Precision Health Initiative (PHI): Impact Assessment & Strategic Foresight

Katy Börner
Victor H. Yngve Distinguished Professor of Information Science
Director, Cyberinfrastructure for Network Science Center
School of Informatics and Computing and
Indiana University Network Science Institute
Indiana University, USA

PHI SICE Meeting
IUB

August 24, 2018

SICE Research Themes (1/2) Taken from Pedrag, need to update

Variant and Genome Interpretation

- Variant effects
- Allele-specific therapeutic models
- Disease risk models

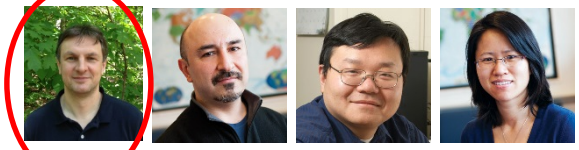

Hahn Radivojac

Partners in PMI GC:

- IUSM
  - Foroud, Li
- IUB
  - Giedroc

Omics Data Analysis and Interpretation

- Cancer genomics
- Proteomics
- Metagenomics
- Metabolomics

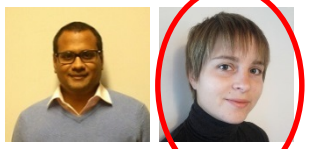

Radivojac Sahinalp Tang Ye

Partners in PMI GC:

- IUSM
  - Yoder
- IUB
  - Giedroc

Clinical Decision Support

- Models for deep reasoning
- Interactive and aid humans
- Learn from data
- Time-based models


Natarajan White

Partners in PMI GC:

- IUSM
  - Foroud
- IUB
  - Radivojac
- Regenstrief
  - Grannis

SICE Research Themes (2/2) Taken from Pedrag, need to update

High-Performance Computing

- Big data storage
- Parallel computing
- Develop Apache Big Data Stack for **all** GCs


Fox


Lumsdaine

Partners in PMI GC:

- Regenstrief
  - Barnett
- IUB
  - Giedroc, Radivojac

Privacy and Security

- Data sharing and privacy protection
- Security from external & internal threats


Myers


Tang

Partners in PMI GC:

- IUSM
  - Foroud, Yoder
- IUB
  - Giedroc
- Regenstrief
  - Barnett, Grannis

Community Engagement + Impact Analysis

- Patient-centric, mobile technologies
- Assessing the success of **all** GCs


Borner


Connelly


Siek

Partners in PMI GC:

- IUB
  - Giedroc
- Regenstrief
  - Grannis

Impact Assessment

IUPUI counter part: Joe Hunt

Provide visual analytics for PHI leads to ease evaluation/reporting burden and to increase their ability to proactively manage teams and research projects.

Among others, create dashboard-like statistics which report the number of

- Team members, including PHI faculty hires and students
- Publications (need acknowledgement text), currently using Google Scholar, soon WoS
- Submitted proposals (need tag like IUNI)
- Awarded funding
- Meetings
- Speakers

Using publication and funding data, visualize collaboration/publication patterns before and after PHI comes into existence--ultimately showing the impact of the GC investment on science.

Co-Author Network BEFORE Precision Health Initiative

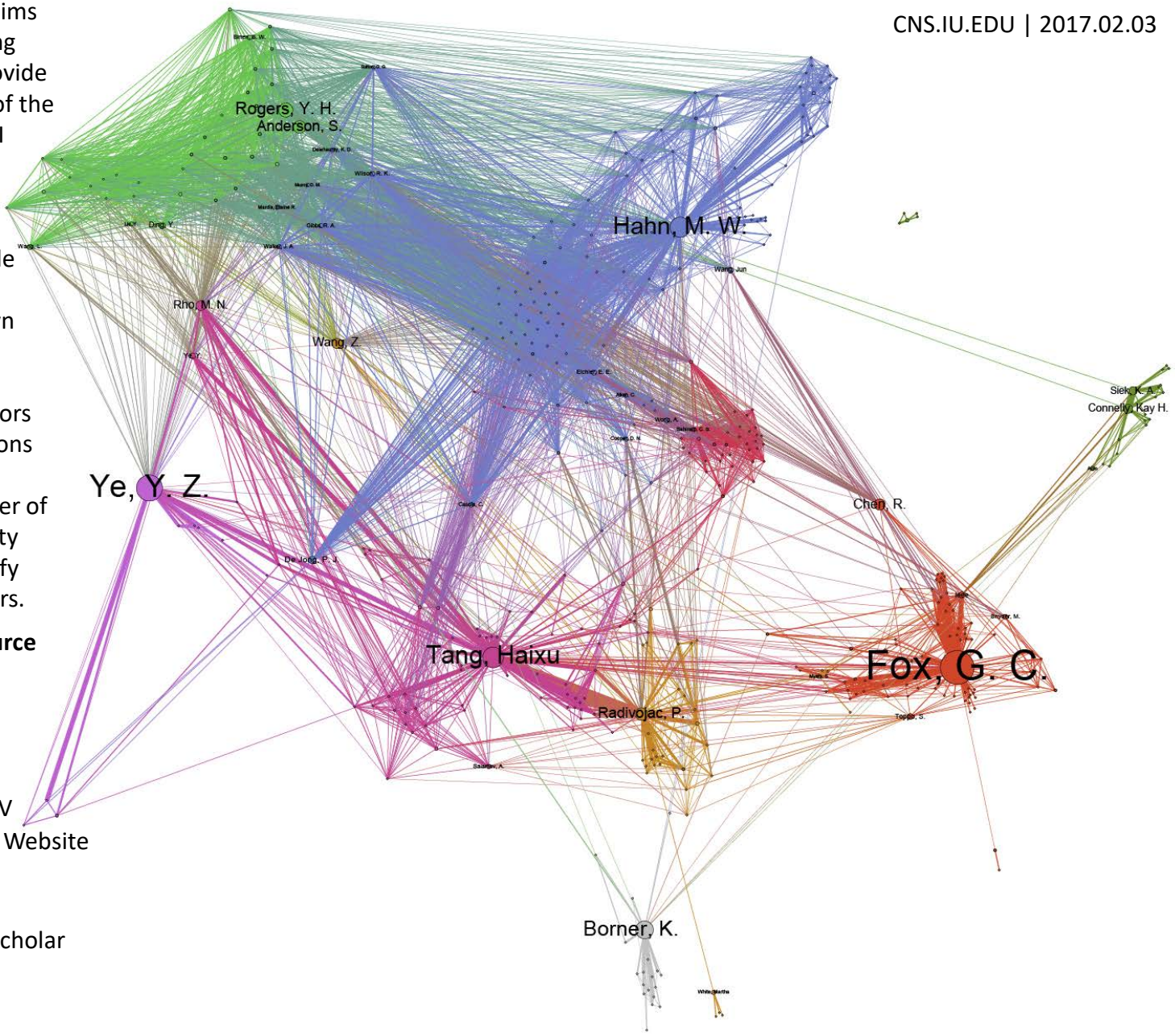
CNS.IU.EDU | 2017.02.03

PHI Impact Assessment work aims to ease the evaluation/reporting burden for PHI leads and to provide a more holistic understanding of the expertise and impact of the PHI team effort. High quality and high coverage data about project activity and outcomes is required to provide actionable insights.

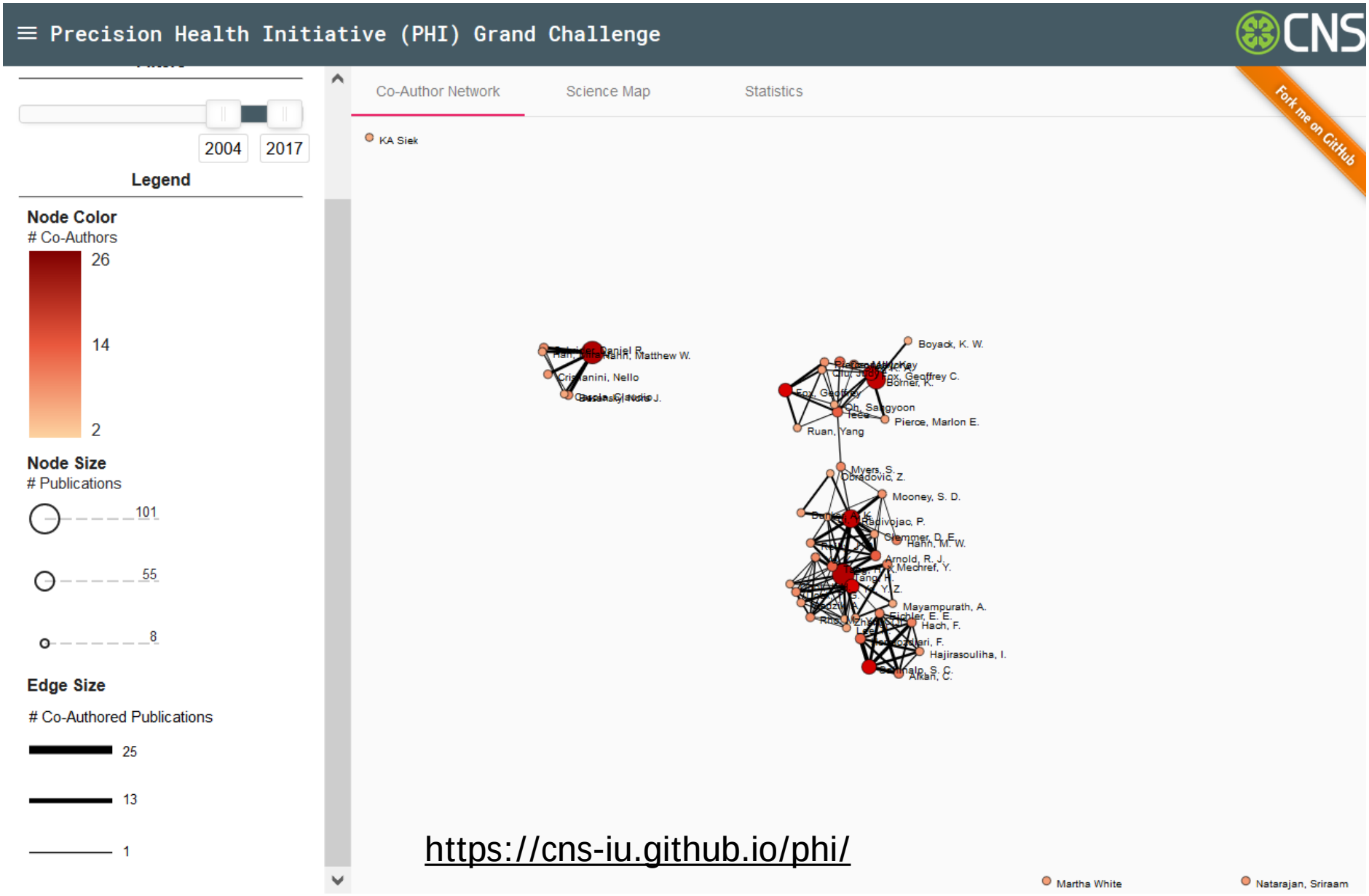
The co-author network shown here was extracted from 800+ papers gathered 10/31 to 11/28/2016. It shows 414 authors with more than three publications and 5,808 collaboration edges.

