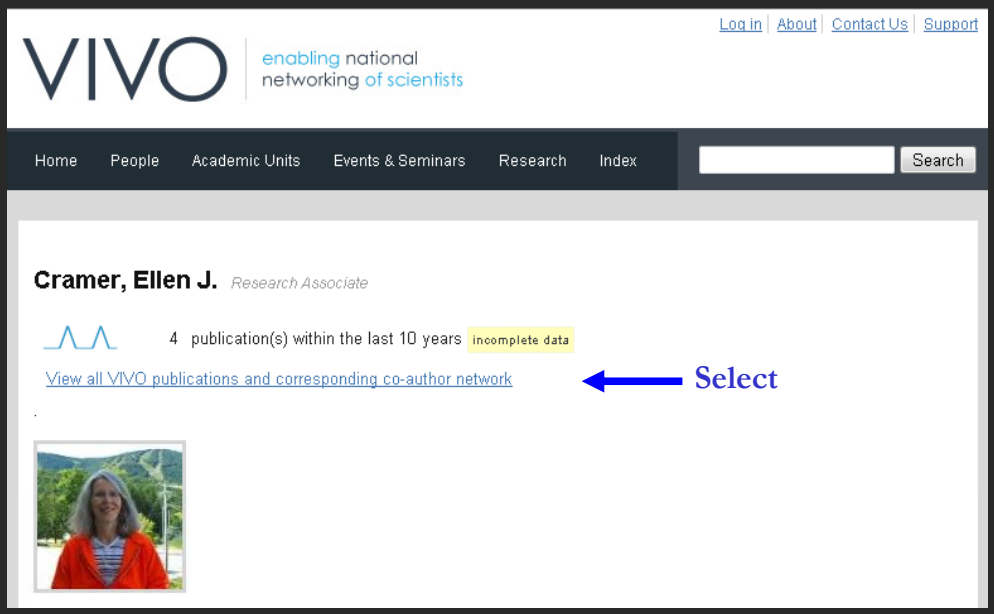
Network visualization can be saved. Data can be downloaded for further processing in MS Excel, Stats packages, network analysis tools.

Initial draft I shown on the left.

VIVO Later Draft of Individual Level Co-Author Visualization

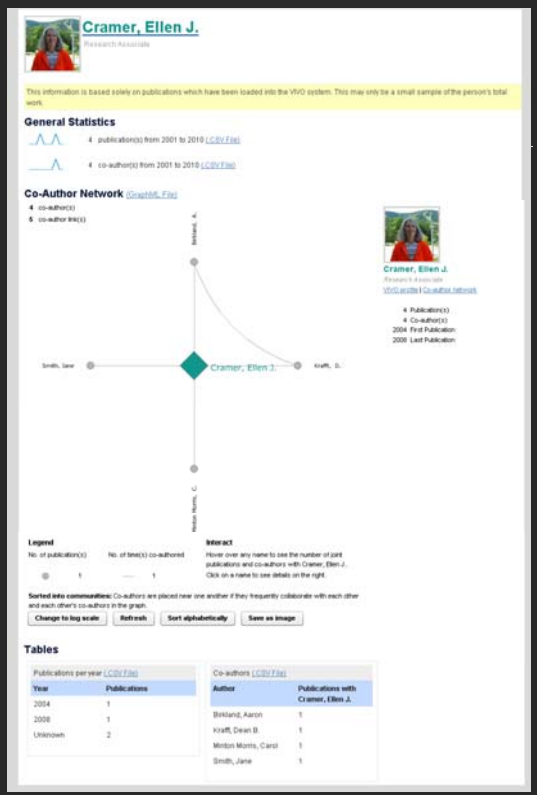


VIVO Release 1 v. 1.1: Individual Level Co-Author Visualization



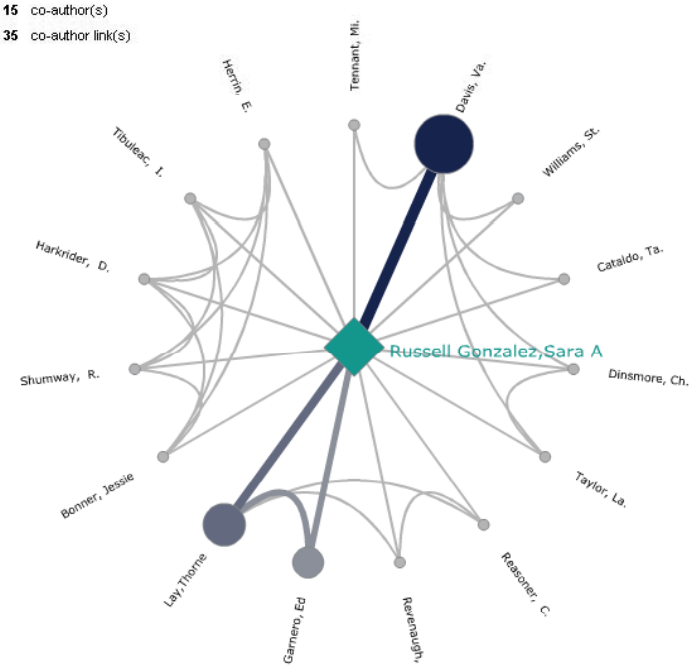
http://usertesting.mannlib.cornell.edu/display/CramerEllenJ

VIVO Release 1 v. 1.1: Individual Level Co-Author Visualization



Co-Author Network (GraphML File)

15 co-author(s)
35 co-author link(s)

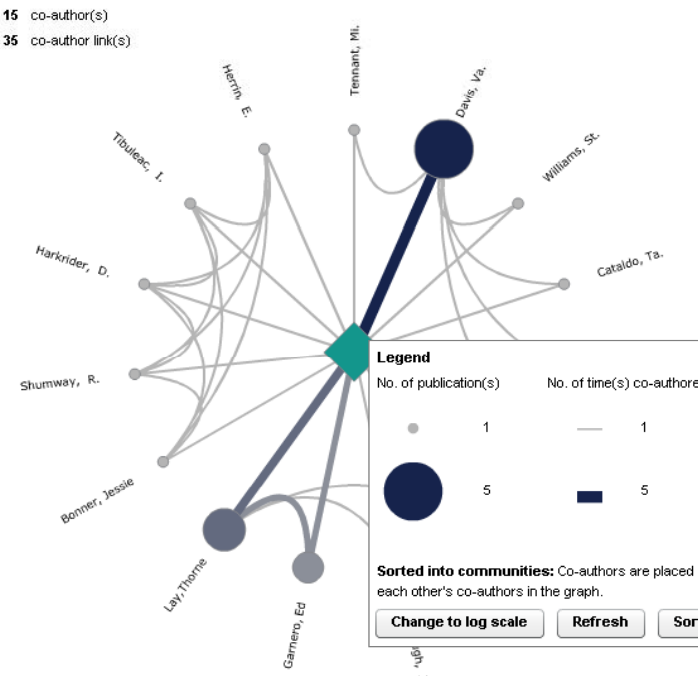


Russell Gonzalez,Sara A
Physical Sciences Librarian
VIVO profile | Co-author network

9 Publication(s)
15 Co-author(s)
1998 First Publication
2010 Last Publication

Co-Author Network (GraphML File)

15 co-author(s)
35 co-author link(s)



Russell Gonzalez,Sara A
Physical Sciences Librarian
VIVO profile | Co-author network

9 Publication(s)
15 Co-author(s)
1998 First Publication
2010 Last Publication

Legend

No. of publication(s)		No. of time(s) co-authored	
●	1	—	1
●	5	▬	5

Interact

Hover over any name to see the number of joint publications and co-authors with Russell Gonzalez,Sara A.

Click on a name to see details on the right.

Thresholding

Only people that co-authored more than 0 paper(s) with Russell Gonzalez,Sara A are shown.

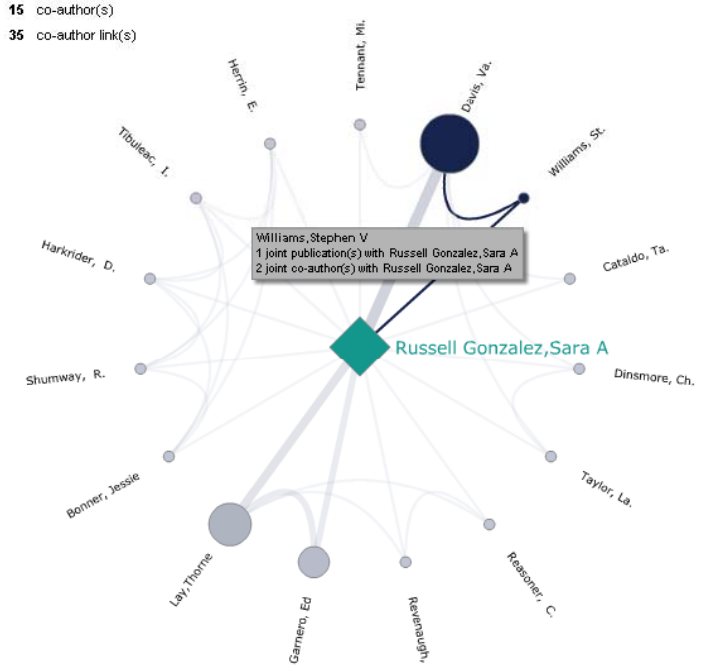
0 out of 0 co-author(s) are shown.

Sorted into communities: Co-authors are placed near one another if they frequently collaborate with each other and each other's co-authors in the graph.

Change to log scale | Refresh | Sort alphabetically | Save as image

Co-Author Network (GraphML File)

15 co-author(s)
35 co-author link(s)



Williams,Stephen V

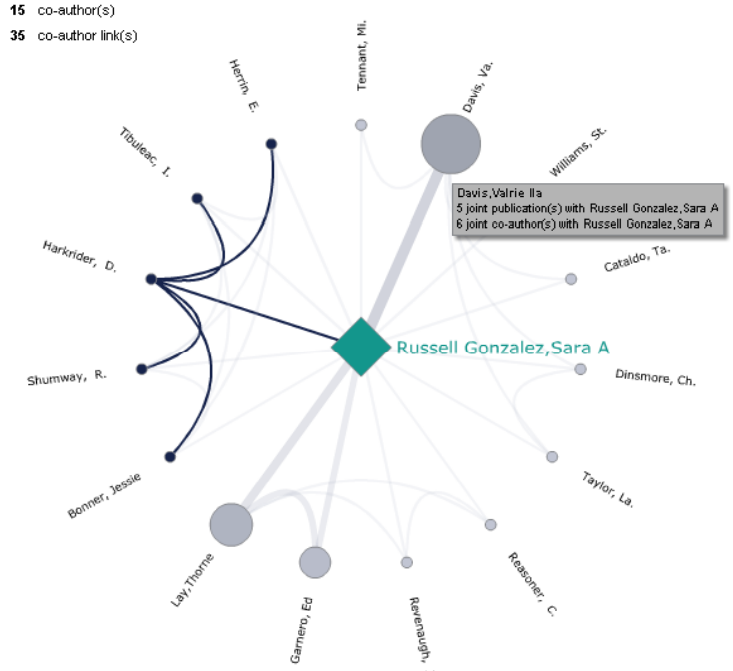
Software Engineer

VIVO profile | Co-author network

1 Joint Publication(s)
2 Joint Co-author(s)
2010 First Publication
2010 Last Publication

Co-Author Network (GraphML File)

15 co-author(s)
35 co-author link(s)

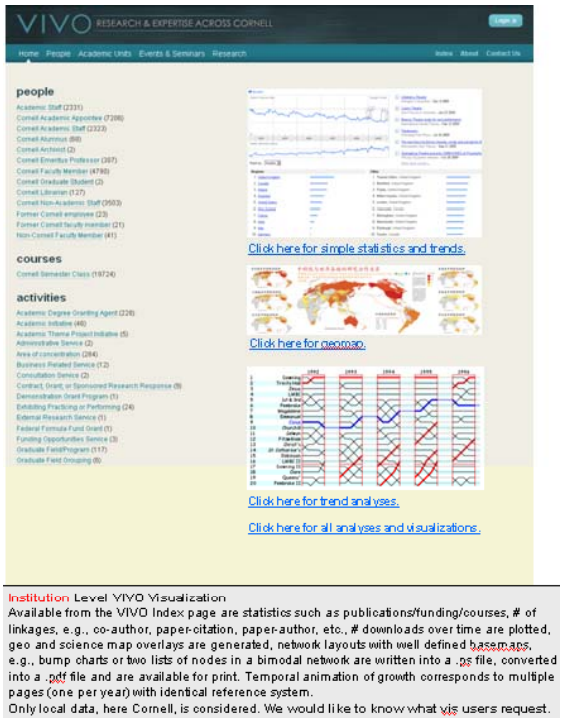


Harkrider, David G.

VIVO profile | Co-author network

1 Joint Publication(s)
5 Joint Co-author(s)
2003 First Publication
2003 Last Publication

VIVO Institution Level Visualizations



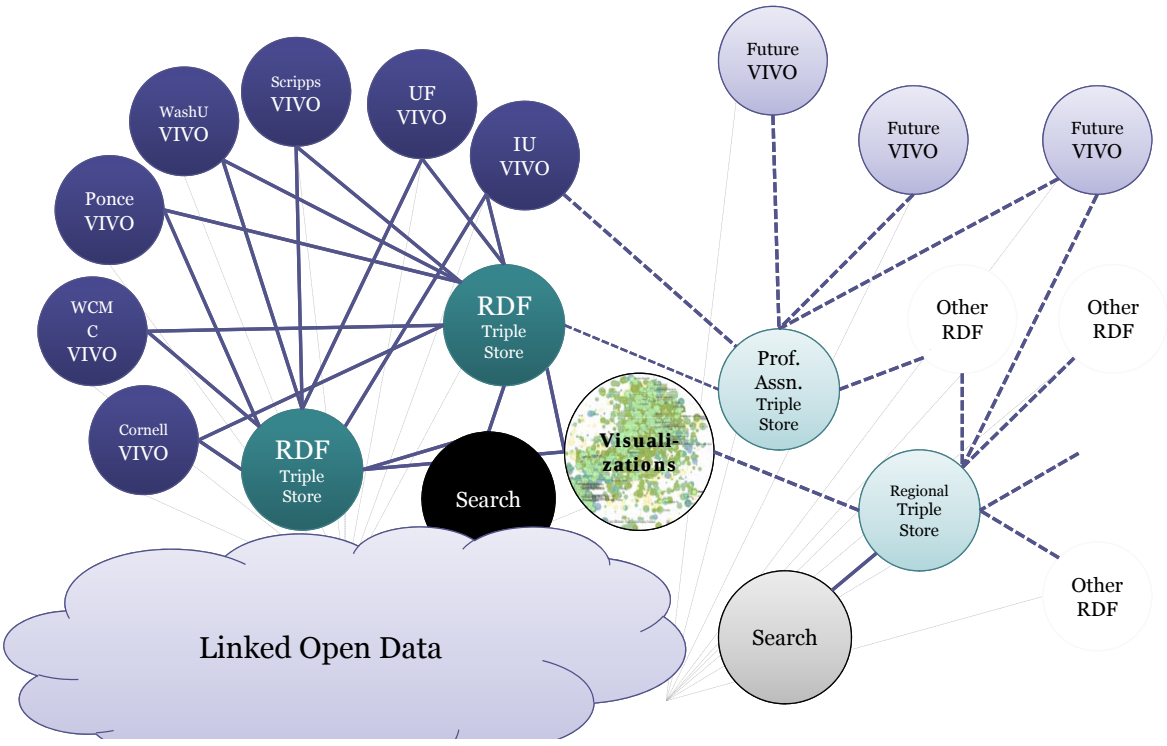
Institution level visualization will be available from the VIVO Index page and comprise statistics such as

- publications/funding/courses,
- # of linkages, e.g., co-author,
- paper-citation, paper-author, etc.,
- # downloads over time are plotted.

