Engineering Research Center Observatory Visualizations

Katy Börner (@katycns)

Joint work with Gerhard Klimeck & Michael Zentner, Purdue University
NSF Supplement, Dec 2015 – Nov 2017

Cyberinfrastructure for Network Science Center
Department of Intelligent Systems Engineering & Department of Information and Library Science
School of Informatics and Computing and IU Network Science Institute
Indiana University, Bloomington, USA

Virtual Presentation to NSF EEC Staff

June 1, 2017
• Dec 09, 2015: Gerhard Klimeck, Michael Zentner, and Katy Börner present *Engineering Research Center Observatory* at Engineering Research Center Observatory Kick-Off Meeting, Washington, DC.

• Dec 12, 2016: Katy Börner presents *Visualizing Nanoscience and Technology* at 2016 NSF Nanoscale Science and Engineering Grantees Conference, Arlington, VA.

• June 1, 2016: Katy Börner presents *ERC Observatory Visualizations*
Goal, Use Cases, Users, Data, Visualizations
Engineering Observatory: Goal

Facilitate near real-time monitoring of Engineering Research Centers (ERCs) in support of better-informed resource allocation, priority setting, and evaluation.

Relevant data streams comprising publications, patents, funding, scientific datasets, code will need to be federated.

Data mining and visualization web services will be provided for different stakeholders (NSF staff, researchers, students) to increase their understanding of temporal, geospatial, topical, and network patterns and trends in engineering.

This collaborative work with the nanoHUB team at Purdue University is funded by NSF, Dec 15 – Nov 17.
Engineering Observatory: Use Cases

Use Cases:

• Day-to-day operations
• Strategic decision making
• Prepare for site visits

Initial Power Users:

• Mehmet Ozturk, Nanosystems ERC for Advanced Self-Powered Systems of Integrated Sensors and Technologies (ASSIST)
• Paul Westerhoff, Nanotechnology Enabled Water Treatment Systems (NEWT)
• Greg Carman, Nanosystems ERC for Translational Applications of Nanoscale Multiferroic Systems (TANMS)
Engineering Observatory: Data & Visualizations

Data
• Planned: Annual progress reports—data cannot be extracted in high quality algorithmically
• Currently: nanoHUB data + bibliography files
• Future: Scholarly Database and IUNI WoS data at IU; data collected in machine-readable format by ERCs

Visualizations
• Evolving collaboration networks obtained from bibliography files, network layout or overlaid on geographic map
• Evolving expertise profiles; overlaid onto the UCSD map of science
• Others TBD.
Interactive Data Visualizations
Importing Data

https://lab1.pharmahub.aws.hubzero.org/erc-citations
Curating Data

After the file is loaded, brief summaries will be shown for each entry. You may click any entry to expand it and give a full, editable view of the properties of that entry.

Here, one should make sure that fields that are pertinent to the visualizations being performed are correct and consistent across entries.

Name disambiguation: If the same author is represented by multiple variations of their name in the visualization, use the “Author Search” feature to find all variations and choose a canonical form for them.

Sex: User is provided the choice to select either Female, Male, or Unreported.

Role: User is able to select the role of each author, e.g., Student, Staff, Postdoc.

Geolocation: The location field accepts several types of input including city names and zip codes. After entering that information, either press ‘Enter’ or move on to the next field and in a few seconds the field will reflect whether geocoding information was found, turning green for success and red for failure.
Downloading and Visualizing Data

At the end of the page there are buttons to download changes as BibTeX or to launch a number of visualizations that highlight different aspects of the data.

Article — European Physical Journal-Special Topics  
Theoretical and technological building blocks for an innovation accelerator  

Article — Journal of Informetrics  
Approaches to understanding and measuring interdisciplinary scientific research (IDR): A review of the literature  
Wagner, C. S. and Roessner, J. D. and Bobb, K. and Klein, J. T. and Boyack, K. W. and Keyton, J. and Rafols, I. and Borner, Katy

Article — Scientometrics  
Mapping interactions within the evolving science of science and innovation policy community  
Zoss, A. M. and Borner, Katy

Download BibTeX

Visualize 29 journals using map of science

Visualize author geolocation network

Visualize co-authorship network

Visualizations open in a new window, ensure you have enabled popups for the hub.
Co-Author Visualization – Network Layout

Each node in the figure represents an author, and author node area size scales with the number of publications. Author nodes with three publications or more are labelled by the author’s name. Two authors are connected if they have authored a publication together and link width scales with the number of joint publications between those authors.
Co-Author Visualization – Geomap (Forthcoming)
Expertise Profile Visualization -- UCSD Map of Science

Maps of science can be used to explore, understand, and communicate the expertise profiles of institutes or nations; to chart career trajectories; to identify emerging research frontiers. They allow us to track the emergence, evolution, and disappearance of topics and help to identify the most promising areas of research.