Node size denotes the number of publications. Blondel community detection was applied to identify (and color) collaboration clusters.

Name	# Pubs	Data Source
Borner	108	WoS
Connelly	32	CV
Fox	180	WoS
Hahn	123	WoS
Myers	29	WoS & CV
Natarajan	15	Personal Website
Predrag	60	WoS
Sahinalp	68	PubMed
Siek	32	Google Scholar
Tang	102	PubMed
White	15	CV
Ye	71	WoS

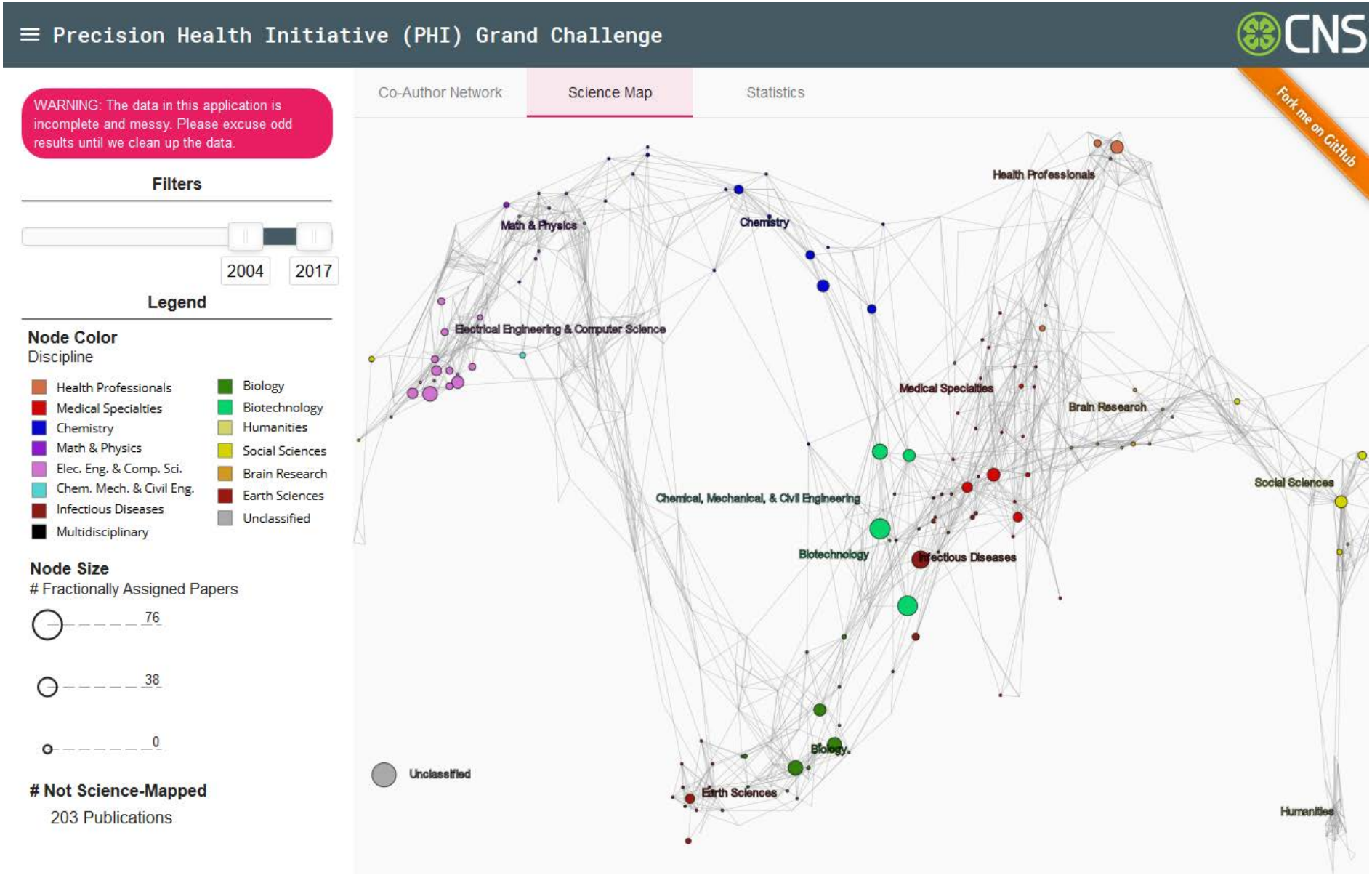


Interactive Co-Author Network in year 2 of PHI



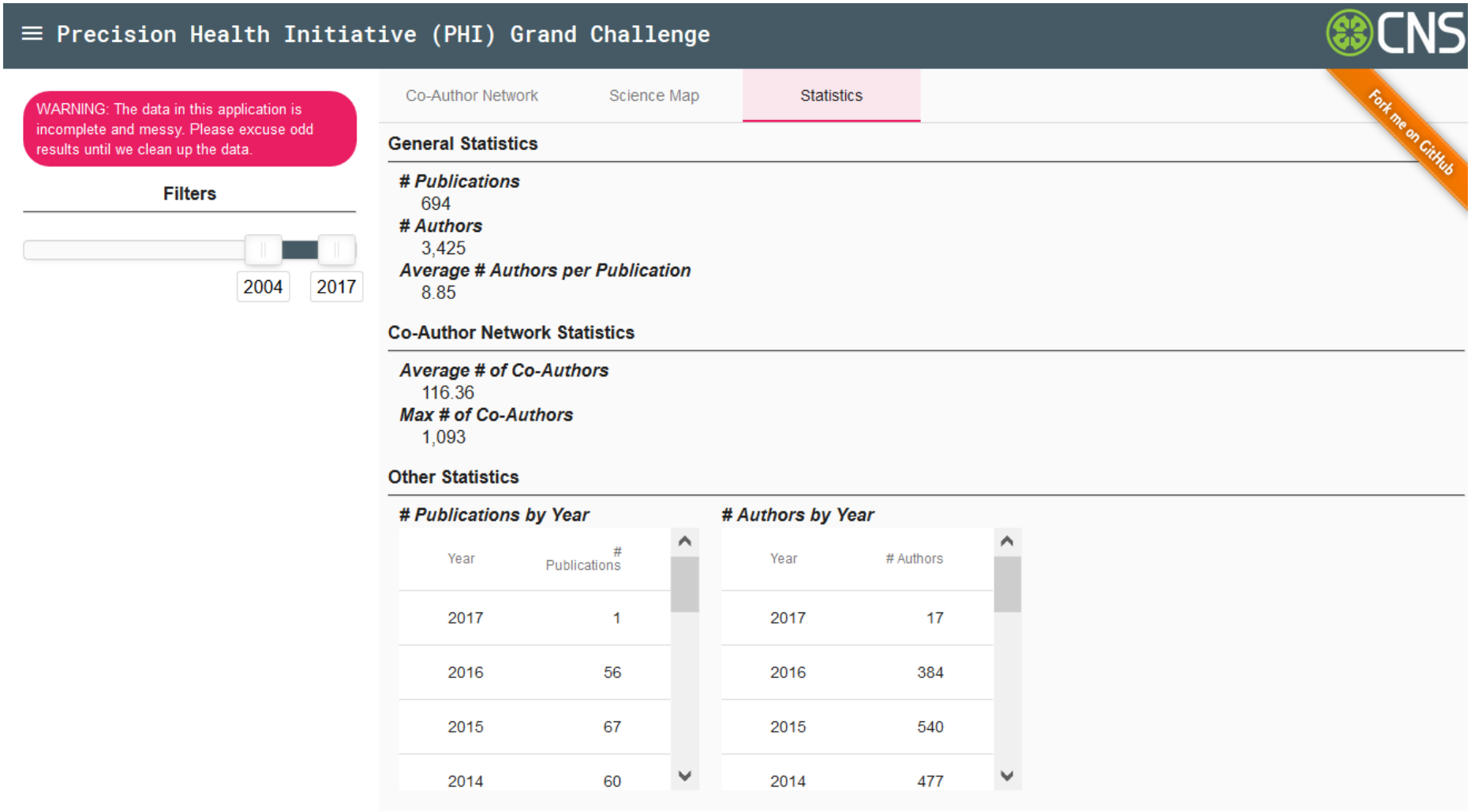
https://cns-iu.github.io/phi/

Interactive Topic (Science) Map in year 2 of PHI



https://cns-iu.github.io/phi/

Statistics of Co-Author Network in year 2 of PHI



https://cns-iu.github.io/phi/

Mapping Longitudinal Scientific Progress, Collaboration and Impact of the Alzheimer's Disease Neuroimaging Initiative (ADNI)

Xiaohui Yao[1,3,4], Jingwen Yan[1,3,4], Michael Ginda[2,3], Katy Börner[2,3], Andrew J Saykin[1,3], Li Shen[1,3,4], for the Alzheimer's Disease Neuroimaging Initiative*

[1] Center for Neuroimaging, Indiana University School of Medicine

[2] Cyberinfrastructure for Network Science Center, Indiana University Bloomington