Geospatial and science map overlays as well as network layouts with well defined base maps, e.g., two lists of nodes in a bimodal network will be written into a PDF file for viewing and printing.

Temporal animation of growth corresponds to multiple pages (one per year) with identical reference system.

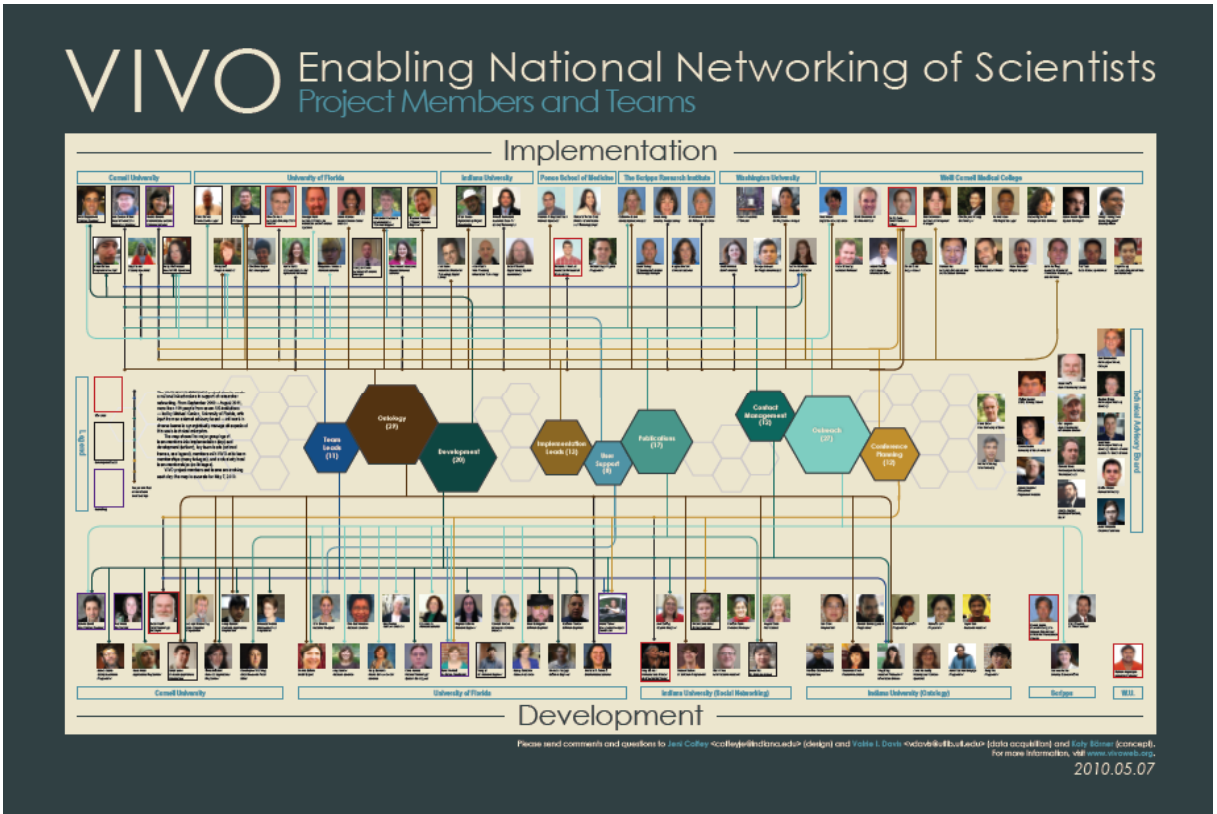
VIVO National Level Visualizations



Graphic Design Visualizations

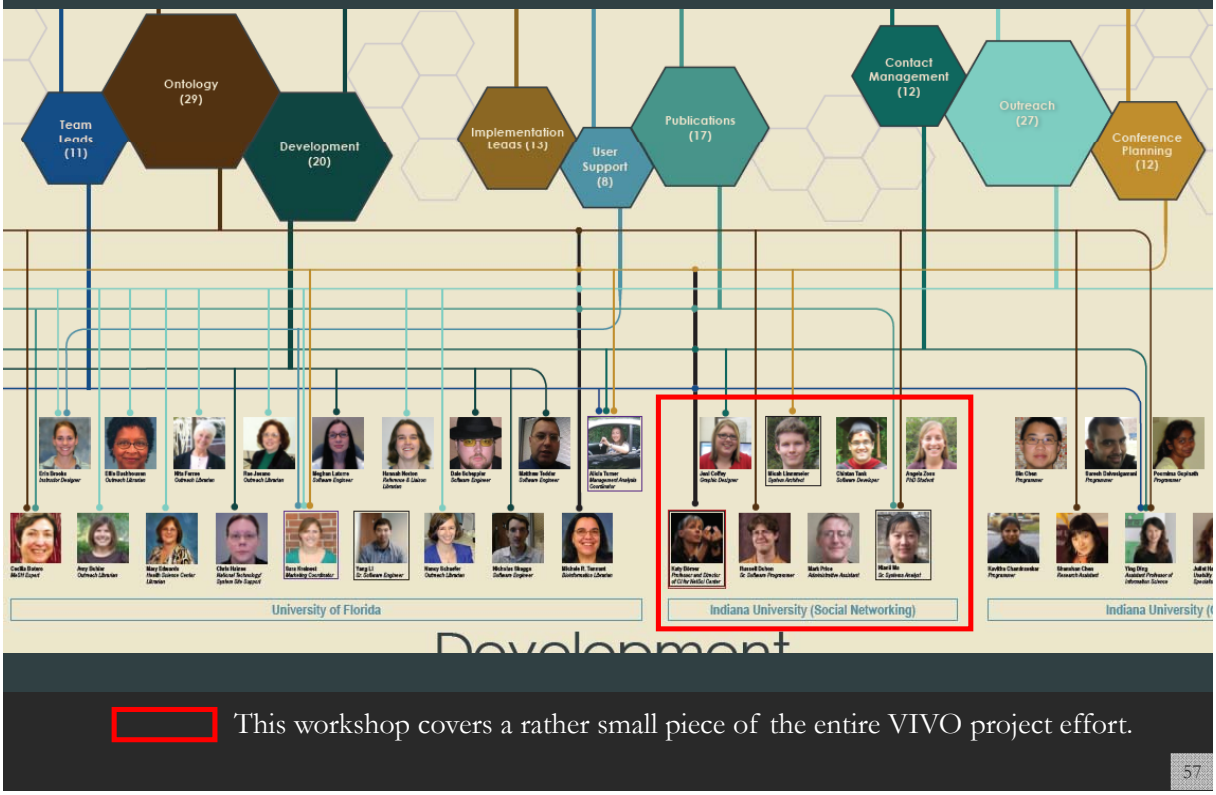
that show

- VIVO Team,
- VIVO Conference Attendees, and
- VIVO Activity.



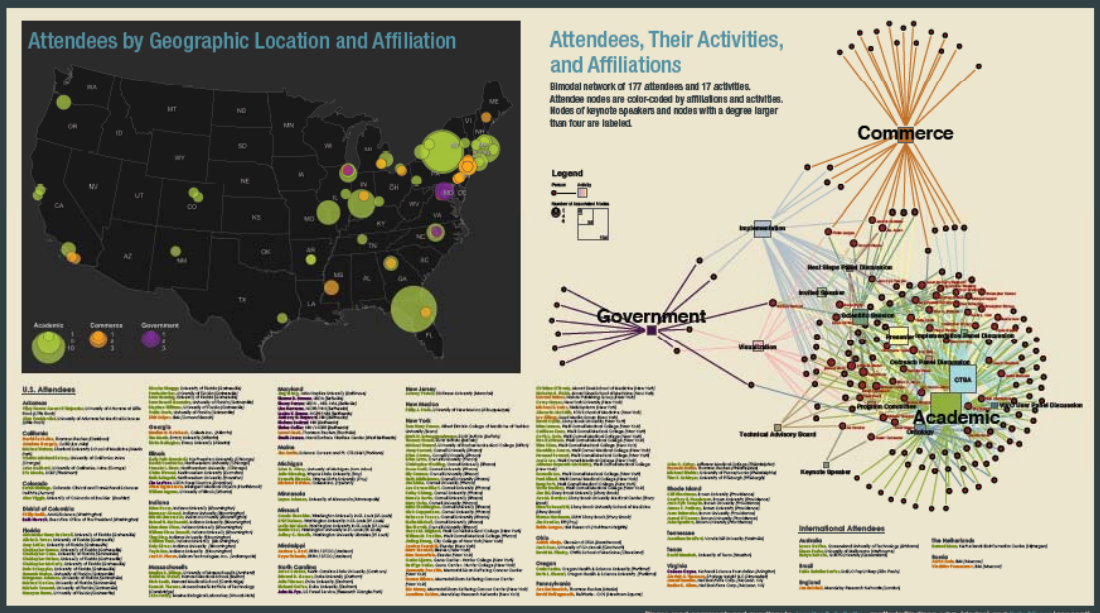
There are more than 120 people working on different aspects of VIVO.

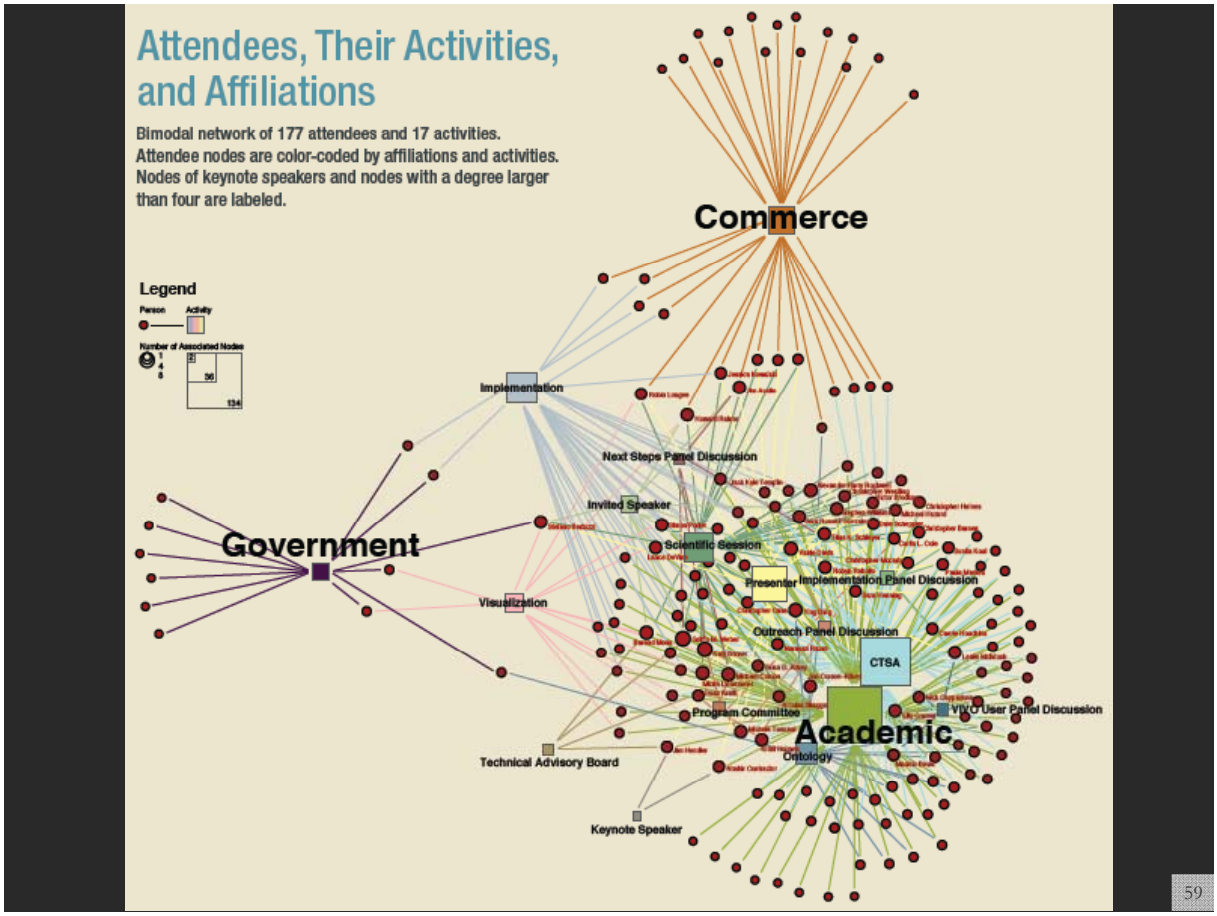
VIVO Enabling National Networking of Scientists
Project Members and Teams

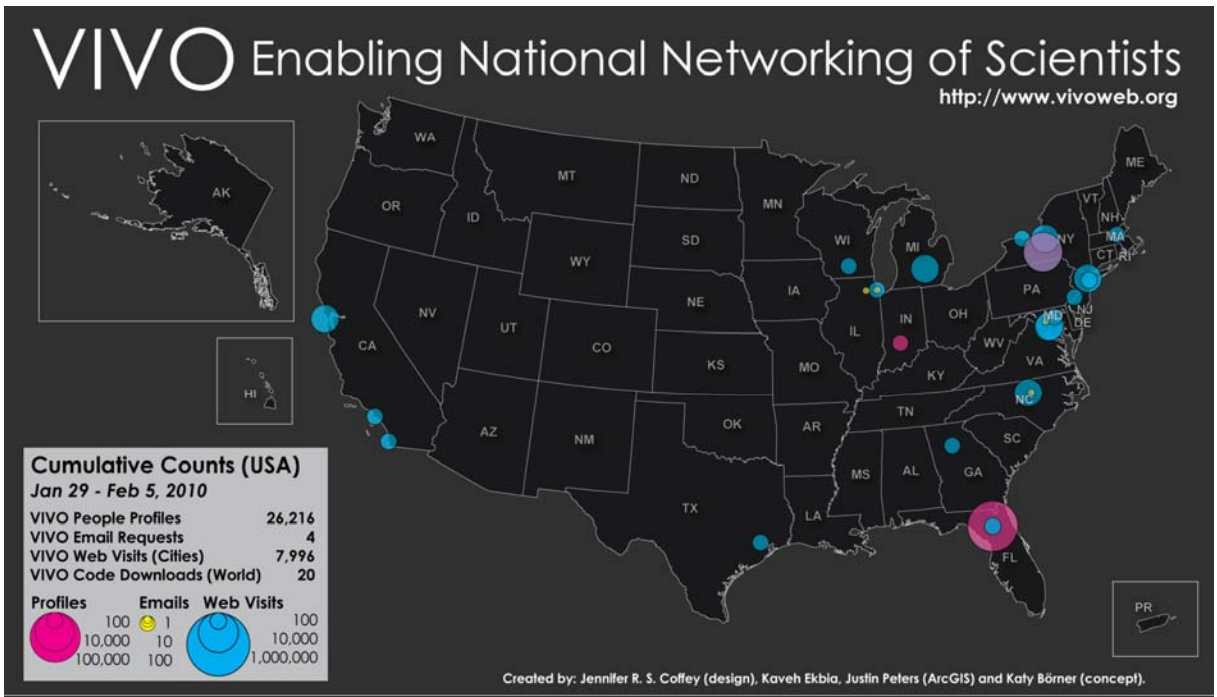


This workshop covers a rather small piece of the entire VIVO project effort.