ASSIST
NEWT (new ERC)

ASSIST (most pubs are in Chem, Mech & Civil Eng, EE and CS, Math and Physics)

TANMS (most interdisciplinary, *Nature* paper)

Legend
Circle area: Fractional Journal Count

How To Read This Map
The UCSD map of science depicts a network of 554 subdiscipline nodes that are aggregated to 13 main disciplines of science. Each discipline has a distinct color and is labeled. Overlaid are circles, each representing all records per unique subdiscipline. Circle area is proportional to the number of fractionally assigned records. Minimum and maximum data values are given in the legend.
Visualizations Used in ERC Annual Reports
The figures below show the evolving NEWT collaboration network based on co-authorship extracted from bibliography files (Figure A). Legends are included in the bottom-left of each visualization to explain the sizing of author nodes and collaboration links. Each node represents an author, and author node area size scales with the number of publications. Author nodes with two publications or more are labelled by the author’s name in the top visualization, while author nodes with three publications or more are labelled on the bottom visualization. Two authors are connected if they have authored a publication together and link width scales with the number of joint publications between those authors. Note that size coding differs for the two visualizations.

Network layouts differ structurally as each is spatially optimized for each of the two periods. The network on top shows the 78 authors that published 15 publications in 2015/2016. Note that one publication has 16 authors resulting in a so-called fully connected clique network that is connected to other subnetworks via Pedro J.J. Alvarez. The network below shows 173 authors that published 48 journal articles in 2016/2017. Comparing the two visualizations reveals that NEWT impact in terms of authors and publications has increased considerably—there are about more than twice the number of authors publishing almost four times more publications within 2016/2017 than in previous years. Plus, there are many larger subnetworks that are more interlinked—showcasing intense collaboration and communication. Authors like Pedro J.J. Alvarez, Jorge Gardea-Torresdey, and Paul Westerhoff have not only many publications to their credit but they also interconnect different subnetworks—effectively serving as gatekeepers. Thickness coding of lines supports the identification of major collaboration (and most likely communication) pathways in this rapidly evolving professional network.
NEWT Reporting Year 1

#Papers

3
1

#Co-authored Papers

11
1
NEWT Reporting Year 2

#Papers
- **16**
- **1**

#Co-authored Papers
- **11**
- **1**
ERC Web Table 1 provides a quantifiable summary of the TANMS's research productivity related to the three testbeds. Research productivity, in addition to testbed development, is a required reporting metric by NSF to assess the health and quality of a center’s program. As can be seen in this table TANMS continues to have a healthy number of journal publications, i.e. 36 publications in the last year from core funding and 35 from associated projects or a total of 229 since center inception. Maybe more complete publication information is provided in the coauthor network visualization provided in Figure 3.2.1-1 for the last couple of years.

Each node in the figure represents an author, and author node area size scales with the number of publications. Author nodes with four publications or more are labelled by the author’s name. Two authors are connected if they have authored a publication together and link width scales with the number of joint publications between those authors. The network on top shows the authors network for 2014/2015. The top-three authors with the most publications are Kang L. Wang, Greg P. Carman, and Pedram Khalil Amiri. The network at the bottom of the figure shows publications for 2016/2017. The top-three authors with the most publications are Kang L. Wang, Gregory P. Carman, and Guoqiang Yu. Comparing these two visualizations reveals that the TANMS impact in terms of authors and publications was unusually high in the initial years and is increasing continuously—i.e. there are many more unique authors publishing within 2016/2017 than in previous years. Plus, there are large, strongly interlinked subnetworks—showcasing intense collaboration and communication. Authors like Kang L. Wang and Greg P. Carman, to their credit, also interconnect different subnetworks—effectively serving as gatekeepers. Thus, TANMS has a vibrant publication record showing extensive collaboration across all campuses. This brief snapshot clearly shows that TANMS system/team approach representing the corner stone of an NERC is alive and well.
Visualizations Used in ERC Site Visit Presentations
This network was generated from 271 ASSIST publications. It comprises 48 authors and 128 edges and covers the years 2012 through March 2017. Nodes are sized by the cumulative number of publications associated with an author. Edges are sized by the cumulative number of ASSIST related papers published by an author.
This network was generated from 271 ASSIST publications. It comprises 48 authors and 128 edges and covers the years 2012 through March 2017. Nodes are sized by the cumulative number of publications associated with an author. Edges are sized by the cumulative number of ASSIST related papers published by an author.
ASSIST Co-authorship Network 2012-2017
CNS @ Indiana University
2017