[3] Indiana University Network Science Institute

[4] School of Informatics and Computing, Indiana University

*Data used in preparation of this article were obtained from the Alzheimer's disease Neuroimaging Initiative (ADNI) database (adni.loni.usc.edu). As such, the investigators within the ADNI contributed to the design and implementation of ADNI and/or provided data but did not participate in data analysis or writing of this report. A complete listing of ADNI investigators can be found at: http://adni.loni.usc.edu/wp-content/uploads/how_to_apply/ADNI_Acknowledgement_List.pdf

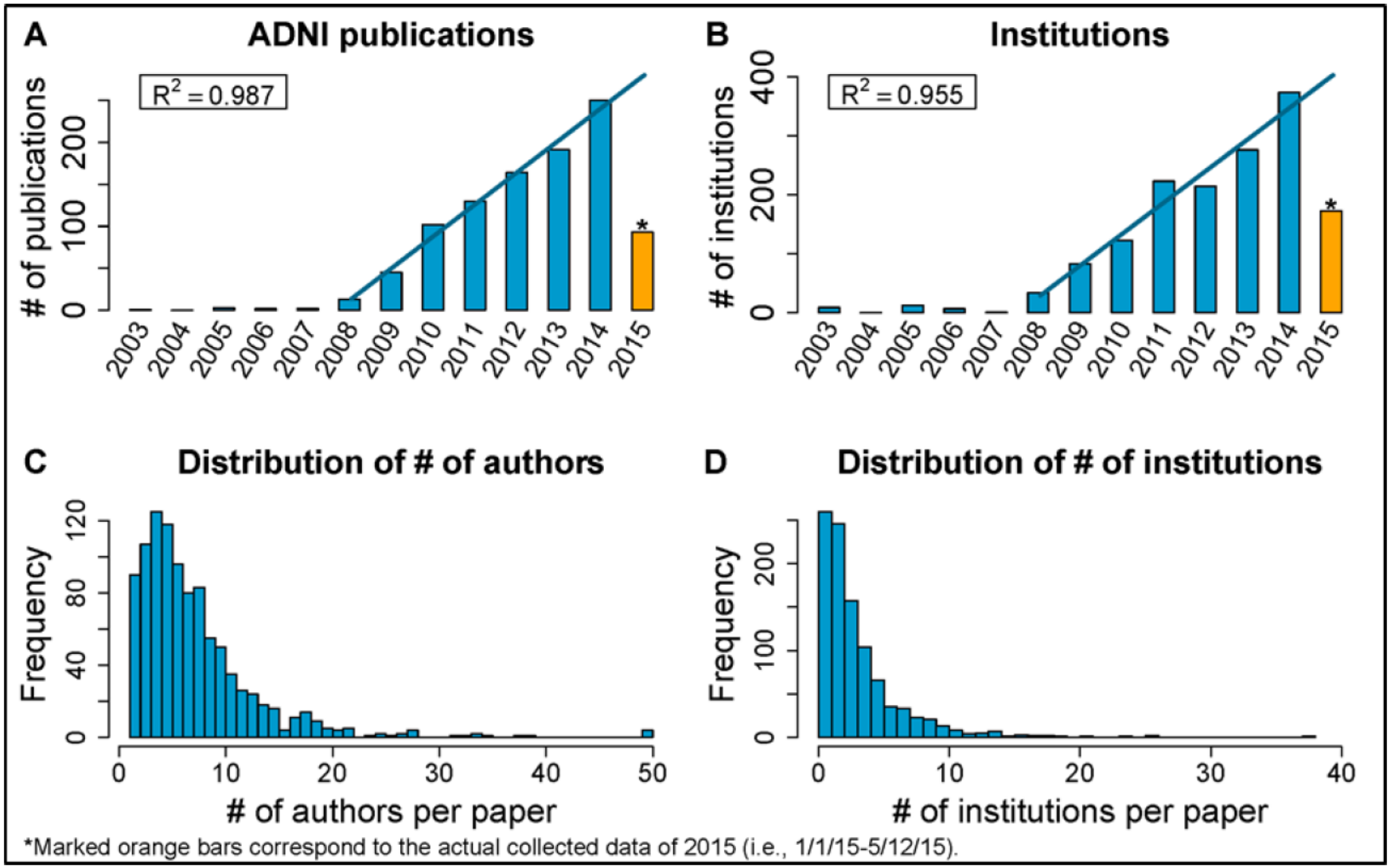


Figure 1. Statistics for ADNI publications between 01/01/2003 and 05/12/2015. (A) Growth of ADNI publications on the year-by-year basis; line indicates a linear regression prediction for the 2015 number using data from 2008 to 2014. (B) Growth of institutions involved in ADNI publications; line indicates a linear regression prediction for the 2015 number using data from 2008 to 2014. (C) Distribution of number of authors per paper. (D) Distribution of number of institutions per paper.

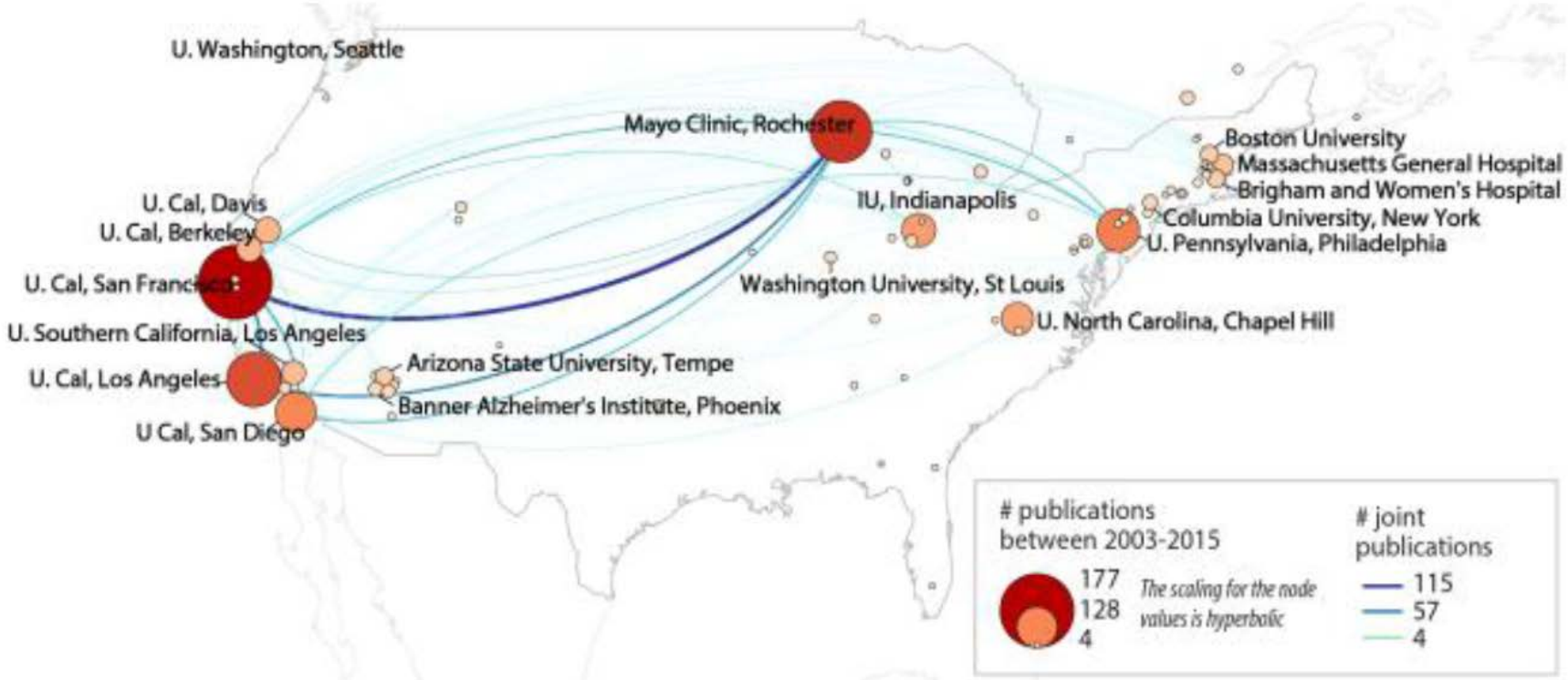
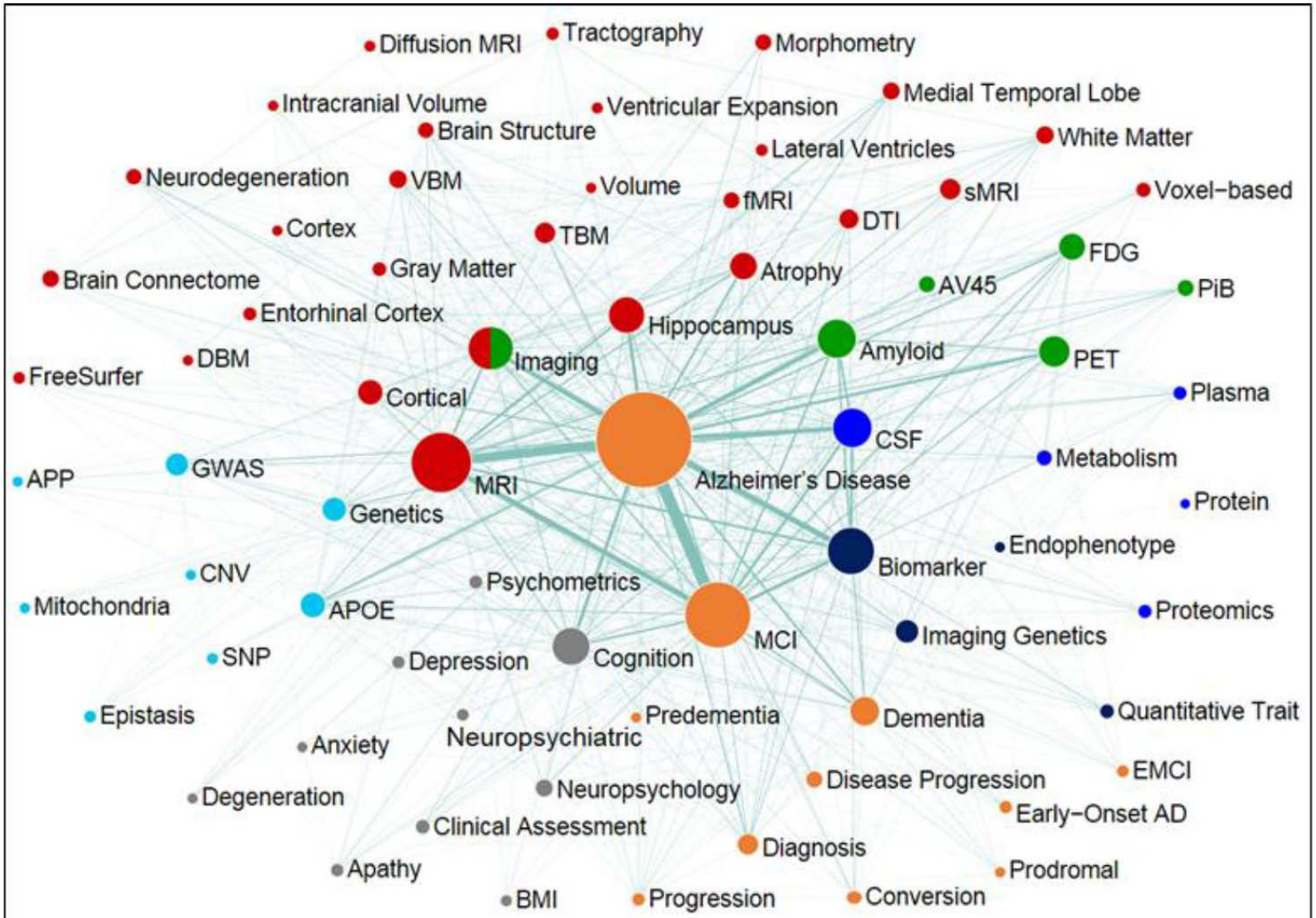