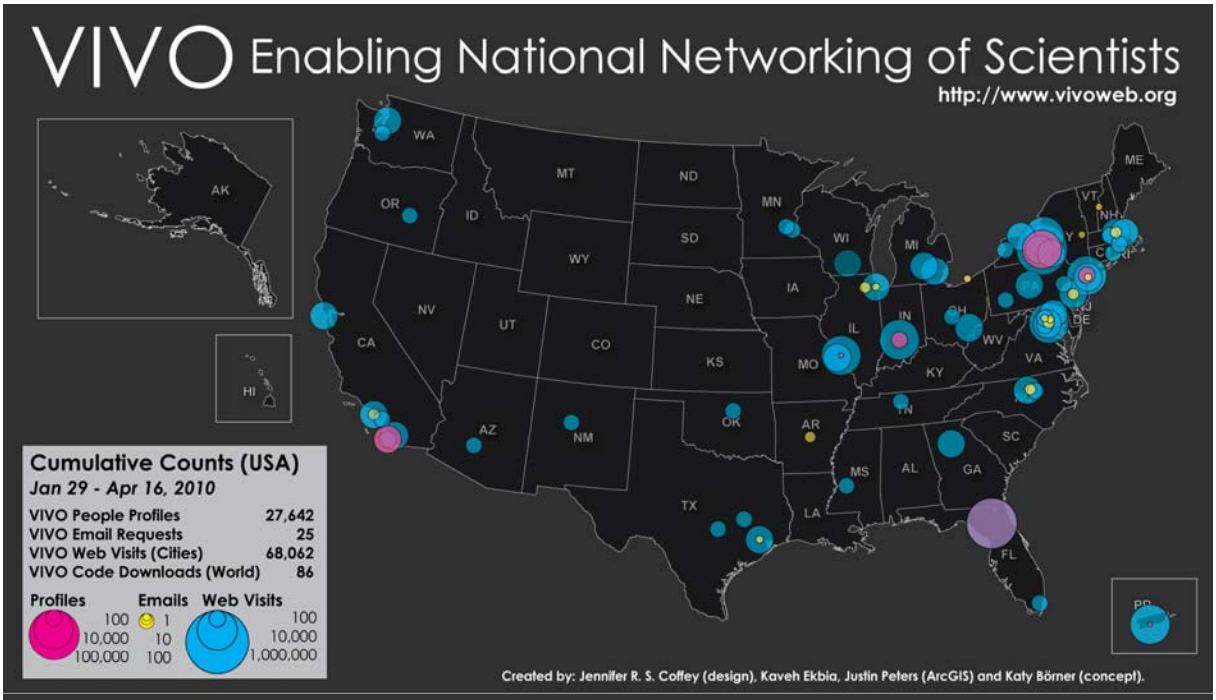
VIVO Enabling National Networking of Scientists
First Annual VIVO National Conference





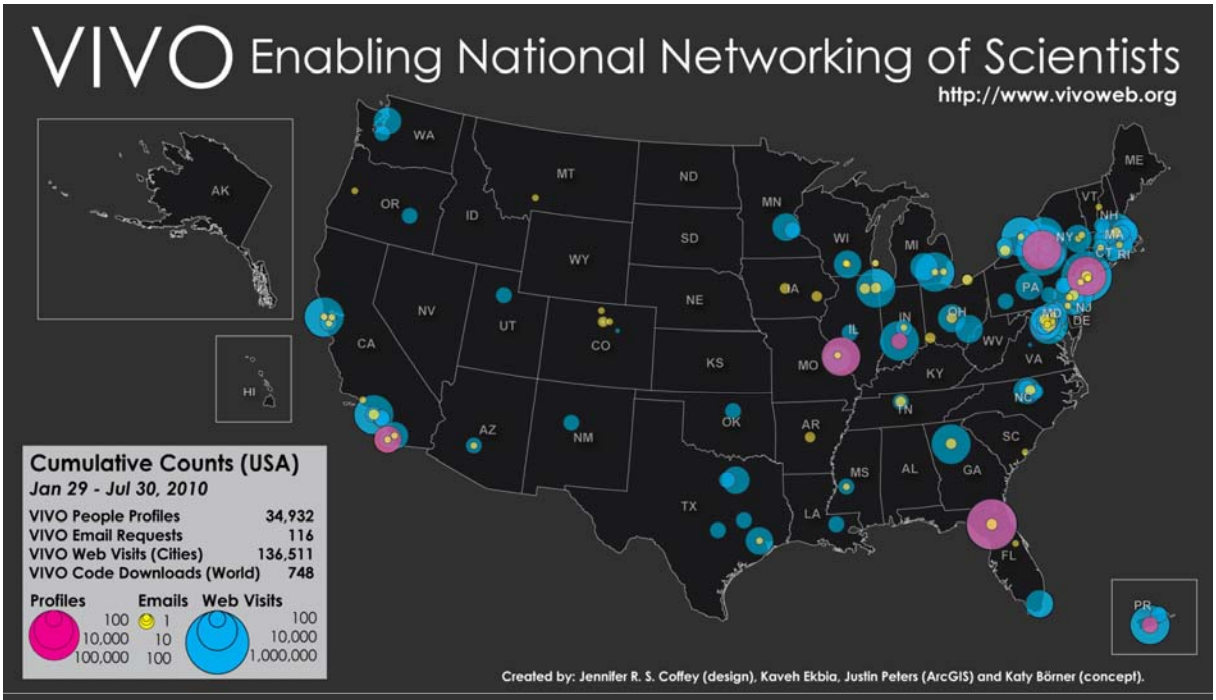


The National Research Network: VIVO: Enabling National Networking of Scientists NIH U24RR029822
Start: Sept 2009
PI: Michael Conlon, University of Florida
Award amount: $12,300,000

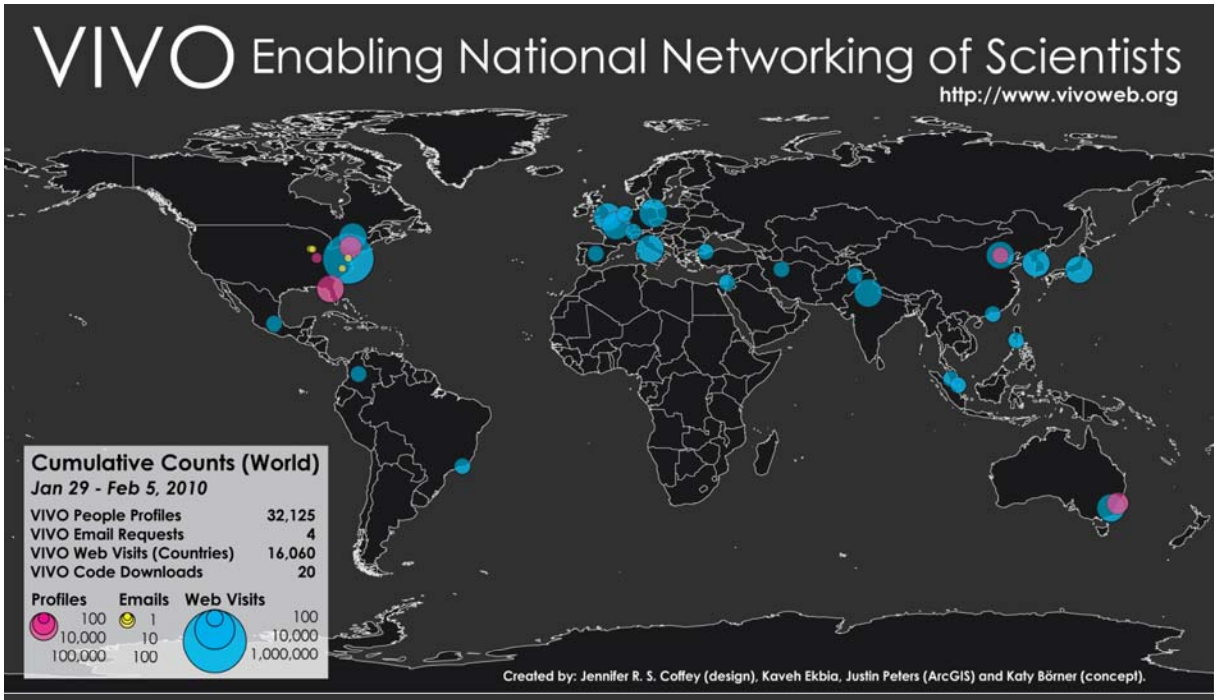


Shown are the
- Number of people profiles in the 7 different installation sites as well as CAS and U of Melbourne
- Email contacts by data and service providers as well as institutions interested to adopt VIVO.
- The number of visitors on http://vivoweb.org

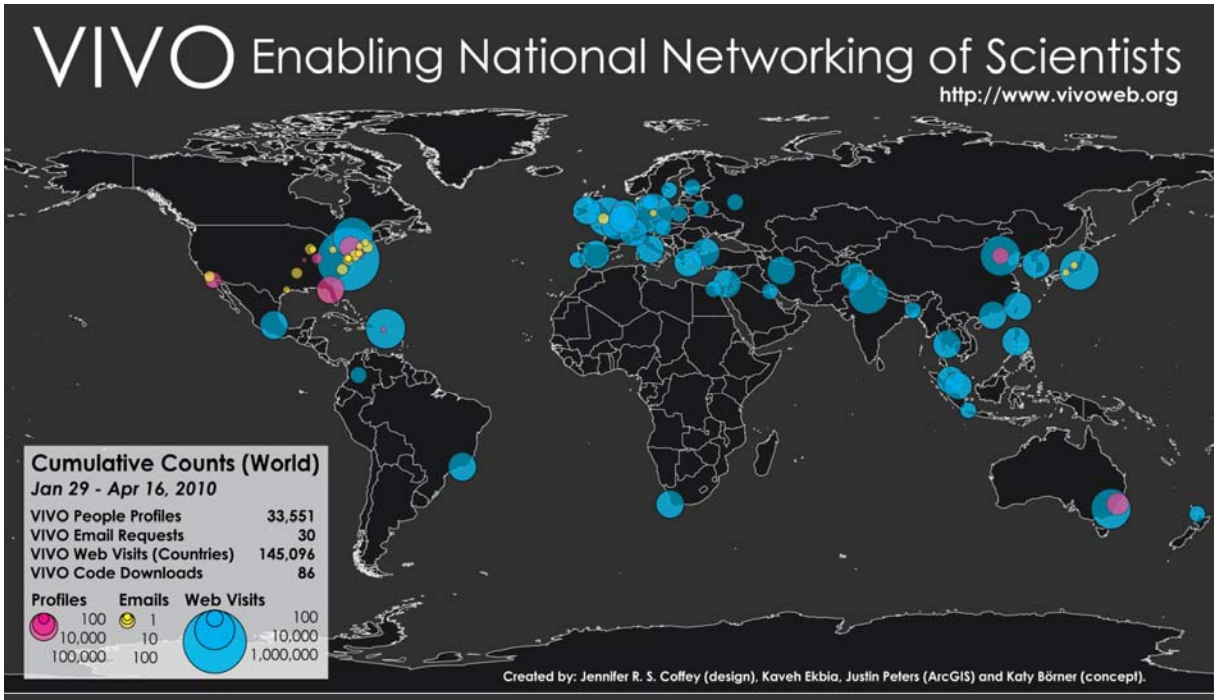
Circles are area size coded using a logarithmic scale.



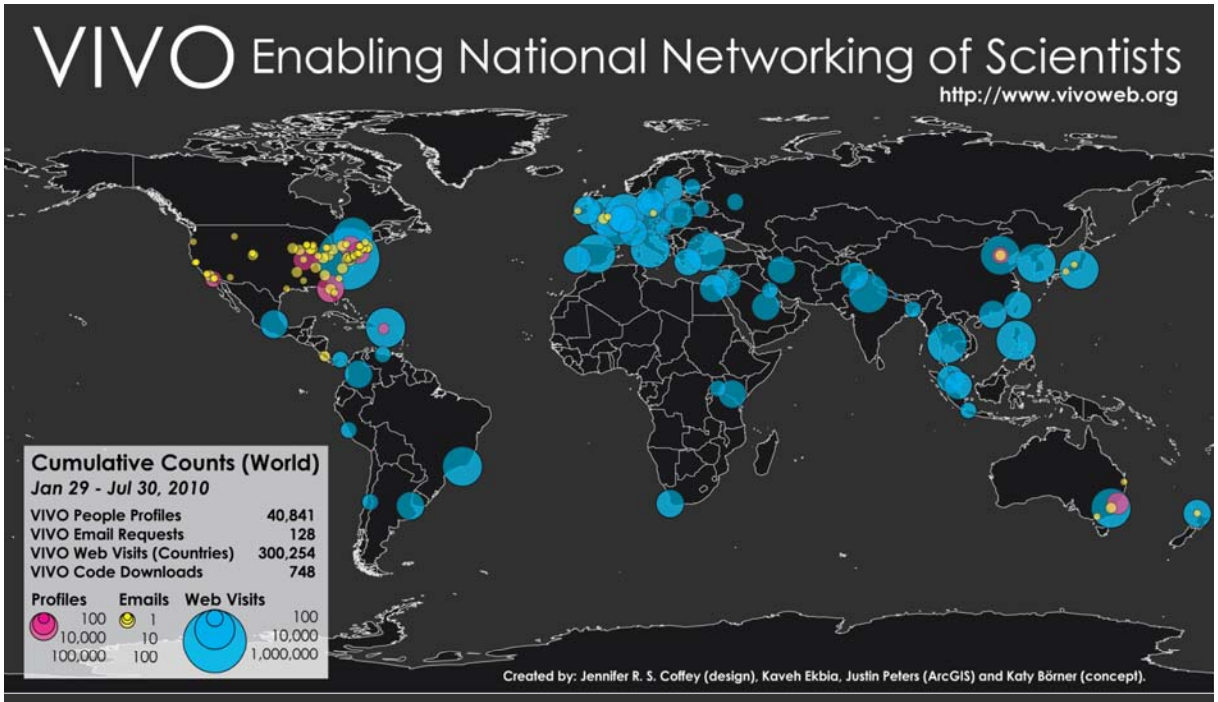
VIVO 1.0 source code was publicly released on April 14, 2010
87 downloads by June 11, 2010.
The more institutions adopt VIVO, the more high quality data will be available to understand, navigate, manage, utilize, and communicate progress in science and technology.



Science is global. World view of VIVO activity.
Web site visits are aggregated at the country level.



Shown are the
-Number of people profiles in the 7 different installation sites as well as CAS and U of Melbourne .
- Email contacts by data and service providers as well as institutions interested to adopt VIVO.
- The number of visitors on http://vivoweb.org
Circles are area size coded using a logarithmic scale.