This network was generated from 271 ASSIST publications. It comprises 48 authors and 128 edges and covers the years 2012 through March 2017. Nodes are sized by the cumulative number of publications associated with an author. Edges are sized by the cumulative number of ASSIST related papers published by an author.
This network was generated from 271 ASSIST publications. It comprises 48 authors and 128 edges and covers the years 2012 through March 2017. Nodes are sized by the cumulative number of publications associated with an author. Edges are sized by the cumulative number of ASSIST related papers published by an author.
This network was generated from 271 ASSIST publications. It comprises 48 authors and 128 edges and covers the years 2012 through March 2017. Nodes are sized by the cumulative number of publications associated with an author. Edges are sized by the cumulative number of ASSIST related papers published by an author.
This network was generated from 271 ASSIST publications. It comprises 48 authors and 128 edges and covers the years 2012 through March 2017. Nodes are sized by the cumulative number of publications associated with an author. Edges are sized by the cumulative number of ASSIST related papers published by an author.
This network was generated from 271 ASSIST publications. It comprises 551 authors and 2,511 edges and covers the years 2012 through March 2017. The 48 faculty associated with ASSIST are colored in green. All other authors are colored in light green.
Next Steps & Future Work
Next Steps

• Fully integrate visualizations into nanoHUB. Current code is running/tested in development environment as production changes to nanoHUB have been frozen.

• Optimize current visualizations based on user feedback. Interactive visualizations are best for exploration; static visualizations are required for slides, reports, publications.

• Provide guidance on ERC activity data acquisition—what data should be captured in which format to support what kind of decision making.

• Explore prototypes of additional visualizations that provide novel insights into ERC usage and the impact of resources on S&T progress, see subsequent two slides.

• Publish results in a scholarly paper.
IAI: Expertise Visualization

This visualization is based on publication datasets retrieved from the Scholarly Database at IU, and is used to identify relevant experts, publications, clinical trials, and awards that match a search term.
XDMOD: Sankey Diagram of IT Resource Impact on Research

Visualization displays the relationship between IT resources, funding, and publications. The width of each line represents grant dollars awarded to researchers at one institution.
Future Work (outside current project scope)

- Study and visualize the structure and dynamics of Engineering, particularly the interplay of
  - S&T progress
  - Education and training (residential and online)
  - Job evolution (OOH: New drafts of the job profiles for Electrical & Electronics Engineers and Electrical & Electronics Engineering Technicians to appear in the upcoming next edition of our Occupational Outlook Handbook)

- Communicate S&T dynamics to general audiences via moderated news broadcasts, see next slide.

Related Work:

Upcoming Colloquia

Unless otherwise indicated, most Sackler colloquia are held at the Arnold and Mabel Beckman Center, in Irvine, California.

Reproducibility of Research: Issues and Proposed Remedies
March 8-10, 2017; Washington, D.C.
Organized by David B. Allison, Richard Shifrin and Victoria Stodden
Registration now open

Science of Science Communication III
November 15-16, 2017; Washington, D.C.
Organized by Karen Cook, Baruch Fischhoff, Alan I. Leshner and Dietram A. Scheufele
Registration will open May 2017

Modelling and Visualizing Science and Technology Developments
December 4-5, 2017; Irvine, CA
Organized by Katy Börner, William Rouse and H. Eugene Stanley
Registration will open August 2017

http://www.nasonline.org/programs/sackler-colloquia/upcoming-colloquia
Atlas Trilogy


• ModSTI Conference slides, recordings, and report are at [modsti.cns.iu.edu/report](http://modsti.cns.iu.edu/report)
All papers, maps, tools, talks, press are linked from cns.iu.edu

These slides are at cns.iu.edu/presentations

CNS Facebook: facebook.com/cnscenter

Place & Spaces: Mapping Science Exhibit Facebook: facebook.com/mappingscience