Figure 3: Co-affiliation network overlaid on a geospatial map shows collaborating organizations affiliated with ADNI in North American based on co-authored publications. Only organizations with at least 4 publications are shown; organizations with at least 30 publications or that are a Core ADNI research institution have been labeled in the network. Organization relationships (edges) with four or more co-authorships are shown.



Supplemental Figure 5: Keyword co-occurrence network focused on major ADNI themes. Nodes represent keywords relevant to major ADNI themes, including MRI, PET, other biological biomarkers, clinical and neuropsychological assessment, genetics, and disease and progression. Edges denote the joint appearance of keywords in a publication. Nodes are colored based on the themes they belonged to, and those across three or more themes are colored in dark blue. Both nodes and edges were scaled proportionally based on Bezier curve. Only nodes with degree > 2 are shown.

Strategic Foresight:
Comparing IU Efforts to Related R&D Efforts

The Kavli HUMAN project (http://kavlihumanproject.org) is similar to the PHI effort in terms of data collection and analysis. It measures everything from microbiome to social distances to medical records for 10,000 New Yorkers over the next 30 years, providing the first statistically balanced, contextually rich picture of human health and development.

PHI publicly promised to address one cancer, one childhood disease, one chronic disease, and one neurodegenerative disease. The candidates as of now are

- Cancer: AML, ALL, or multiple myeloma
- Childhood: Fanconi anemia, which is a rare disease.
- Chronic: diabetes or osteoarthritis, probably not obesity
- Neurodegenerative: Alzheimer's

Many teams in the US and world-wide are working on cures for these diseases.

Compare IU Efforts to Related R&D Efforts cont.

It seems highly desirable to monitor progress by other teams using Medline publication, NIH funding, WIPO patents, and US clinical trials data, in order to semi-automatically

- identify relevant publications, funded grants, clinical trials, or patents,
- key experts (e.g., to hire the best faculty for the PHI team effort and/or invite them to come visit IU and give talks), and
- to collaborate with the best teams.

Results will be published as internal white papers and in scholarly journals.

Related CNS Work and Synergies

“Expertise Visualization” for NIH’s Clinical and Translational Science Awards (CTSA) Program Hubs. The online service lists key experts, publications, funding awards and clinical trials that match user-specified search queries. *Collaborative work with Intelligent Automation, Inc. work; NIH SBIR Phase II project entitled “SMS-VAT: A Scalable Multi-Scale Visual Analytical Tool.”*

“Engineering Observatory” that facilitates near real-time monitoring of Engineering Research Centers (ERCs) in support of informed decision making. Relevant data streams comprising course data, publications, patents, 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. User evaluations will be performed to validate and optimize the new functionality. *This collaborative work with the nanoHub team at Purdue University is funded by NSF.*

“XDMoD Value Analytics” aims to improve our understanding of the interplay between compute resource availability, resource consumption, and scientific outputs. The overall goal of this line of research is to provide data-driven, objective input to regular evaluation exercises but also to support near-real time proactive management and resource allocation decision making related to optimizing the usage of advanced computing infrastructure. *This collaborative work with UITS@IUB is funded by an NSF EAGER Award.*

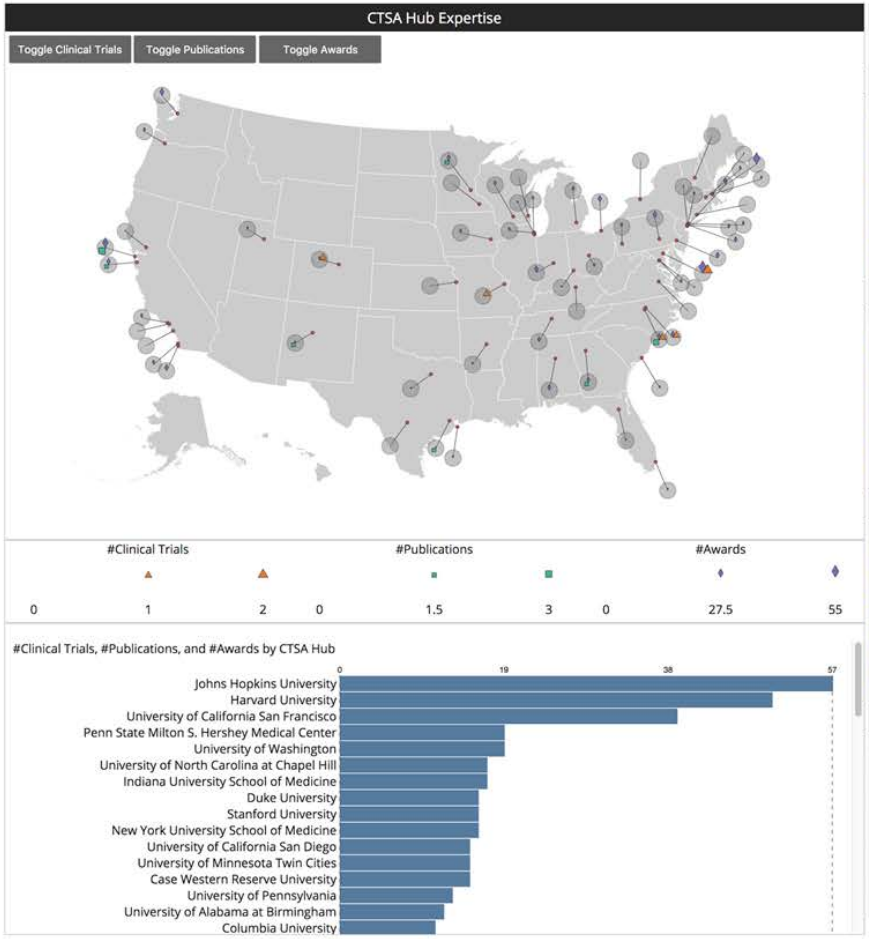
“Visualizing Healthcare System Dynamics in Big Biomedical Science.” NIH U01CA198934 (Griffin Weber, Harvard University, Katy Börner) June 15 - May 18.

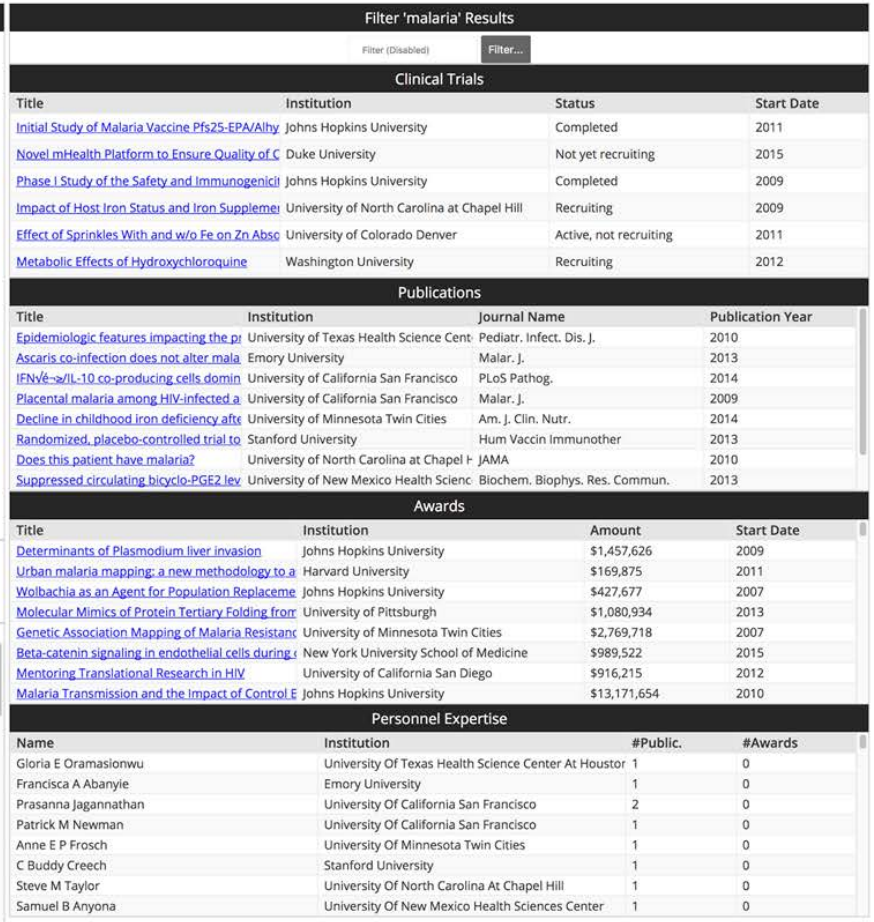
“Expertise Visualization” for NIH’s Clinical and Translational Science Awards (CTSA) Program Hubs. The online service lists key experts, publications, funding awards and clinical trials that match user-specified search queries. *Collaborative work with Intelligent Automation, Inc. work; NIH SBIR Phase II project entitled “SMS-VAT: A Scalable Multi-Scale Visual Analytical Tool.”* Sept. 14 - Aug. 16.