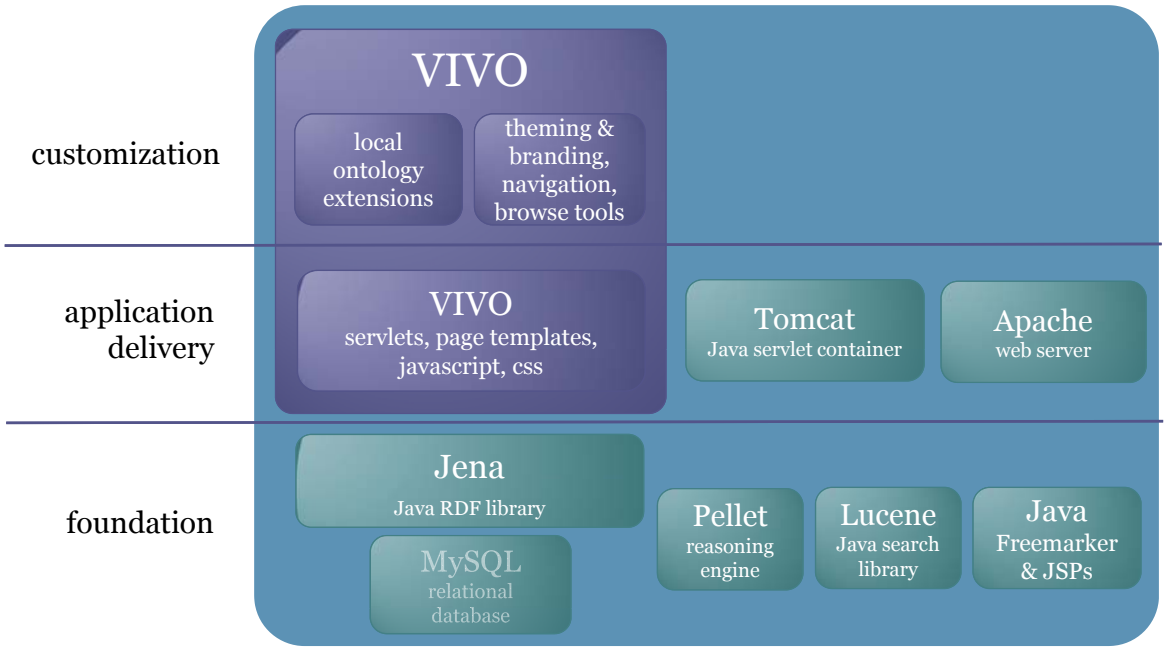
VIVO 1.0 source code was publicly released on April 14, 2010
87 downloads by June 11, 2010.
The more institutions adopt VIVO, the more high quality data will be available to understand, navigate, manage, utilize, and communicate progress in science and technology.

15 min break

Accessing and Using VIVO Data

- VIVO Architecture
- Accessing VIVO Data
- Developing for VIVO
- How we use VIVO data: The Visualization Pipeline

VIVO technical components



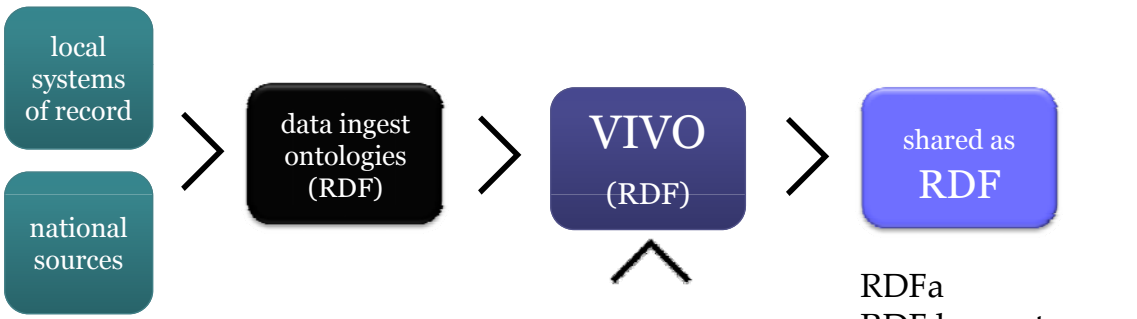
VIVO's Three Functional Layers

end users

curators

ontology editing
& data flow

Local Data Flow



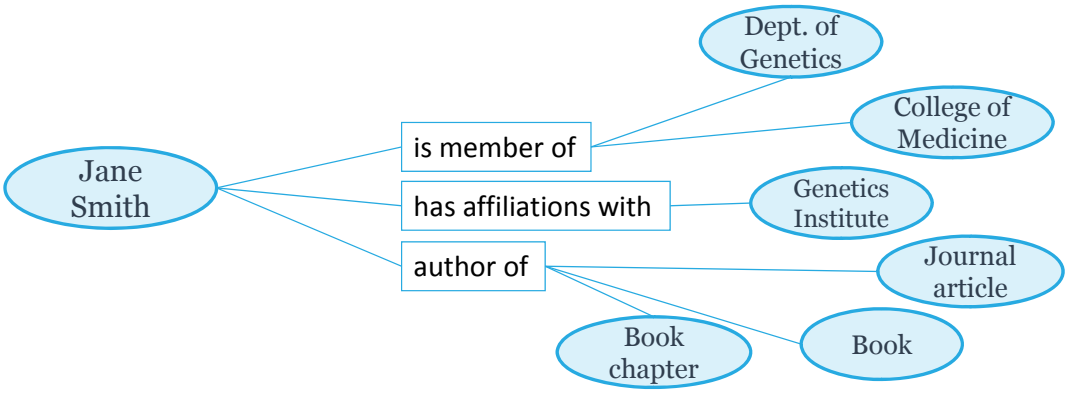
Peoplesoft
Grants DB
PubMed
Publishers

RDFa
RDF harvest
SPARQL endpoint

Researchers
Librarians
Administrative Staff
Self-Editors

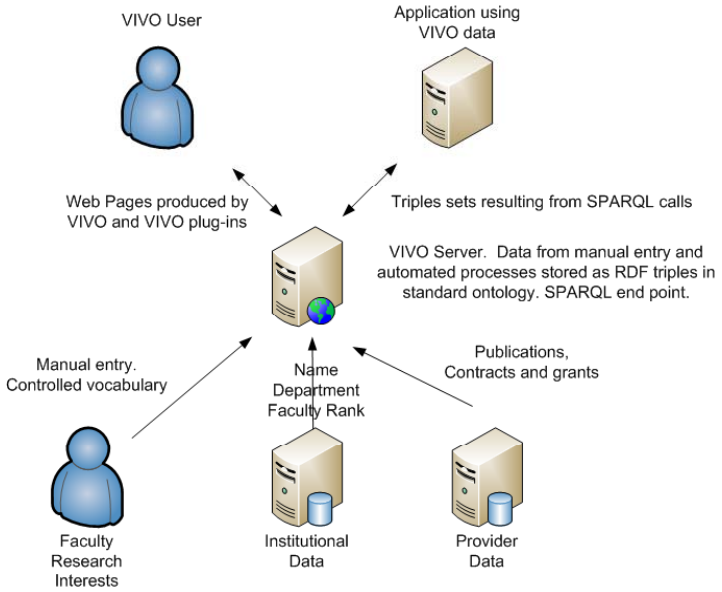
Storing Data in VIVO

- Information is stored using the **Resource Description Framework (RDF)**.
- Data is structured in the form of "triples" as subject-predicate-object.
- Concepts and their relationships use a **shared ontology** to facilitate the harvesting of data from multiple sources.



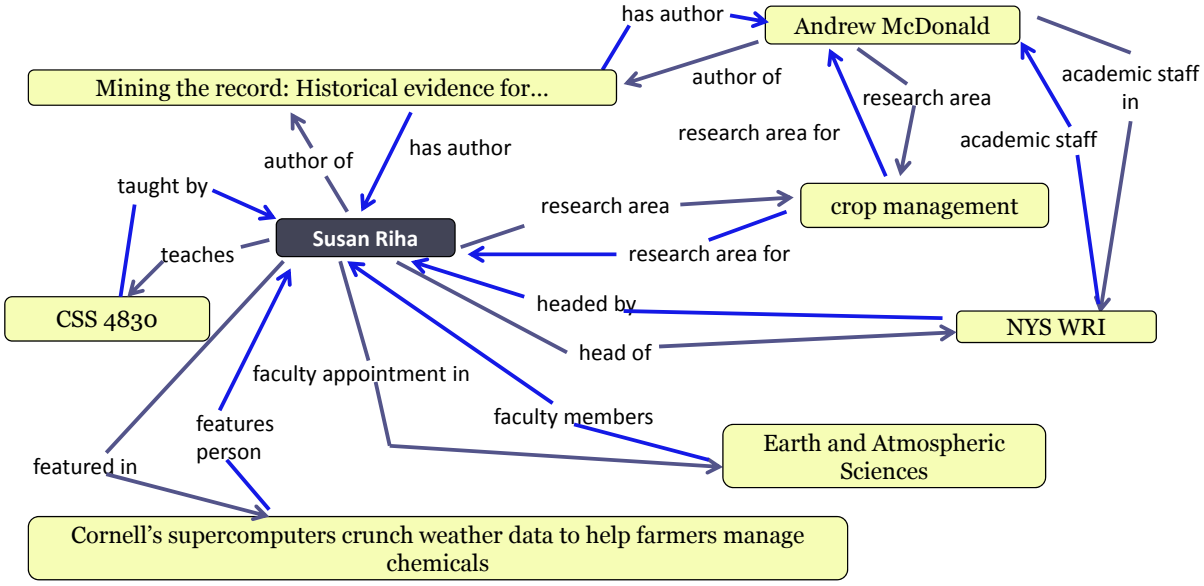
Institutional Architecture

- Three sources of VIVO information
  - User data
  - Institutional data
  - Provider data
- Two formats for output
  - Web Pages for users
  - Resource Description Framework for applications



Data Representation Using RDF Triples

Detailed relationships for a researcher at Cornell U.
Open source code (BSD) and ontology available at http://vivoweb.org.



Query and Explore

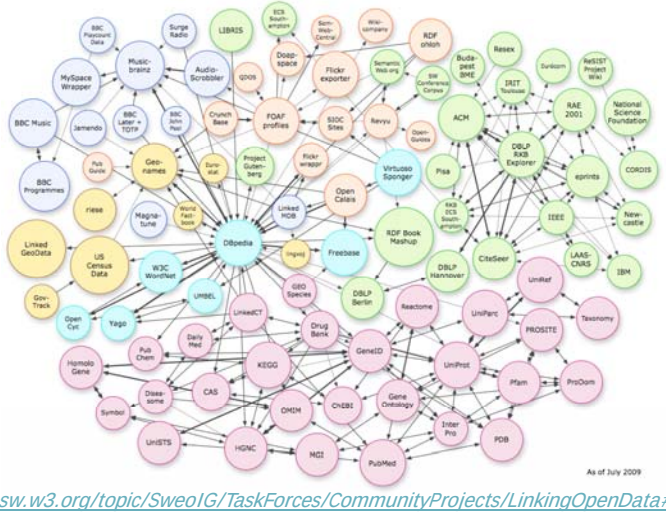
- By individual
  - Everything about an event, a grant, a person
- By type
  - Everything about a class of events, grants, or persons
- By relationship
  - Grants with PIs from different colleges or campuses
- By combinations and facets
  - Explore any publication, grant, or talk with a relationship to a concept or geographic location
  - Explore orthogonally (navigate a concept or geographic hierarchy)

Accessing and Using VIVO Data

- VIVO Architecture
- Accessing VIVO Data
- Developing for VIVO
- How we use VIVO data: The Visualization Pipeline

Methods of Accessing VIVO Data

- Linked Open Data (via RDF or N3)
- SPARQL Endpoints
- Read "The Semantic Web: An Introduction" http://infomesh.net/2001/swintro



author: Chris Bizer
taken from: http://esw.w3.org/topic/SweoIG/TaskForces/CommunityProjects/LinkingOpenData#dbpedia-lod-cloud

Open Linked Data (via RDF or N3)

- Advantage: Accessible by anyone on the Web.
- Disadvantage: Difficult to work with large amounts of data quickly/easily.

- N3 example:
  - http://vivo-vis-test.slis.indiana.edu/vivo/individual/Person72/Person72.n3
- RDF example:
  - http://vivo-vis-test.slis.indiana.edu/vivo/individual/Person72/Person72.rdf

SPARQL Endpoints

- Advantage: Working with data is easier/faster (using SPARQL queries).
- Disadvantage: May not be accessible to everyone.

Accessing and Using VIVO Data

- VIVO Architecture
- Accessing VIVO Data
- Developing for VIVO
- How we use VIVO data: The Visualization Pipeline

Developing a VIVO Application

- What you need
  - Someone who understands Semantic Web technologies.
  - Someone who can program (ideally these two are the same person!)
  - A computer with access to the web.
  - The rest is project dependent.

Development Scenarios

- Customize a VIVO instance.
- Create a new Data Ingest option (Data Harvester).
- Create an External VIVO Application
- Create an Internal VIVO Application
- Customized VIVO instance
- Internal VIVO Application (Visualization)

Customize a VIVO Instance

- Do this when you want to change:
  - Theming and Branding
  - Navigation
  - Browse tools
  - …
- This is supported: and relatively easy: Most (all?) current adopters do this.
- Resources: See the site administrator's guide on vivoweb.org: http://www.vivoweb.org/support/user-guide/administration

New Data Ingest Option (Data Harvester)

- Do this when you want to:
  - Ingest data from a new external data provider, or your own internal data sources.
  - If the data is in a relational database, it is already partly supported.
- Resources:
  - Check out the separate data ingest tool and source code at http://sourceforge.net/projects/vivo/files/
  - Get in touch with the harvester development team (University of Florida) at http://sourceforge.net/projects/vivo/support

Create an 'External' VIVO Application

- Do this when you want to:
  - Integrate VIVO data with your application.
  - Write an application that exists independently from a VIVO instance.
- Resources:
  - Get in contact with the VIVO team. http://vivoweb.org/contact should route you to the right contacts, depending on your project specifics.
  - Learn about the semantic web or examine our ontology at http://www.vivoweb.org/support

Create an 'Internal' VIVO Application

- Do this when you want to:
  - Write something more tightly integrated with VIVO.
- How to integrate depends on application needs.
- Resources:
  - Get in contact with the VIVO team. http://vivoweb.org/contact should route you to the right contacts, depending on your project specifics.
  - Download VIVO's source code and get in contact with VIVO developers at http://sourceforge.net/projects/vivo/

Create an 'Internal' VIVO Visualization

- Do this when you want to:
  - Write a visualization of VIVO data to be included inside a VIVO instance.
- Visualization architecture is modular: Easy to drop in code.
- Resources:
  - Contact mwlinnem@indiana.edu
  - Download VIVO's source code and get in contact with VIVO developers at http://sourceforge.net/projects/vivo/
- More on this shortly…

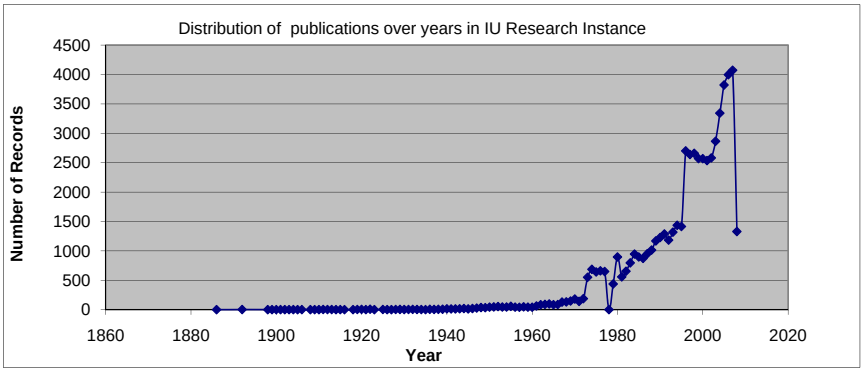
Needed for VIVO Vis Development

- Human Resources:
  - Web design/visualization skills (front-end)
  - Web development/Java skills (back-end)
  - Relational database expertise if you are involving non-semantic web data.
- On the server:
  - The VIVO server itself (includes Tomcat, MySQL, and more; see VIVO installation guide)
  - Plenty of memory (depending on how much data you intend to load)
- On the desktop:
  - Eclipse IDE for Java development
  - Flash/Flex/Flare and Google Visualization API for completed visualizations (more on this later)

IU Research Instance

http://vivo-vis.slis.indiana.edu

- Contains test data for visualization development.
  - Scopus publication data from May 2008 (grant data from NSF in the works).
  - Data was converted from CSV to RDF format.