Visualization: IAI Expertise Visualization

Project: IAI

demo.cns.iu.edu/client/iai/expertise.html?set=malaria



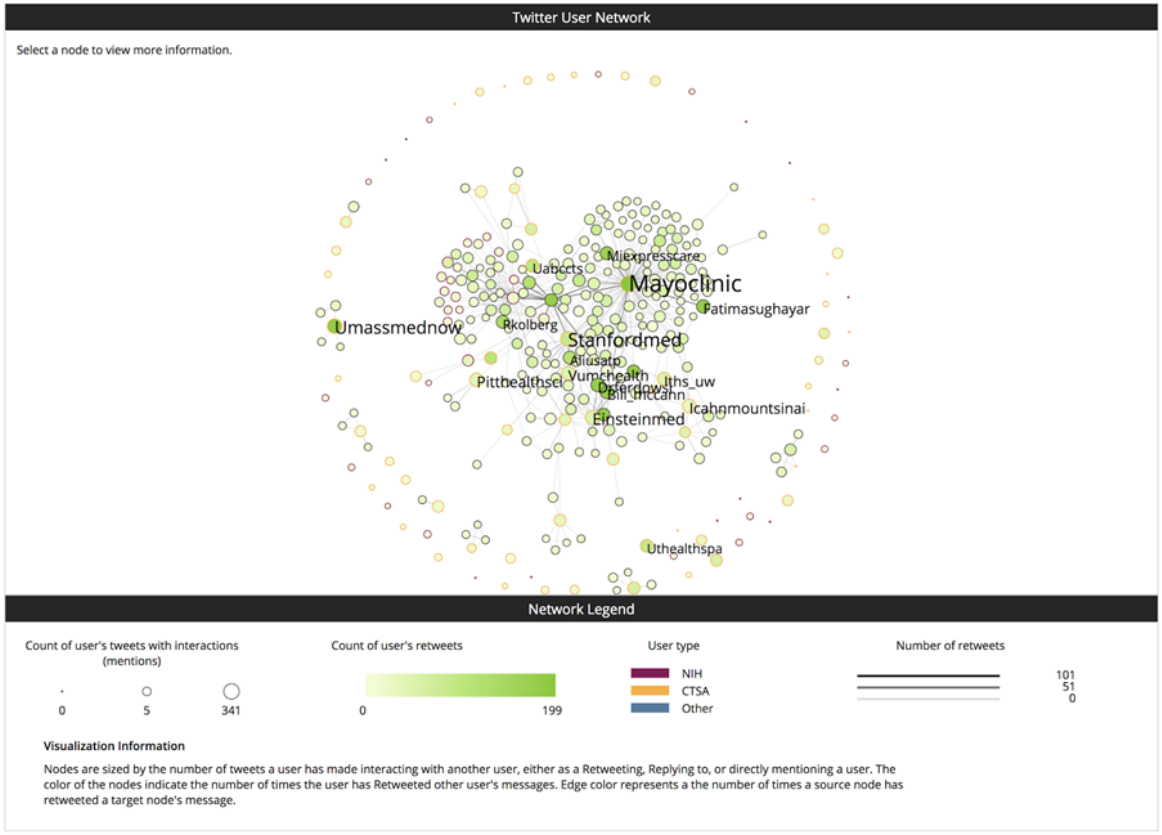


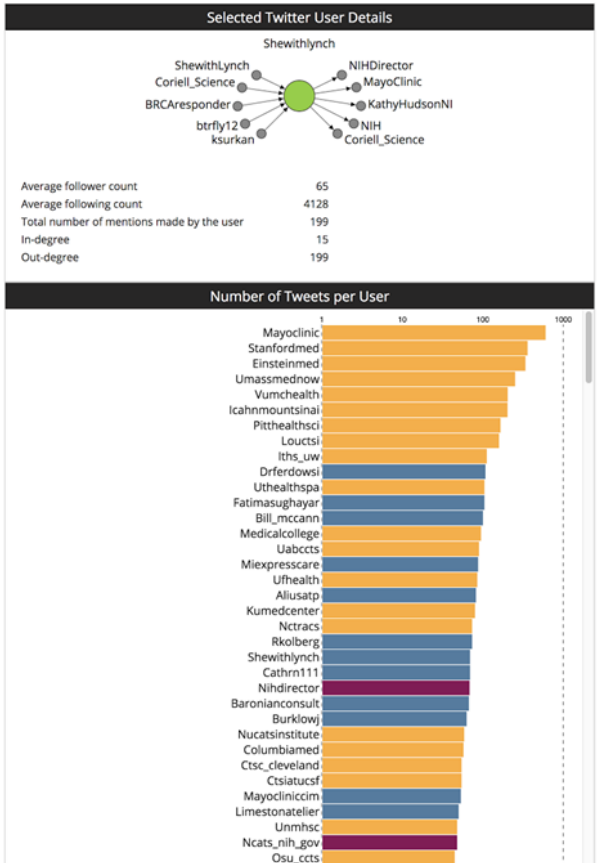
This visualization is based on publication datasets retrieved from the Scholarly Database at IU, and is used to identify relevant experts, publications, clinical trials, and awards that match a search term.

Visualization: Twitter Network

Project: IAI

demo.cns.iu.edu/client/iai/twitter.html





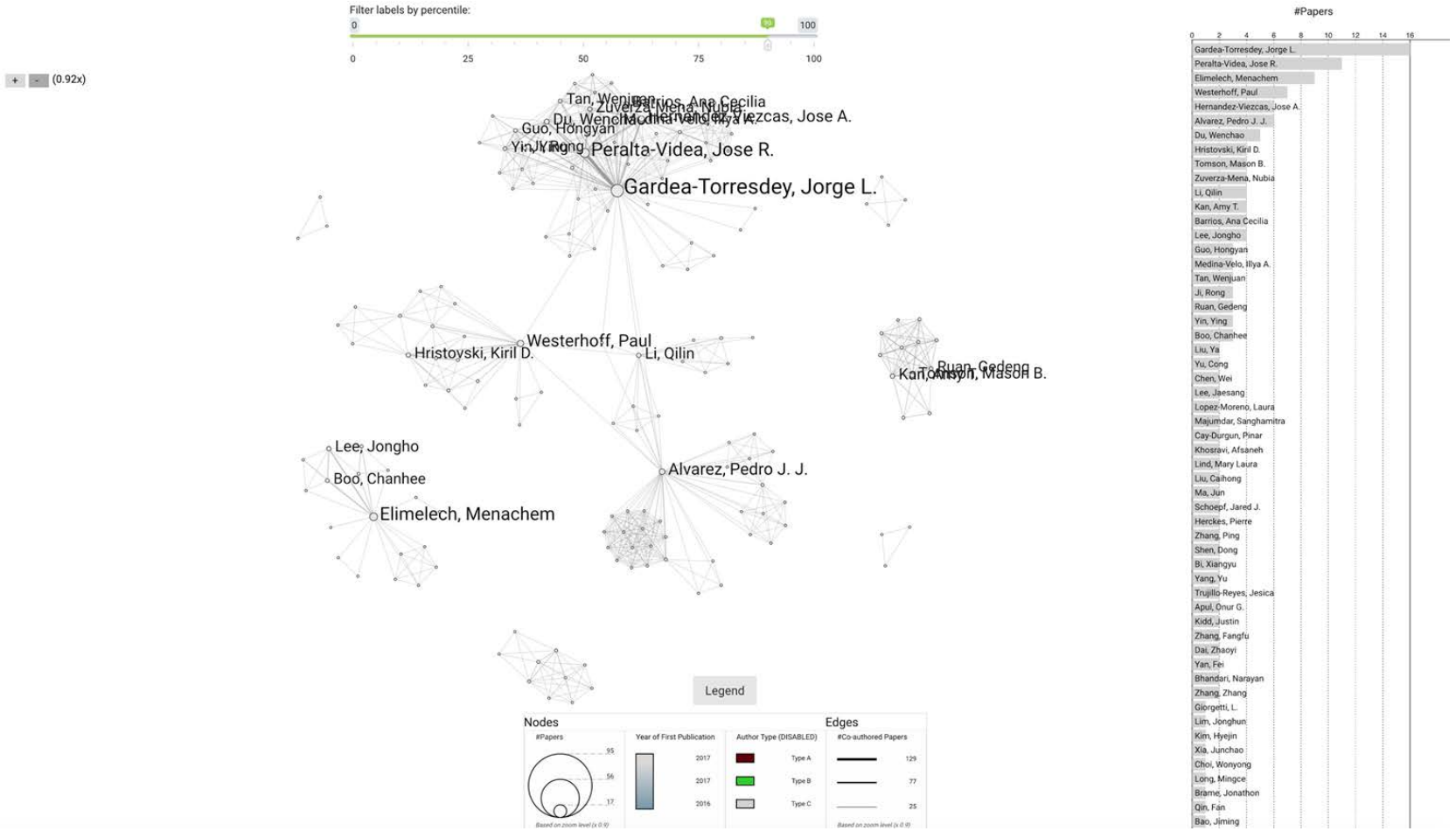
This visualization shows CTSA hub and NIH activities on Twitter. The accounts and tweets associated with CTSA and NIH were collected between August 2015 - Sept. 2015, then processed and analyzed to create a social network based on the interaction behaviors of users. The layout of the user network is force-directed, meaning that nodes that are close to each other have a stronger connection to one another indicating which Twitter users interact with each other most often.

“Engineering Observatory” that facilitates near real-time monitoring of Engineering Research Centers (ERCs) in support of informed decision making. Relevant data streams comprising course data, publications, patents, 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. User evaluations will be performed to validate and optimize the new functionality. *This collaborative work with the nanoHub team at Purdue University is funded by NSF.* Dec 15 – Nov 17.