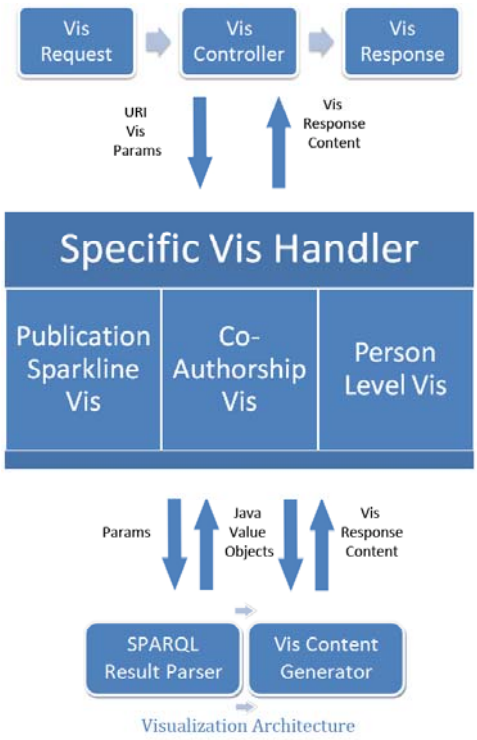


Accessing and Using VIVO Data

- VIVO Architecture
- Accessing VIVO Data
- Developing for VIVO
- How we use VIVO data: The Visualization Pipeline

VIVO Visualization Architecture

1. User requests the visualization
2. Request is received by the VIVO application
3. Vis Controller gets control
4. Vis Controller delegates the control of flow to the specific handler for the requested visualization
5. The handler passes request information to the SPARQL Result Parser
6. SPARQL Result Parser queries the semantic web data store
7. Results of the query are converted into Java objects
8. Java objects are used to generate response in the requested format
9. Request handler renders the generated response



Visualization Architecture

Co-Author Visualization Pipeline

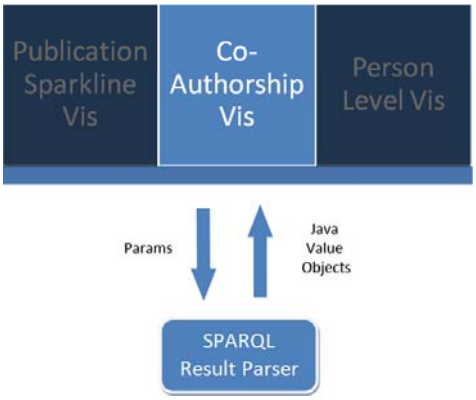
- Breakdown of serving the co-authorship visualization request received at, http://vivo-vis.slis.indiana.edu/vivo1/visualization?uri=http://vivoweb.org/ontology/core/Person72&vis=coauthorship&render_mode=standalone

- It has following parameters,
  - **/visualization** – URL prefix
  - **uri**=<Unique URI of a person e.g. http://vivoweb.org/ontology/core/Person72>
  - **vis**=coauthorship
  - **render_mode**=standalone
- Primer on some of the parameters
  - **vis** – Required. "Which visualization to be rendered?" E.g. Publication count visualization (person_pub_count), Co-authorship visualization (coauthorship), Person level visualization (person_level) etc.
  - **uri** – Required. "Unique URI of an individual" E.g. http://vivoweb.org/ontology/core/Person72
  - **render_mode** – Optional. "How to render the data?" E.g. data (for downloadable file), dynamic (for AJAX Calls) & standalone (for visualization to reside on it's own page)
  - **vis_mode** - Optional . "Modifier for the visualization being rendered" E.g. "Publication count visualization" can be rendered in 2 ways - short & full.

Co-Author Visualization Pipeline

- Since "/visualization" is provided in the URL VIVO application releases the control of flow to "Visualization Controller"
- Since the "coauthorship" is provided as value for "vis" parameter the controller delegates the control to Co-Authorship visualization request handler.

Co-Author Visualization Pipeline



- Captured parameters are passed to the SPARQL Result Parser
- Proceed only if URI is valid
- Plug parameters in the pre-built SPARQL query & run it

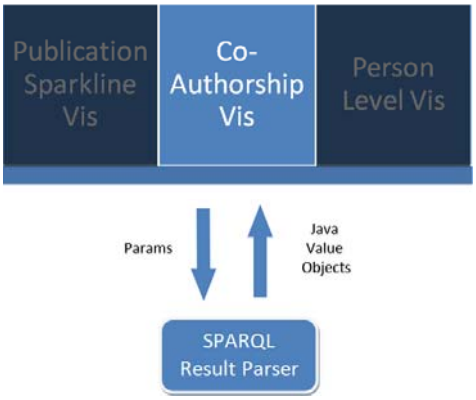
see next slide for pre-built SPARQL query

Co-Author Visualization Pipeline

- Pre-built SPARQL query

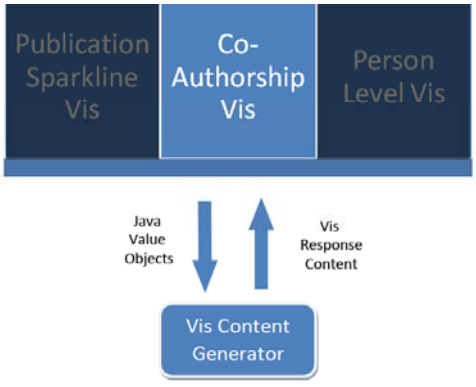
```
SELECT      (str(<URI>) as ?authPersonLit)
            (str(?authorLabel) as ?authorLabelLit)
            (str(?coAuthorPerson) as ?coAuthPersonLit)
            (str(?coAuthorPersonLabel) as ?coAuthPersonLabelLit)
            (str(?document) as ?documentLit)
            (str(?documentLabel) as ?documentLabelLit)
            (str(?documentMoniker) as ?documentMonikerLit)
            (str(?documentBlurb) as ?documentBlurbLit)
            (str(?publicationYear) as ?publicationYearLit)
            (str(?publicationYearMonth) as ?publicationYearMonthLit)
            (str(?publicationDate) as ?publicationDateLit)

WHERE {
            <URI>           rdf:type                        foaf:Person .
            <URI>           rdfs:label                      ?authorLabel  .
            <URI>           core:authorInAuthorship         ?authorshipNode .
            ?authshipNode rdf:type                          core:Authorship .
            ?authshipNode core:linkedInformationResource ?document .

            ?document       rdf:type                        bibo:Document .
            ?document       rdfs:label                      ?documentLabel .

            ?document       core:informationResourceInAuthorship        ?coAuthorshipNode .
            ?coAuthorshipNode           core:linkedAuthor               ?coAuthorPerson .
            ?coAuthorPerson             rdfs:label                      ?coAuthorPersonLabel .

            OPTIONAL { ?document core:year ?publicationYear } .
            OPTIONAL { ?document vitro:moniker ?documentMoniker } .
            OPTIONAL { ?document vitro:blurb ?documentBlurb } .
            OPTIONAL { ?document vitro:description ?documentDescription } .
            OPTIONAL { ?document core:yearMonth ?publicationYearMonth } .
            OPTIONAL { ?document core:date ?publicationDate }

}
```

Co-Author Visualization Pipeline



- Captured parameters are passed to the SPARQL Result Parser
- Proceed only if URI is valid
- Plug parameters in the pre-built SPARQL query & run it
- Using the query results create the java objects

Co-Author Visualization Pipeline



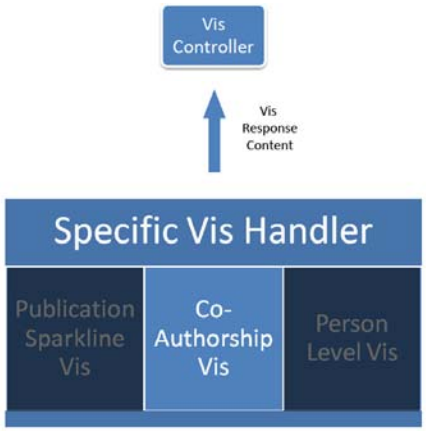
If **render_mode** is "standalone" (or illegal render_mode provided) then,

- Handler saves all the structured data to be used by the front-end
- The application response object has hooks for plugging in front-end content with the structured data

If **render_mode** is "data" (for a CSV or GraphML file)

- **vis_mode** parameter is checked next to identify which data to be served.
- If "sparkline" is provided then serve the CSV file data
  - Traverse through the structured data and generate CSV content
- Else by default GraphML file data is served
  - Traverse through the structured data and generate GraphML content

Co-Author Visualization Pipeline

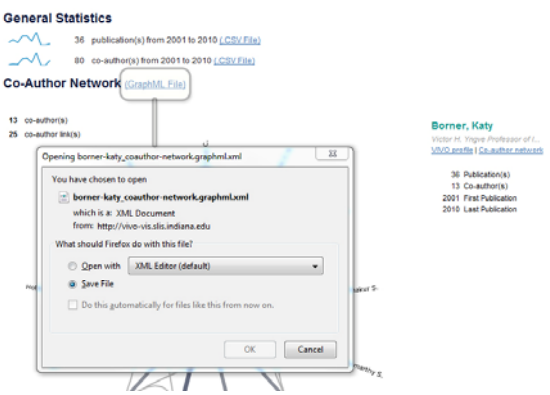


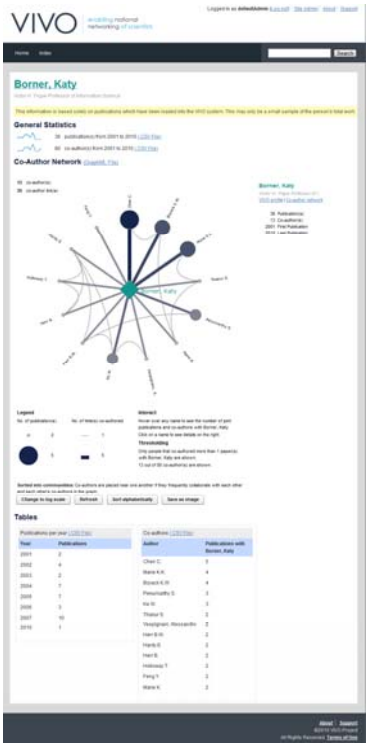
- Application response object is updated by Vis Response Content
- Visualization controller resumes control of flow

Co-Author Visualization Pipeline

Vis Response

- Visualization Controller renders the visualization content
- User sees the output as a web page or file download prompt





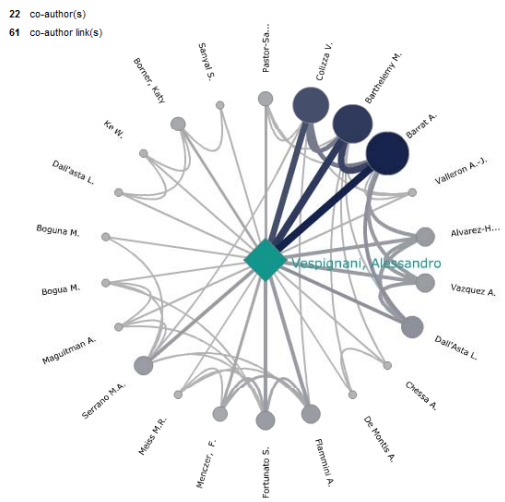
Visualization Details

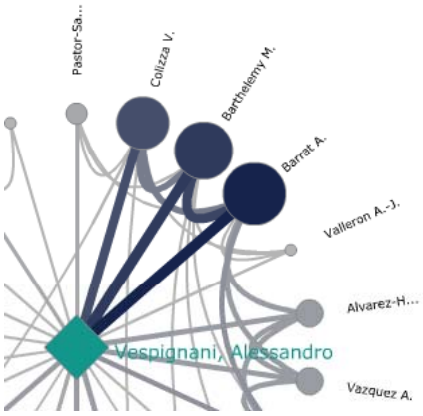
- Co-author visualization
- Sparklines
- Data download and analysis

VIVO
Enabling a National Network of Scientists

Concept

- Based on "personal networks" or "ego-centric networks" from social sciences.
  - Visualization from the perspective of the "ego".
- Frequent co-authors stand out.
- Sub-communities of collaborators are visible.





Code Libraries Used

- Flash
  - Powerful features
  - Consistent rendering across all browsers.
  - 97% of users have Flash installed[1].
  - But it doesn't work on the iPhone/iPad.
- Flare - http://flare.prefuse.org
  - A data visualization library for Flash.
- Adobe Flex - http://www.adobe.com/products/flex
  - A library for creating rich internet applications with Flash.
  - Used for interface components like buttons and text areas.

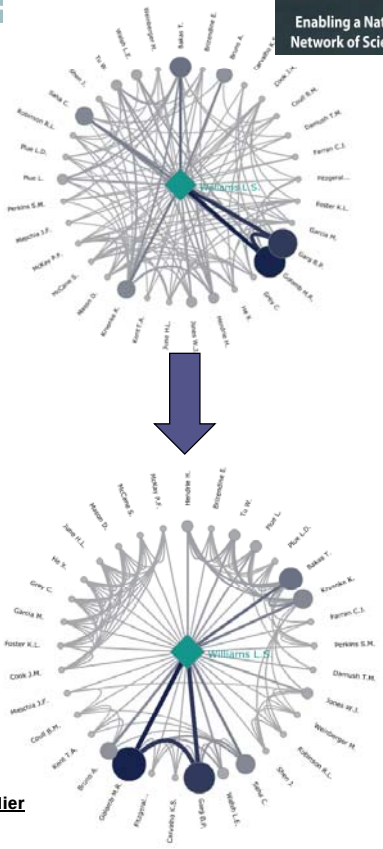
[1] - http://www.adobe.com/products/player_census/flashplayer/version_penetration.html

Data Access Details

The co-author visualization gets its data **after** the page is loaded.

1. The co-author visualization is given a URL to the data it needs when the page is loaded (flashvars).
2. The co-author visualization requests the network data from the URL.
3. The server returns the network data to the visualization.
4. The co-author visualization begins to load in the page, using the network data.

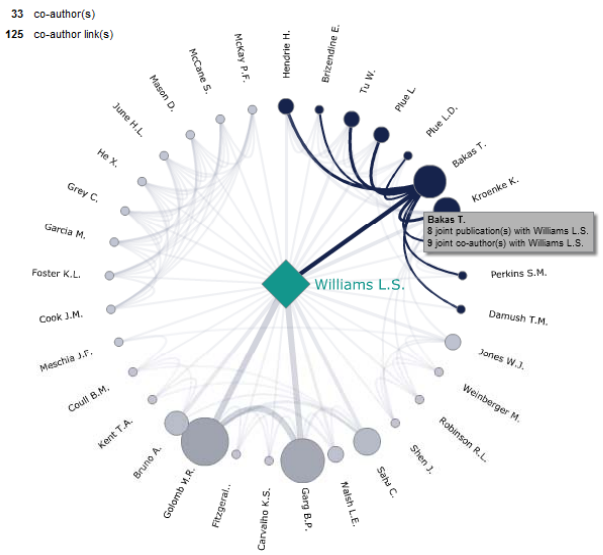
Network Clustering



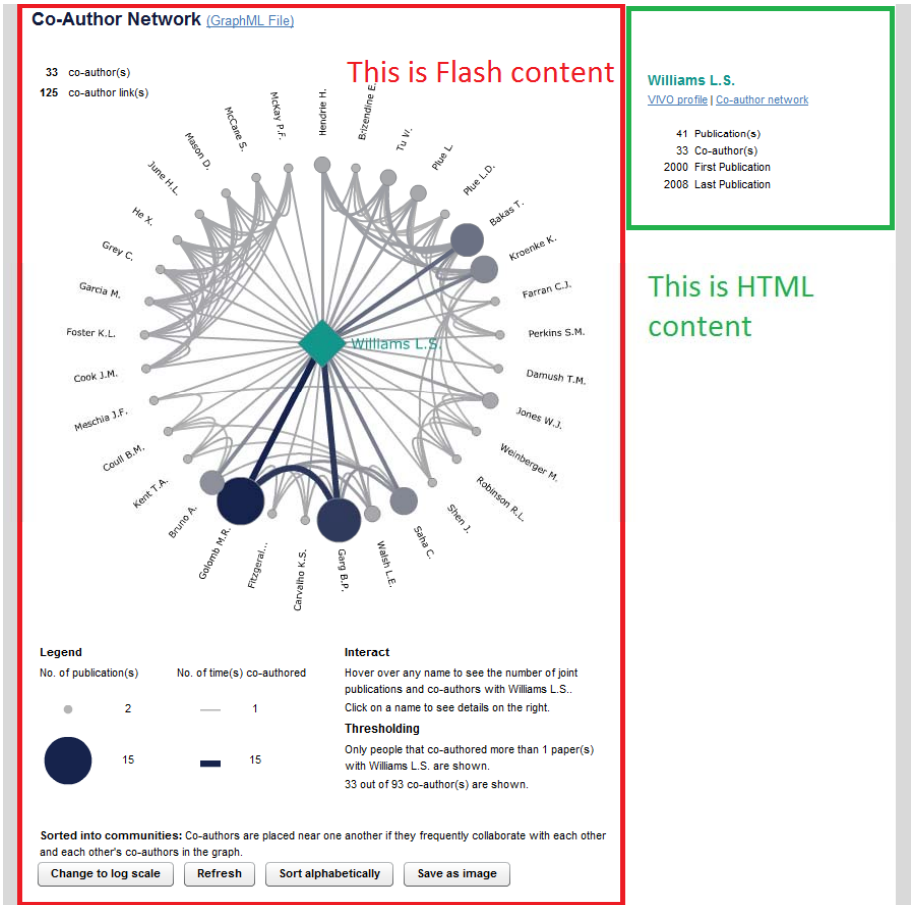
- Uses 'Agglomerative Hierarchical Clustering'
  - Assign each author to a separate cluster.
  - Evaluate all pair-wise distances between clusters (distance metrics are described in Distance Metrics Overview).
    - We use Cosine similarity, where authors are more similar if they tend to collaborate with the same authors with around the same frequency.
  - Construct a distance matrix using the distance values.
  - Look for the pair of clusters with the shortest distance.
  - Remove the pair from the matrix and merge them.
  - Evaluate all distances from this new cluster to all other clusters, and update the matrix.
  - Repeat until the distance matrix is reduced to a single element.

Process adapted from http://www.improvedoutcomes.com/docs/WebSiteDocs/Clustering/Agglomerative_Hierarchical_Clustering_Overview.htm

Interactivity



- Clicking an author node…
  - Shows the # of publications and co-authors shared with the ego.
  - Highlights that author's collaboration links.
  - Brings up extra information and links in the sidebar.



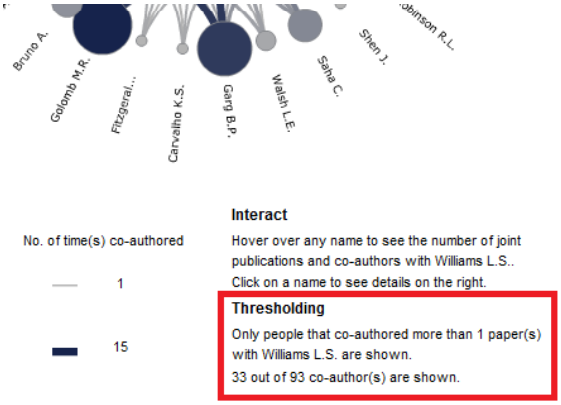
VIVO
Enabling a National Network of Scientists

Flash + HTML

- Adobe Flex lets Flash content communicate with external Javascript functions through the "External Interface" API (application programming interface).
- Javascript functions in the page modify the HTML content to show the information for the person who was clicked.

Thresholding

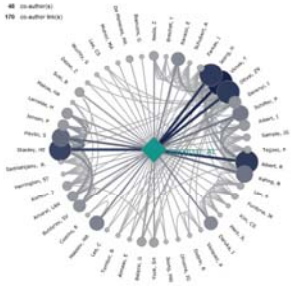
- Problem: Readability decreases significantly with >50 authors.
- Solution: Remove authors who have collaborated with the ego once, then twice, then three times, etc… until there are less than 50 authors shown.
- Result: Only the most important authors are included.
  - (full author list is via .csv download link).



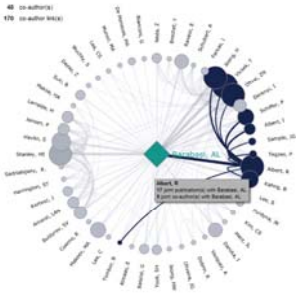
Other Features

- Log or Linear Scaling
  - Users can switch between logarithmic and linear scaling for node and edge sizes.
  - This is provided by Flare.
- Save as image
  - Users can save a screenshot of the visualization.
  - This is provided by Flex.
- Download graphml file.
  - Users can download the full network in a standard format for further processing and analysis (more later).
  - The file you download is identical to the data provided to the co-author visualization from the VIVO server.

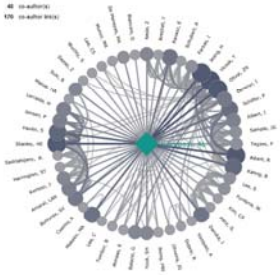
Different Layouts of a Network



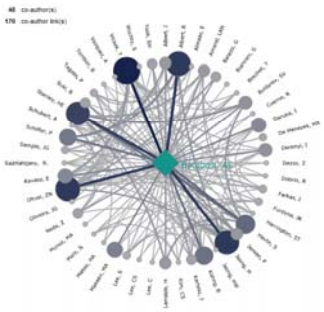
Clustering



Co-author Selected



Log scale



Alphabetical sorting

Visualization Details

- Co-author visualization
- Sparklines
- Data download and analysis

Concept

- Proposed by Edward Tufte
  - "small, high resolution graphics embedded in a context of words, numbers, images"
  - Also, "data-intense, design-simple, word-sized graphics"
- Information graphic characterized by small-size & data intensity
- E.g. OMB data shows the ebb and flow of the deficit from 1983 – 2003.
- More information at wikipedia entry for sparkline & Tufte's article on sparkline.

36 publication(s) from 2001 to 2010 (.CSV File)

173 publication(s) within the last 10 years incomplete data

View all VIVO publications and corresponding co-author network.

80 co-author(s) from 2001 to 2010 (.CSV File)

Why Sparklines?

- To succinctly convey activity of an individual
- Sparklines by definition convey this in a simple and condensed way.
- Currently we have,
  - Number of publications over the years 36 publication(s) from 2001 to 2010 (.CSV File)
  - Number of Co-authors over the years 80 co-author(s) from 2001 to 2010 (.CSV File)
- But, Quantity != Quality
- Also, "less is good" might be the mantra in the domain of that individual.
- But, Pros >> Cons

Code Libraries Used

- Google visualization API
  - Google provides visualizations[1] built using it's API in JavaScript.
    - Visualizations are free for anyone to use, and relatively easy to work with.
    - Works in most modern browsers, including iPhone/iPad.
  - Provides ways to manipulate data like creating views based on certain conditions e.g. consider only rows of data that having dates later than *1st January, 2001*
  - We use Sparkline Visualization API[2]
  - Sparkline images contained in HTML table are rendered using jQuery JavaScript library API.

see next slide for code snippet

[1] - http://code.google.com/apis/visualization/documentation/gallery.html
[2] - http://code.google.com/apis/visualization/documentation/gallery/imagesparkline.html

Used Code Libraries

- Google visualization API
  - Code snippet & Output

```
google.load("visualization", "1",
                {packages:["imagesparkline"]});

google.setOnLoadCallback(drawChart);
```

Load Google Sparkline API activator

```
var data = new google.visualization.DataTable();
```

Create JS data object

```
data.addColumn('string', 'Year');
data.addColumn('number', 'Publications');
data.addRows(10);
data.setValue(0, 0, '2001');
data.setValue(0, 1, 2);
data.setValue(8, 0, '2009');
data.setValue(8, 1, 0);
<more  data entry>
data.setValue(9, 0, '2010');
data.setValue(9, 1, 1);
```

Populate JS data object

Publications	

```
var chart = new google.visualization
                        .ImageSparkLine($('#chart_div'));
```

Create sparkline object

```
chart.draw(data, {width: 63,  height:21,
                  showAxisLines: false,
                  showValueLabels: false,
                  labelPosition: 'left'});
```

Draw sparkline using data object & other options

Code Libraries Used

- jQuery
  - a fast and concise JavaScript library[1]
  - simplifies HTML document traversing, event handling, animating, and AJAX interactions
  - We use it extensively to,
    - fetch data dynamically via AJAX calls,
    - display new tables on the fly
    - to activate the Visualizations
  - Also because it provides uniform behavior on all major browsers.

[1] - http://jquery.com/

Different Sparkline Render Modes

- 2 render modes - short & full.
- Short sparkline,
  - On the main profile page it is a good idea to display a sparkline that considers the publications from just the last 10 years.
  - Below is an example of a "short" sparkline for an individual

24 publication(s) within the last 10 years incomplete data

- Full sparkline,
  - On the person level visualization page we display a sparkline that considers the publications spanning the entire career of an individual.
  - Below is an example of a "full" sparkline for the same individual

25 publication(s) from 1995 to 2010 (.CSV File)

VIVO Sparklines

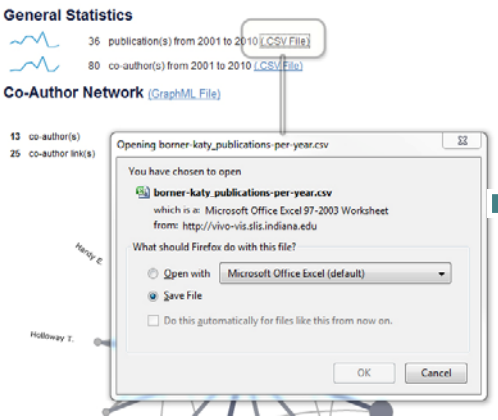
- When rendering the sparkline all the years in a particular time span are considered, not just the actual publication years.
  - E.g. an individual first published in 1995 & then in 2005 & 2010 then the sparkline visualization will consider the intervening years to have 0 publications.
  - So in effect you will have 15 data points (1995, 1996, 1997 ... 2009, 2010) rather than just 3 (1995, 2005, 2010).
- For cases when we cannot determine publication year we consider the publication to have "Unknown" year.
- Any sparkline will have a minimum time span of 10 years
  - E.g. an individual first published in 2005 but we will display a sparkline starting 10 years before the current year, which in current scenario means from 2001.

Data Used for Visualization

Server	• Java objects are created from SPARQL result parser
Server	• Traverse java objects, and • Generate JavaScript (JS) code for creating JS objects once activated
Server	• Check if short sparkline is requested • If so, then generate JS code for filtering out publications that were published before 10 years from current date
Client	• JS code is activated • JS objects are created
Client	• JS objects are passed to the Google sparkline API • Sparkline API returns with sparkline image embedded in an HTML table

Data Download

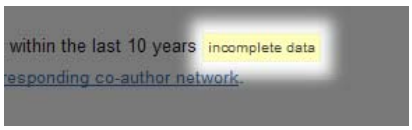
- Data used to render the sparkline visualization is made available in a simple .CSV format.
- Link to the file is just along side the respective sparkline visualization. Usually annotated as "(.CSV File)".
- Same data as used for rendering the sparkline but formatted as .CSV & made available to the user in a file.
- User is prompted to save or open the file for consumption.



Year	Publications
2001	2
2002	4
2003	2
2004	7
2005	7
2006	3
2007	10

"Incomplete Data" Disclaimer

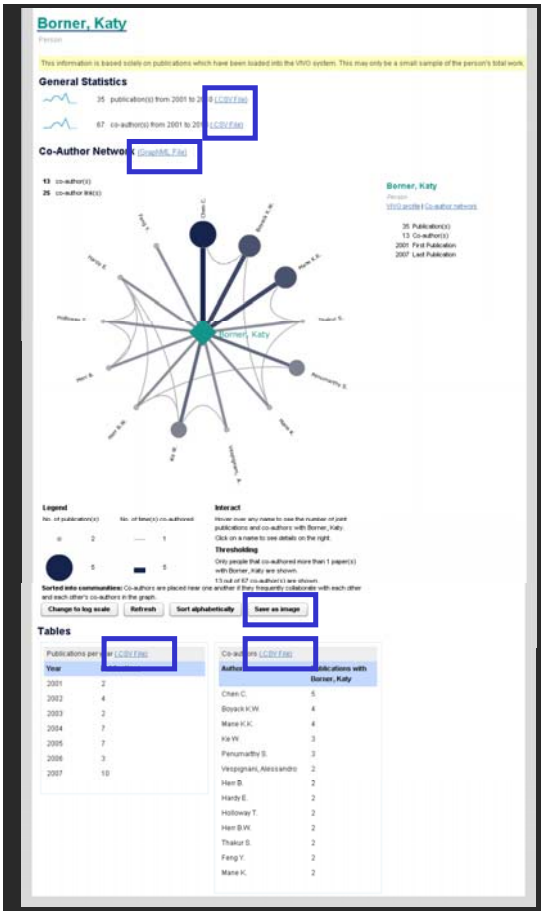
- Accuracy of the visualization is only as good as the underlying data
- There are always exceptions like,
  - not all publications entered for a particular individual,
  - wrong individual is tagged with a publication etc.
- To guard against these discrepancies we display a disclaimer along side the visualizations.



- At later stages we intend to put up a link to which a user can follow-up if it notices discrepancies in the data.