Visualization: Co-Authorship Network

Project: ERC

nanohub.org/citations/curate

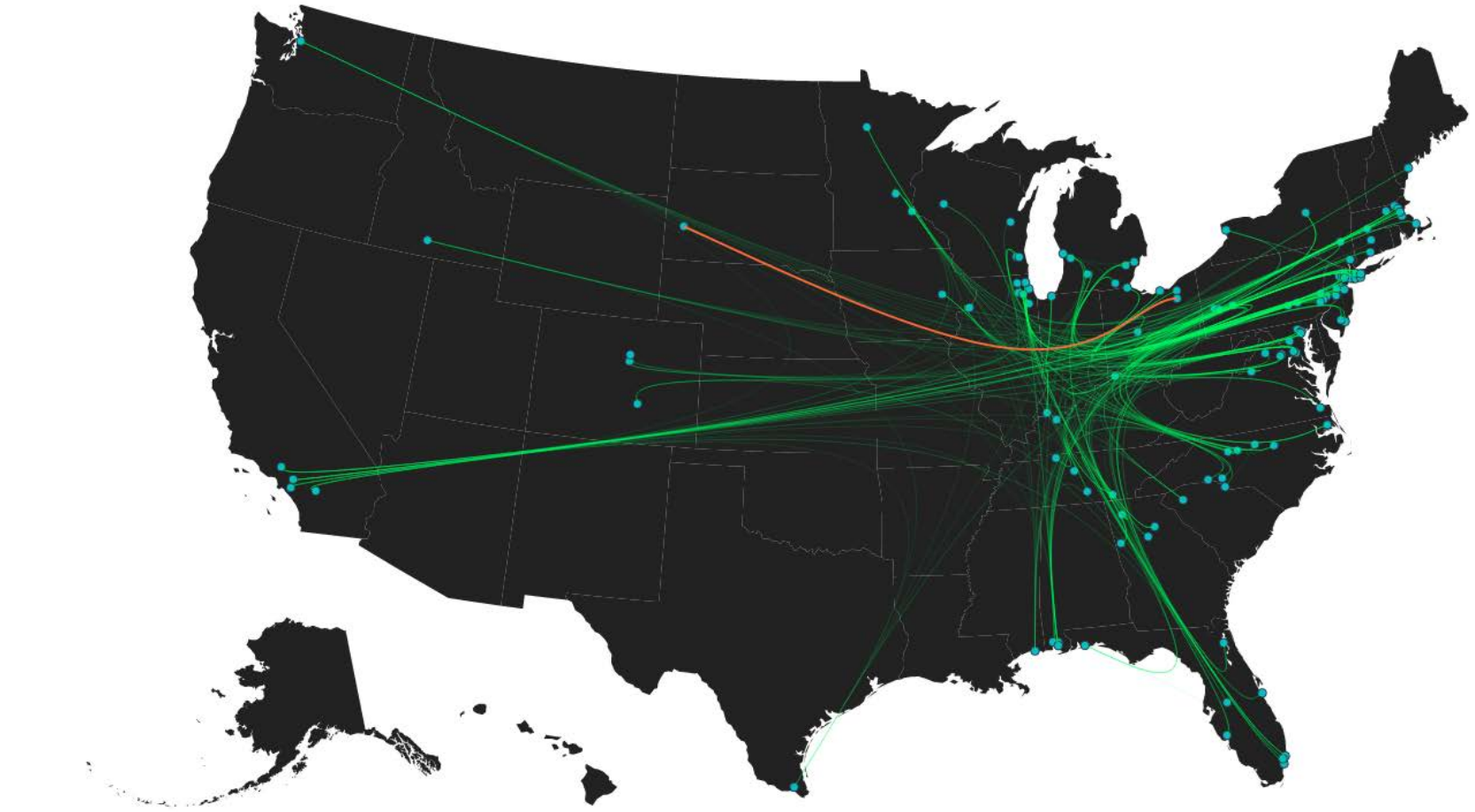


This figure shows the collaboration network of one ERC, based on co-authorship extracted from bibliography files. Each node is an author, and 2 authors are connected if they have authored a publication together.

Visualization: Geographic co-authorship visualization

Project: ERC

nanohub.org/citations/curate

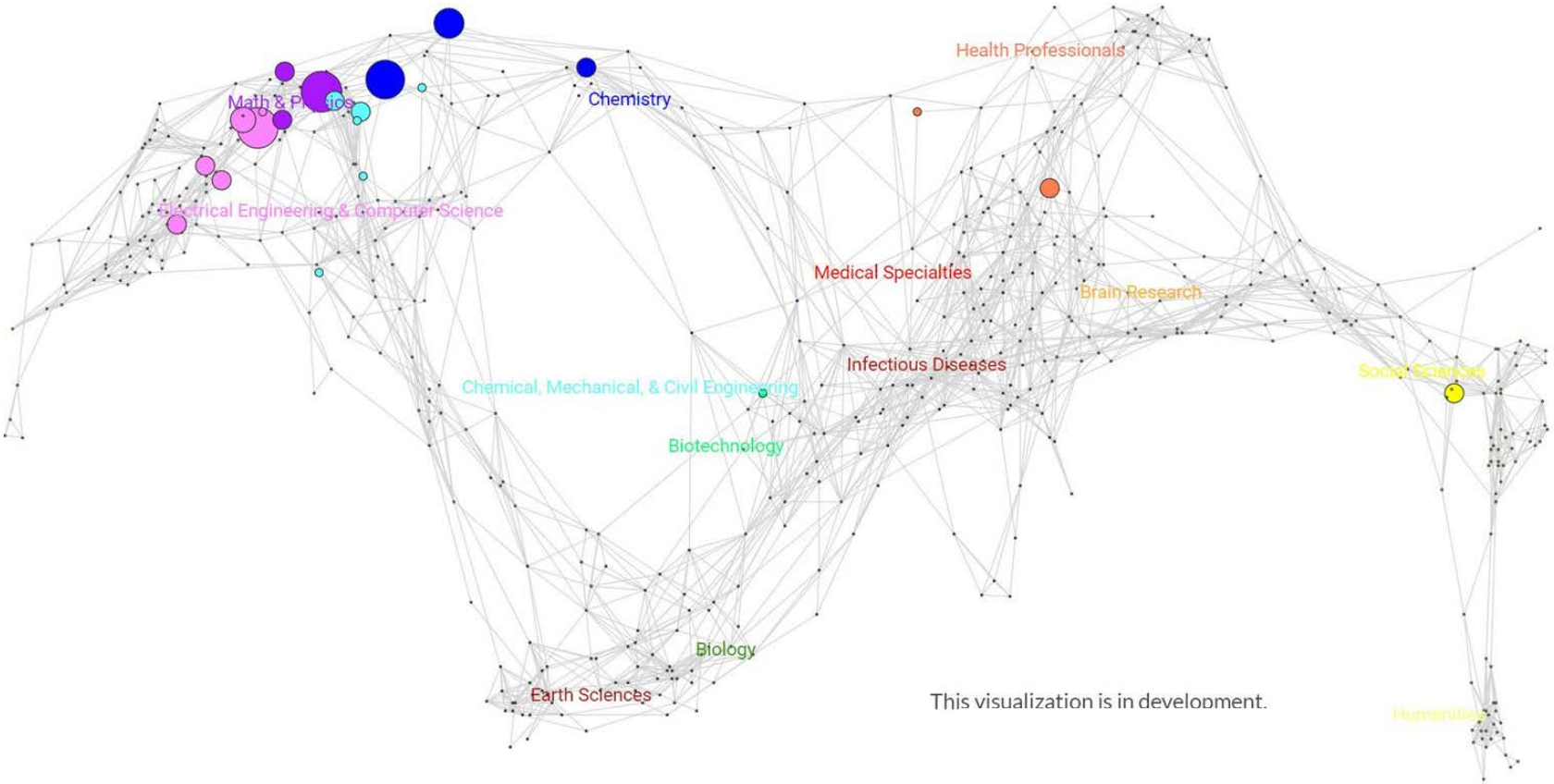


This map shows the co-authorship network overlaid on a geospatial map of the US. each node represents an author and two authors are connected if they have authored a paper together.