Visualization Details

- Co-author visualization
- Sparklines
- Data download and analysis



Download Data

General Statistics

- 36 publication(s) from 2001 to 2010 (.CSV File)
- 80 co-author(s) from 2001 to 2010 (.CSV File)

Co-Author Network

(GraphML File)

Save as Image (.PNG file)

Tables

- Publications per year (.CSV File)
- Co-authors (.CSV File)

http://vivo-vis.slis.indiana.edu/vivo1/visualization?uri=http%3A%2F%2Fvivoweb.org%2Fontology%2Fcore%2FPerson72&vis=person_level&render_mode=standalone

36 publication(s) from 2001 to 2010 (.CSV File)

Year	Publications
2001	2
2002	4
2003	2
2004	7
2005	7
2006	3
2007	10
2010	1

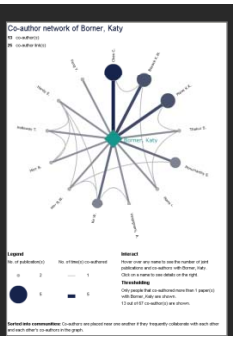
80 co-author(s) from 2001 to 2010 (.CSV File)

Year	Count	Co-Author(s)
2001	1	Chen C.
2002	3	Chen C.; McMahon T.; Feng Y.
2003	2	Chen C.; Boyack K.W.
2004	17	Sengupta A.; Penumarthy S.; Thakur S.; Sooriamurthi R.; Maru J.T.; Shiffrin R.M.; Mane K.; Moor K.A.;

Co-author network (GraphML File)

```
<?xml version="1.0" encoding="UTF-8"?>
    <graphml xmlns="http://graphml.graphdrawing.org/xmlns"
  xmlns:xsi="http://www.w3.org/2001/XMLSchema-instance"
  xsi:schemaLocation="http://graphml.graphdrawing.org/xmlns
  http://graphml.graphdrawing.org/xmlns/1.0/graphml.xsd">
<key id="label" for="node" attr.name="label" attr.type="string" />
<key id="number_of_authored_works" for="node" attr.name="number_of_authored_works" attr.type="int" />
<key id="num_unknown_publication" for="node" attr.name="num_unknown_publication" attr.type="int" />
<key id="num_latest_publication" for="node" attr.name="num_latest_publication" attr.type="int" />
<key id="latest_publication" for="node" attr.name="latest_publication" attr.type="int" />
<key id="profile_url" for="node" attr.name="profile_url" attr.type="string" />
```

Save as Image (.PNG file)

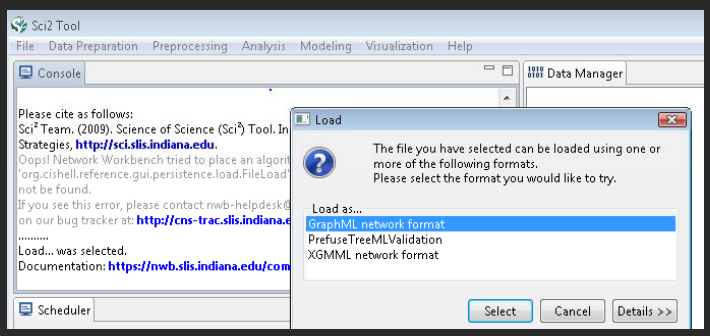


Publications per year (.CSV File), see top file.

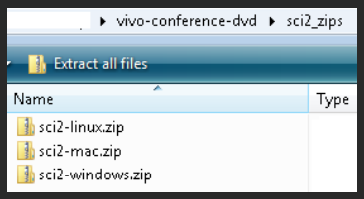
Co-authors (.CSV File)

Co-Author	Count
Andrienko G.	1
Andrienko N.	1
Ben-Miled Z.	1
Blackwell A.	1
Boyack K.W.	4
Bozicevic M.	1
Brodbeck D.	1
Burkhard R.A.	1
Chen C.	5

Load Co-Author Network (GraphML File)

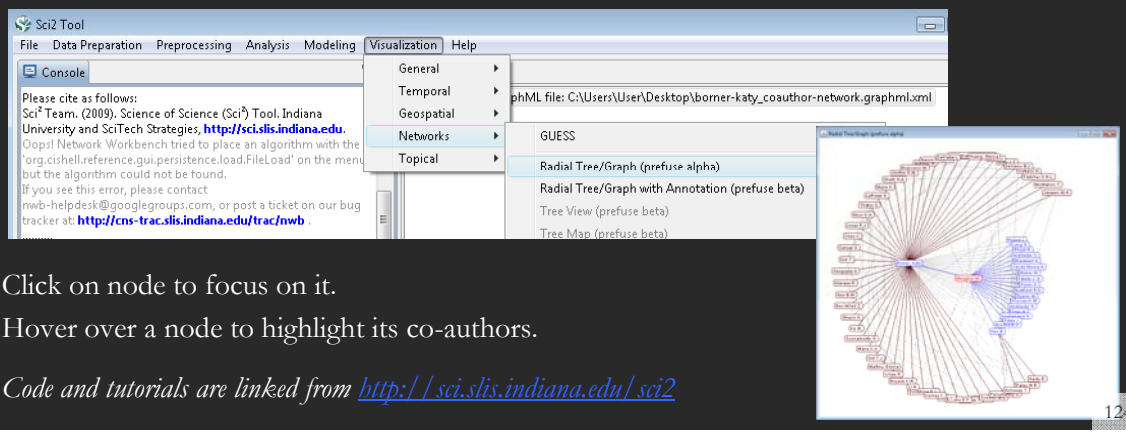


Unzip and Run Sci2 Tool



Network Analysis Toolkit
Nodes: 81
Edges: 390

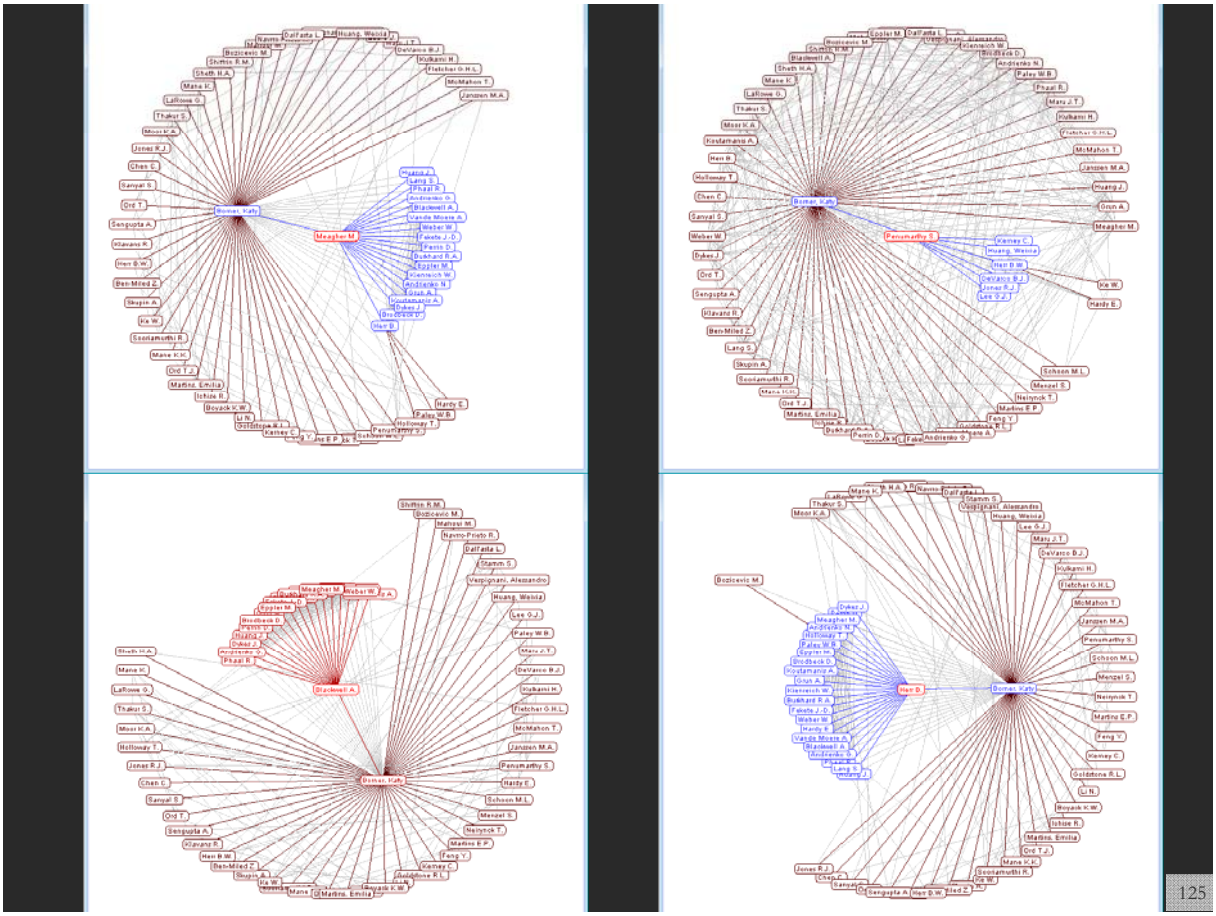
Visualize the file using Radial Graph layout.



Click on node to focus on it.

Hover over a node to highlight its co-authors.

Code and tutorials are linked from http://sci.slis.indiana.edu/sci2



Hands-on

- Creating and Executing SPARQL queries
- Creating visualizations based on SPARQL query results

Playing with SPARQL

- Please visit our SPARQL query end-point specially created for the workshop at http://vivo-vis-test.slis.indiana.edu/vivo/workshop/sparqlquery/
  - This link will be disabled after the workshop due to technical reasons. In that case you will find the results already saved out for you in the "files" folder of the DVD.
  - You can also access the hands-on workshop material here.
  - We will be referring to this material throughout our tutorial.

- Now let us try different queries,
  - Get name of every person in the system
  - Get titles of all the publications in the system
  - Get titles of all the publications for a particular person in the system
  - Get all the co-authors for a particular person in the system

- Later we will try saving out the results in different formats.

Playing with SPARQL

- Get name of every person in the system

```
PREFIX rdf:   <http://www.w3.org/1999/02/22-rdf-syntax-ns#>
PREFIX rdfs:  <http://www.w3.org/2000/01/rdf-schema#>
PREFIX foaf: <http://xmlns.com/foaf/0.1/>

SELECT    (str(?person) as ?personURI)
          (str(?personName) as ?personNameLit)
WHERE
{
  ?person rdf:type foaf:Person .
  ?person rdfs:label ?personName .
}
```

Playing with SPARQL

- Get titles of all the publications in the system

```
PREFIX rdf:   <http://www.w3.org/1999/02/22-rdf-syntax-ns#>
PREFIX rdfs:  <http://www.w3.org/2000/01/rdf-schema#>
PREFIX bibo: <http://purl.org/ontology/bibo/>

SELECT    (str(?publication) as ?publicationURI)
          (str(?publicationTitle) as ?publicationTitleLit)
WHERE
{
  ?publication rdf:type bibo:Document .
  ?publication rdfs:label ?publicationTitle .
}
```


Playing with SPARQL

- Get titles of all the publications for a particular person in the system

```
PREFIX rdf:   <http://www.w3.org/1999/02/22-rdf-syntax-ns#>
PREFIX rdfs:  <http://www.w3.org/2000/01/rdf-schema#>
PREFIX bibo: <http://purl.org/ontology/bibo/>
PREFIX foaf: <http://xmlns.com/foaf/0.1/>
PREFIX core: <http://vivoweb.org/ontology/core#>

SELECT (str(?publicationTitle) as ?publicationTitleLit)

WHERE {
          <http://vivo-trunk.indiana.edu/individual/Person72> rdf:type foaf:Person .
          <http://vivo-trunk.indiana.edu/individual/Person72> rdfs:label ?authorLabel .
          <http://vivo-trunk.indiana.edu/individual/Person72> core:authorInAuthorship
          ?authorshipNode .

          ?authorshipNode rdf:type core:Authorship ; core:linkedInformationResource
                    ?publication .
          ?publication rdf:type bibo:Document .
          ?publication rdfs:label ?publicationTitle .
}
```

- <http://vivo-trunk.indiana.edu/individual/Person72> denotes the particular person

Playing with SPARQL

- Get all the co-authors for a particular person in the system

```
PREFIX rdf:   <http://www.w3.org/1999/02/22-rdf-syntax-ns#>
PREFIX rdfs:  <http://www.w3.org/2000/01/rdf-schema#>
PREFIX bibo: <http://purl.org/ontology/bibo/>
PREFIX core: <http://vivoweb.org/ontology/core#>
PREFIX foaf: <http://xmlns.com/foaf/0.1/>

SELECT DISTINCT          (str(?authorLabel) as ?authorLabelLit)
                         (str(?coAuthorLabel) as ?coAuthorLabelLit)
WHERE {
            <http://vivo-trunk.indiana.edu/individual/Person72> rdf:type foaf:Person .
            <http://vivo-trunk.indiana.edu/individual/Person72> rdfs:label ?authorLabel .
            <http://vivo-trunk.indiana.edu/individual/Person72> core:authorInAuthorship
                        ?authorshipNode .
            ?authorshipNode rdf:type core:Authorship ; core:linkedInformationResource ?publication .
            ?publication rdf:type bibo:Document .
            ?publication rdfs:label ?publicationTitle .
            ?publication core:informationResourceInAuthorship ?coAuthorshipNode .
            ?coAuthorshipNode core:linkedAuthor ?coAuthor.
            ?coAuthor rdfs:label ?coAuthorLabel.
            FILTER (<http://vivo-trunk.indiana.edu/individual/Person72> != ?coAuthor)
}
```

- <http://vivo-trunk.indiana.edu/individual/Person72> denotes the particular person

Playing with SPARQL

- Now let us try saving out the results in different formats.
  - Plain text (use option "RS_TEXT")
  - Downloadable CSV file (use option "CSV")
  - JSON (use option "RS_JSON")
  - RDF (use option "RS_RDF")
- We will be using results in JSON format to be utilized in next section.
  - JSON is used because of it's human readability advantage.
  - Also, jQuery has a built-in parser for it so we don't have to write one to parse the results during the workshop.

Hands-on

- Creating and Executing SPARQL queries
- Creating visualizations based on SPARQL query results

Creating Visualizations

VIVO Enabling a National Network of Scientists

- Use "template.html" as the base for each visualization we create.
  - This has clearly demarcated area for playing around with the code.
    - Demarcations are in the form,
      - <!–

        START/END OF WORKSHOP SPECIFIC MOD AREA

        -->
    - We would be working only in that portion of the file.
- The general flow of our code would be,
  - Get data
  - Use data to create JS objects
  - Use JS objects to create visualization object
  - Render visualization in specific portion in the final page
  - Specify exception case
- We will start with a simple word cloud visualization.

Creating Visualizations

- How to get data to be used in the visualization?

```
/* jQuery provides a nice API ".getJSON" to grab JSON content from a file and make it available for
manipulation. */
$.getJSON('all-publications-titles.txt', function(jsonData) {

/* Here after we grab the JSON data we traverse through all the values. */
        $.each(jsonData.results.bindings, function(index, item){

/* Here we print the index i.e. nth publication & the title for all the publications in a pre-defined area – a DIV
with id "information" - in the page.
Note that "item.publicationTitleLit.value" is used to access the publication title in the parsed JSON object.
*/
        $('#information').append(index + " -> " + item.publicationTitleLit.value + "<br />");
        });
});
```

- Check "template-pub-wc-A.html" in "archive" folder to see this code snippet in action.

Creating Visualizations

- How to create JS objects?
  - We will be using google visualization API for creating objects.

```
/* google provides ways to create data objects & also functions for manipulation. Here we create a a
"datatable" object and then fill our JSON data in it.*/
var data = new google.visualization.DataTable();
data.addColumn('string', 'Title');

        $.each(jsonData.results.bindings, function(index, item){
        $('#information').append(index + " -> " + item.publicationTitleLit.value + "<br />");

/* Here we add the publication title value to the "data" object which will be used later to actually generate the
visualization. */
        data.addRow([item.publicationTitleLit.value]);
        });
});
```

- Check "template-pub-wc-B.html" in "archive" folder to see this code snippet in action.