Visualization: UCSD Map of Science

Project: ERC

nanohub.org/citations/curate



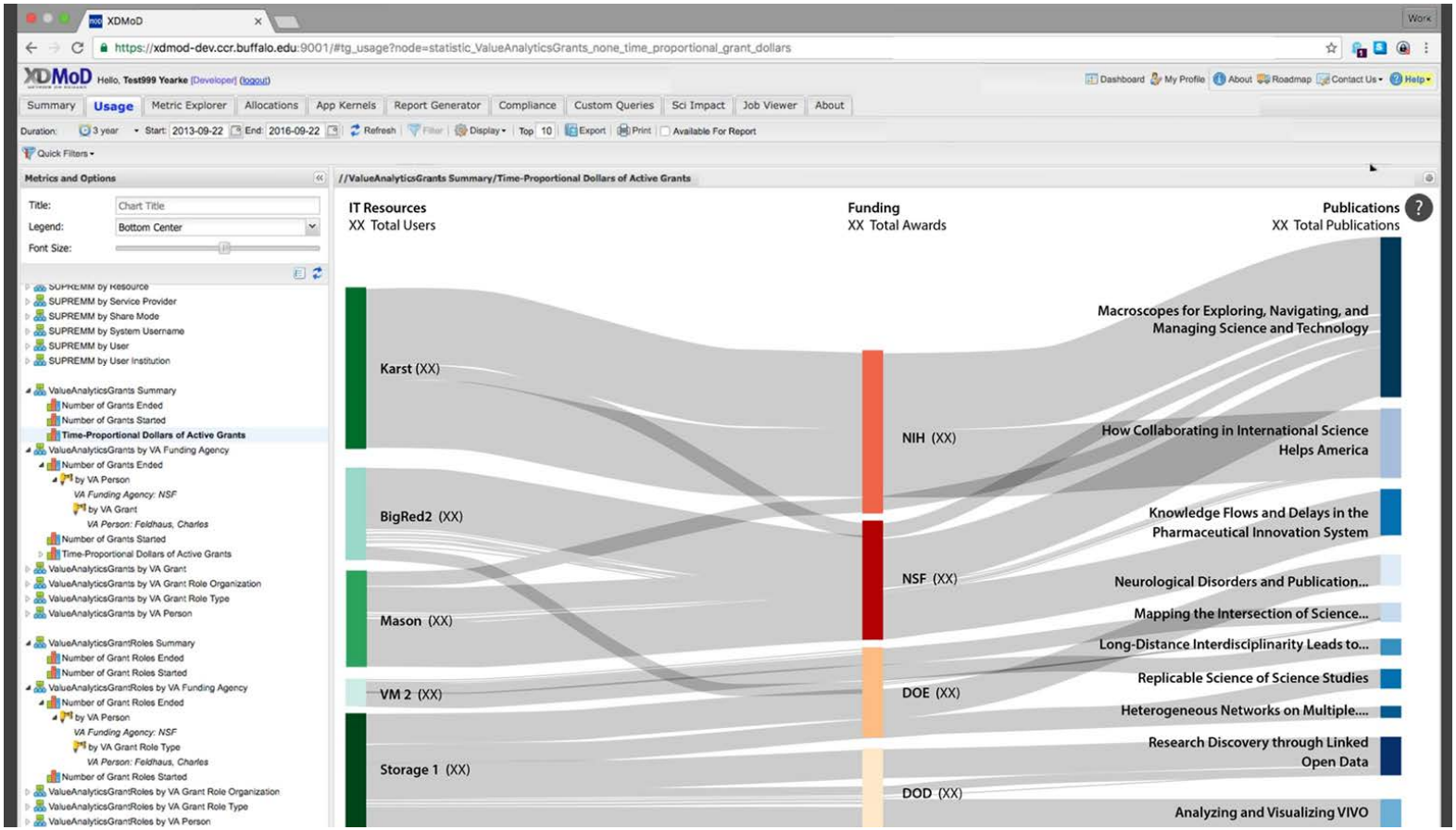
This organizes and visually represents 554 sub disciplines of science and their relationships to one another. Sub disciplines are grouped into 13 overarching disciplines that are color coded (red for medicine, green for biology, etc.) and labelled. Using a journal name based or keyword based mapping process, data overlays can be computed. For example, expertise profiles for an individual or an institution are generated by reading a bibtex or EndNote file with relevant publications, identifying unique journal names, and overlaying geometric symbols such as circles atop the sub discipline(s) that are associated with each journal. This Map of science can be used to explore, understand, and communicate the expertise profiles of an institute or nation.

“XDMoD Value Analytics” aims to improve our understanding of the interplay between compute resource availability, resource consumption, and scientific outputs. The overall goal of this line of research is to provide data-driven, objective input to regular evaluation exercises but also to support near-real time proactive management and resource allocation decision making related to optimizing the usage of advanced computing infrastructure. *This collaborative work with UITS@IUB is funded by an NSF EAGER Award.* April 16 - March 18.

Scrivner, Olga, Gagandeep Singh, Sara Bouchard, Scott Hutcheson, Ben Fulton, Matt Link, and Katy Börner. 2018. "XD Metrics on Demand Value analytics: Visualizing the impact of internal information Technology investments on external Funding, Publications, and collaboration networks". *Frontiers Research Metrics and Analytics* 2 (10).

Visualization: Sankey Diagram

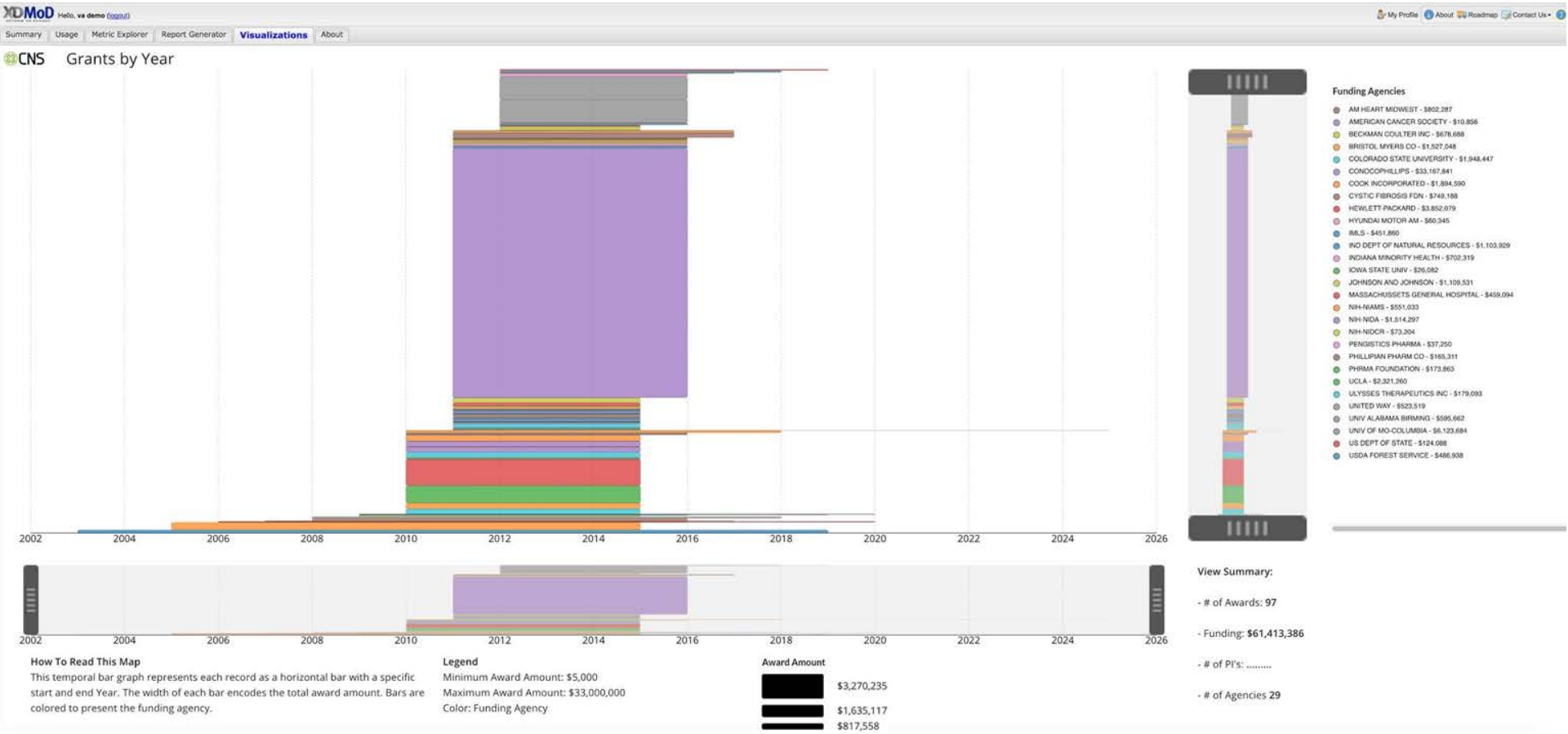
Project: XDMoD



This Sankey diagram displays a multivariate analysis of the relationship between IT resources, funding agencies, and publications. The width of each line represents grant dollars awarded to researchers. The configuration model allows for easy metric switching.

Visualization: Temporal Bar Graph

Project: XDMoD

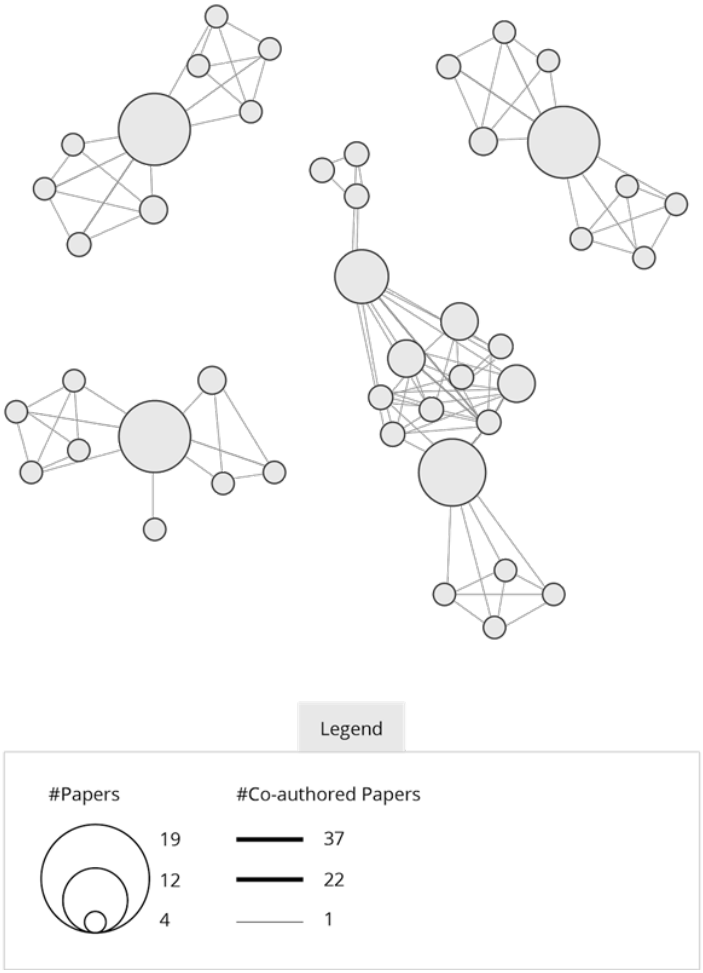


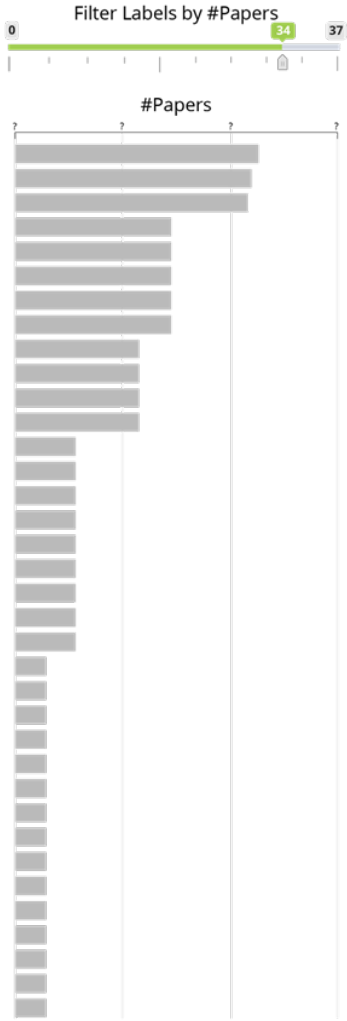
This temporal bar graph represents each record as a horizontal bar with a specific start and end year. The width of each bar encodes the total award amount. Bars are colored to represent the funding agency (legend of funding agencies are listed on the right). This graph shows funding duration, amounts, and types over time.