Creating Visualizations

- How to use JS objects to create & then render visualization object?
  - Specific visualization library we use to generate word cloud requires google's data objects as input.

```
/* This selects the id of the DIV object which will serve as the placeholder for the eventual visualization. */
var outputDiv = document.getElementById('wcdiv');

/* This creates the visualization object. */
var wc = new WordCloud(outputDiv);

/* This feeds the data to the visualization object along with a parameter for "stop word". Stop words are used
so that the visualization is not taken ovber by words ina publication's title that do not provide any insight like
"the, a, an" etc. So the words mentioned here will not be considered for generating the Word Cloud. */
wc.draw(data, {stopWords: 'a an and is or the of for to in by on with as but S at it not '
                        + 'than do use like from its are how au who via among that this can into'});
```

- Check "template-pub-wc-B.html" in "archive" folder to see this code snippet in action.

Creating Visualizations

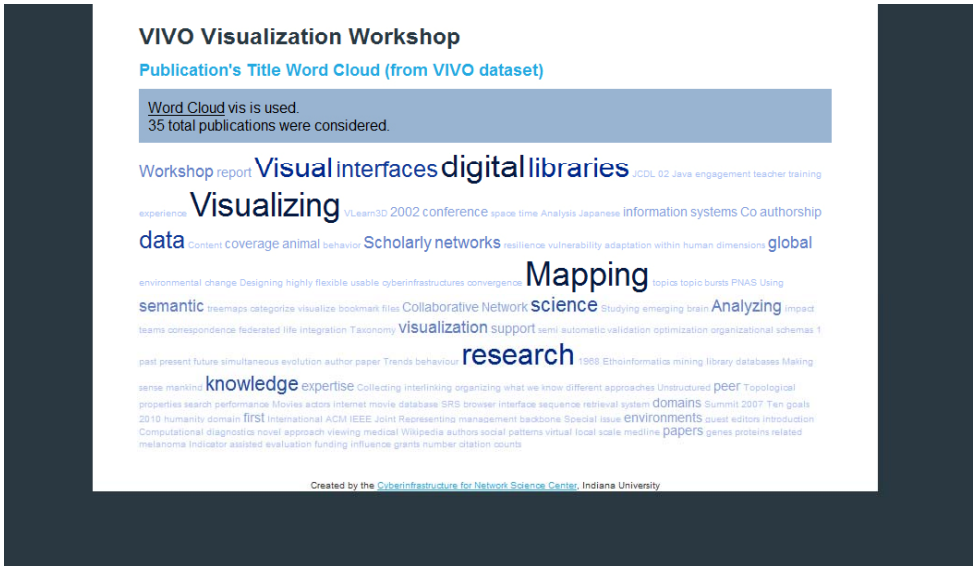
- How to handle exceptions?
  - There might be chances that creating a visualization on the fly might slow your machine down. In such cases it is better to try out smaller dataset first.
  - We can control the number of records we want to consider for the word cloud by mentioning a simple condition during JS object creation.

```
/* Here we specify number of publications we want to use for our visualization. */
var numPublicationsConsidered = 100;

/* This will keep track of the total publications present in our visualization. This will be different form
"numPublicationsConsidered" only when total publications in the dataset is less than 100 (in this case). */
var totalPublicationsPresent = 0;

/* We introduce this condition when we are traversing the JSON data such that only create new data object if
number of publications considered till now is less than a predefined value. */
        if (index < numPublicationsConsidered) {
                data.addRow([item.publicationTitleLit.value]);
                totalPublicationsPresent = index;
        } else { return false; }  /* This will cause JSON data traversal to cease.*/
```

- Check "template-pub-wc-C.html" in "archive" folder to see this code snippet in action.

Creating Visualizations

- What is the final output?



- Check "publication-word-cloud.html" in the root folder to see this visualization in action.

Creating Visualizations

- Similarly we will now create our next visualization – Person Stats.
  - This is used to show some of the stats for a person viz., total number of publications & co-authors that the person has published with.
  - We have already covered the basics earlier on.
  - Here I will point out the differences.
  - We will get data from 2 different files, one for publication titles & other for co-authors.

```
/* We process the file containing all the co-authors a person has published with. */
$.getJSON('katy-coauthors.txt', function(jsonData) {
/* Manipulate co-authors data */
});

/* We process the file containing all the publications of a person. */
$.getJSON('katy-publications-titles.txt', function(jsonData) {
/* Manipulate co-authors data */
});
```

- Check "template-person-A" in "archive" folder to see this code snippet in action.

Creating Visualizations

- How to get the stats?
  - Our main goal is to draw horizontal bars based on the number of publication (& co-authors)
  - So we will collect that data next.

```
/* This will store the stats that are required to render the visualization. */
var personName;
var totalPublications = 0;
var totalCoAuthors = 0;

/* This will render the name of the person in out visualization. Note that "authorLabelLit" stores name of the person in our JSON file. */
personName =  jsonData.results.bindings[0].authorLabelLit.value;

/* While traversing through co-authors JSON we increment a counter for each co-author. */
$.each(jsonData.results.bindings, function(index, item){
        totalCoAuthors++;
});

/* Similarly while traversing through publications JSON we increment a counter for each publication. */
totalPublications++;
```

- Check "template-person-B" in "archive" folder to see this code snippet in action.

Creating Visualizations

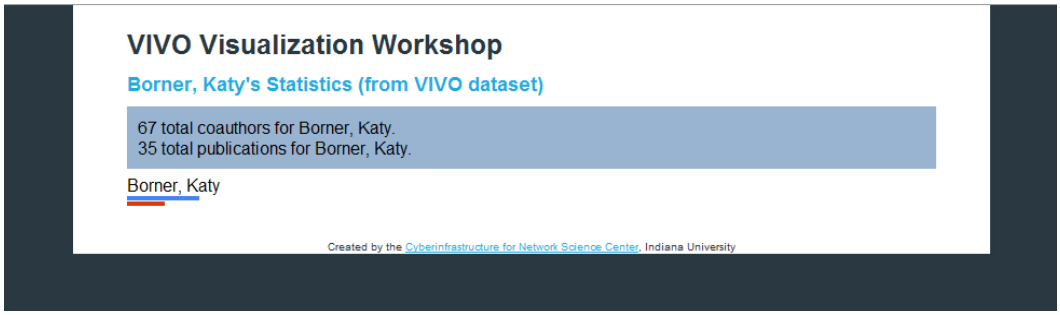
- How to render the visualization?
  - We will be rendering 3 items,
    - Name of the person
    - Horizontal bar denoting number of co-authors
    - Horizontal bar denoting number of publications

```
/* This will set content of an HTML object denoted by class "person-name" to name of the person. */
$('.person-name').text(personName);

/* This will set width of an HTML object denoted by id "actual-coauthor-count" to be proportional to "totalCoAuthors". We might want to adjust or normalize that value before setting the width. */
$('#actual-coauthor-count').css('width', getAdjustedWidth(totalCoAuthors));

/* This will set width of an HTML object denoted by id "actual-publication-count" to be proportional to "totalPublications". We might want to adjust or normalize that value before setting the width. */
$('#actual-publication-count').css('width', getAdjustedWidth(totalPublications));

/* This function can be used to adjust the value. This will come handy especially in cases when number of co-authors for a person will exceed 1000. Currently 1 co-author translates to 1pixel of width of the bar. In this case it will mean that our horizontal bar might go off the screen. Ideally we would not want that. So we will use this function to resolve this. */
function getAdjustedWidth(originalWidth) {
        var newHeight = originalWidth * 1;
        return newHeight;
}
```

- Check "template-person-C" in "archive" folder to see this code snippet in action.

Creating Visualizations

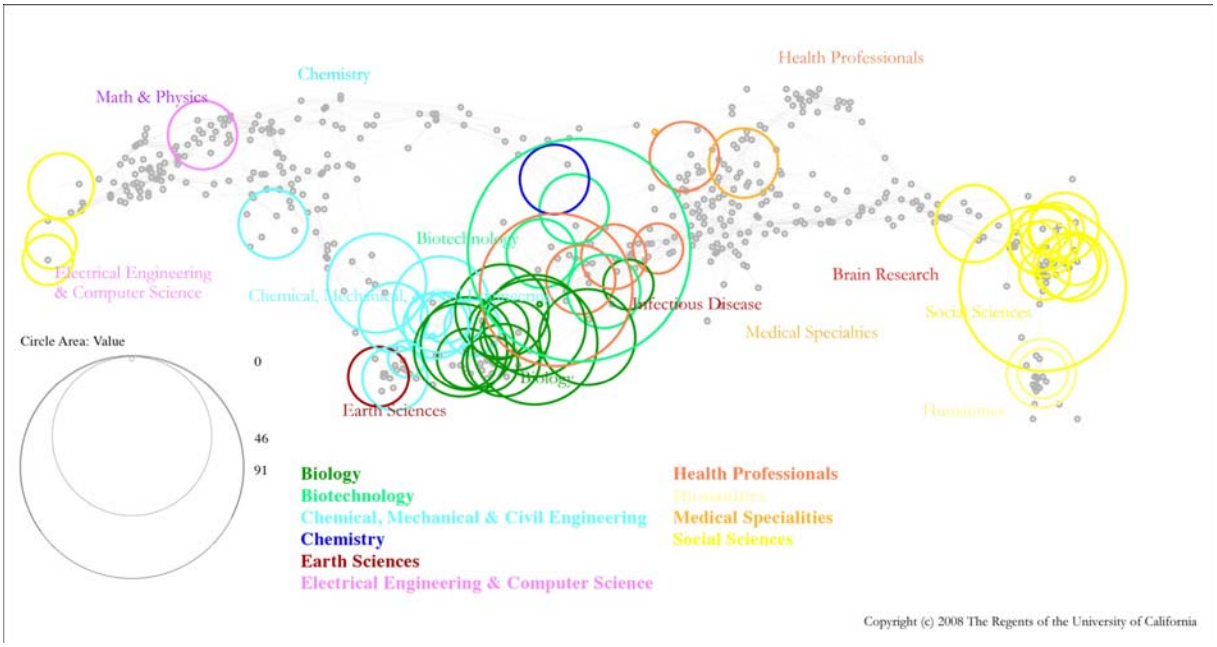
- What is the final output?



- Check "person-stats" in the root folder to see this code snippet in action.

Outlook

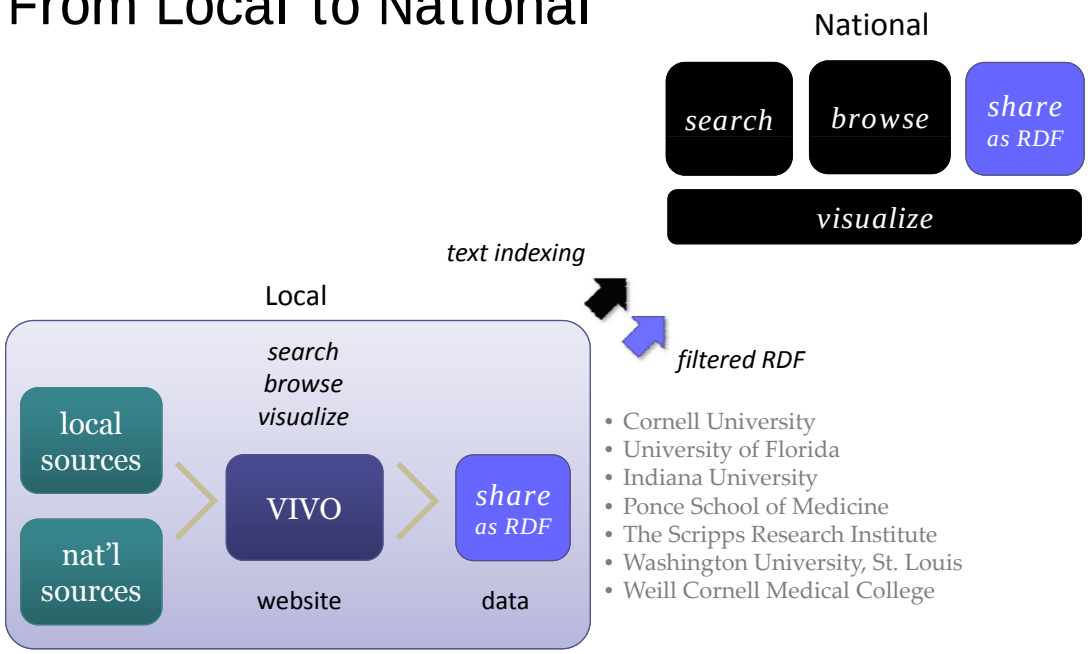
- Institution Level Visualizations Under Development
  - Science Maps
  - Comparison visualization
- National Level Visualizations



(Generated using dummy data. The values shown here are not real).

(Generated using dummy data. The values shown here are not real).

From Local to National



- Cornell University
- University of Florida
- Indiana University
- Ponce School of Medicine
- The Scripps Research Institute
- Washington University, St. Louis
- Weill Cornell Medical College

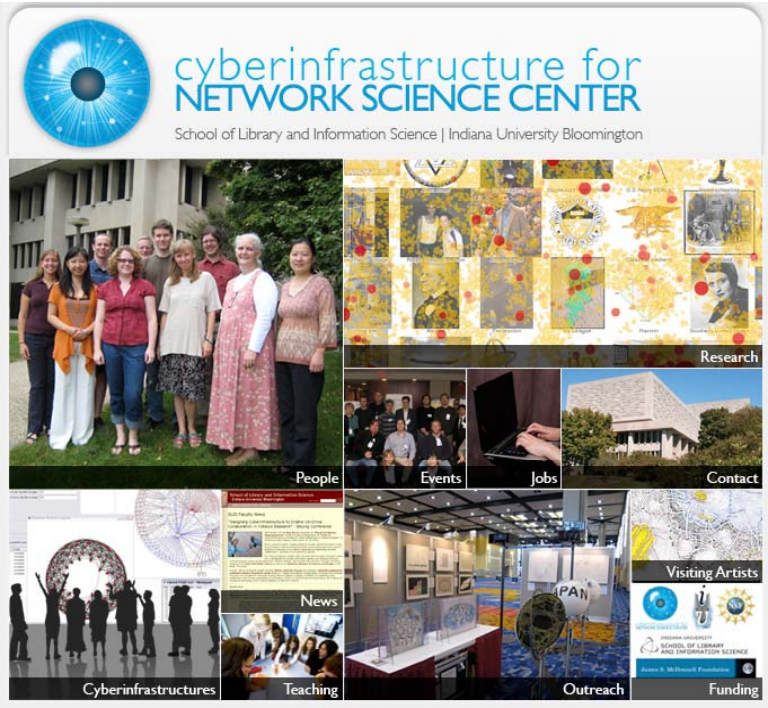
VIVO Documentation, Code, and Data

Documentation and Code:

- VIVO Web Site: http://www.vivoweb.org
- VIVO Support: http://www.vivoweb.org/support
- VIVO Ontology: http://www.vivoweb.org/download#ontology
- Sourceforge for source code: http://sourceforge.net/projects/vivo

Data from VIVO Instances:

- http://vivo.iu.edu [Test Environment]
- http://vivo-on-vivo.mannlib.cornell.edu [release 1.0 -- Test environment]
- http://vivo.med.cornell.edu [release 1.0 -- production system]
- http://vivo.scripps.edu [release 1.0 -- production system]
- http://vivo.ufl.edu [release 1.0]
- http://vivo.psm.edu [release 1.0]
- http://vivo.cornell.edu [original web site]
- http://vivo.wustl.edu [release 1.0 -- production system]
- http://vivo-vis.slis.indiana.edu [Test Environment, 1.1]



All papers, maps, cyberinfrastructures, talks, press are linked from http://cns.slis.indiana.edu