Visualization: Co-PI

Project: XDMoD

This network represents collaboration patterns based on NSF funding data. Each node represents a principal investigator (PI or Co-PI). It is labelled by the PI's name, size coded by the total value of all awards for each PI, and color coded by the number of collaborators. An edge between two PIs denotes that they have a grant together with edge thickness denoting the number of times they collaborated and edge color reflecting their success in terms of total sum of all their joint awards. The listing on right rank orders PIs by total dollar amounts.



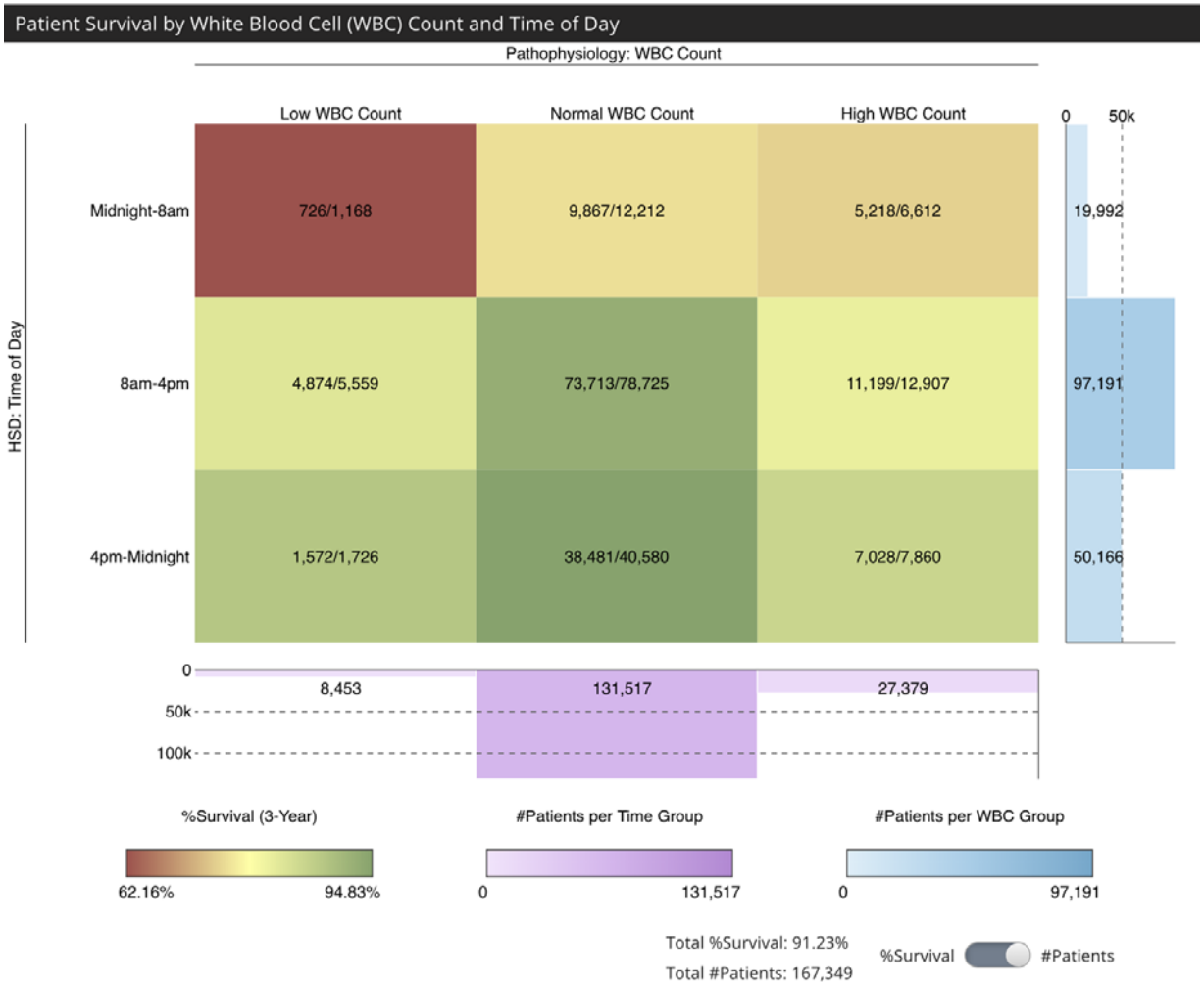


“Visualizing Healthcare System Dynamics in Big Biomedical Science.” NIH U01CA198934 (Griffin Weber, Harvard University, Katy Börner) June 15 - May 18.

Visualization: Heatmap

Project: HSD

demo.cns.iu.edu/client/hsd/static/heatmap_group.html



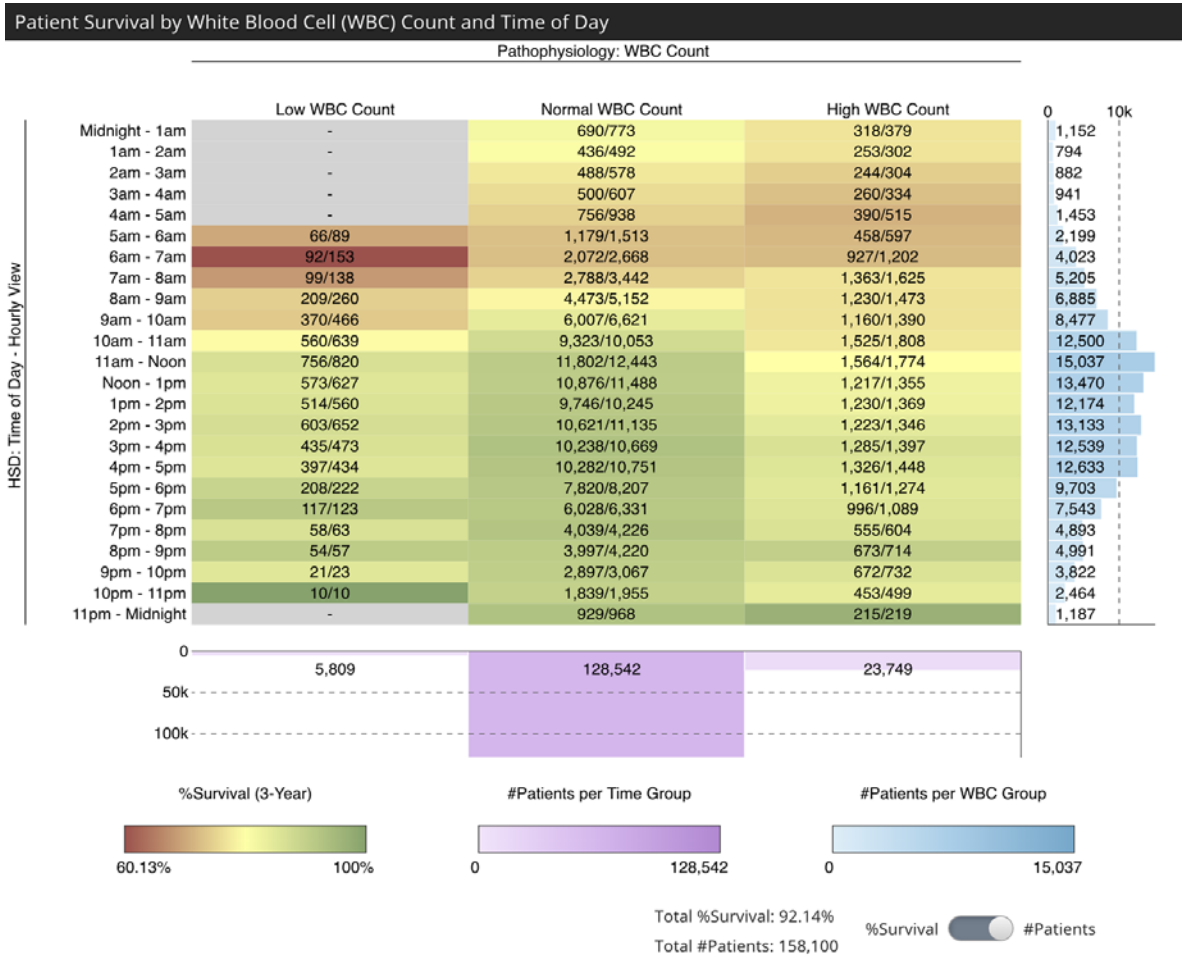
This visualization shows how white blood cell (WBC) laboratory tests correlate with three-year survival rates. The HSD dimension of the data (rows) is the time of the day of the test; and three-year survival rate (numbers and colors in the boxes) is an outcome variable. **Aggregation level for the HSD time of day are shown—three 8 hour blocks**. The lowest survival rates are for patients with a low WBC value in the morning (specifically at 6am).

In this project, we created data visualizations to explain HSD to users and to help them incorporate it into in their research.

Visualization: Heatmap

Project: HSD

demo.cns.iu.edu/client/hsd/static/heatmap_hour.html



This visualization shows how white blood cell (WBC) laboratory tests correlate with three-year survival rates. The HSD dimension of the data (rows) is the time of the day of the test; and three-year survival rate (numbers and colors in the boxes) is an outcome variable. **Aggregation level for the HSD time of day are shown— 24 hourly blocks on the right.**. The lowest survival rates are for patients with a low WBC value in the morning (specifically at 6am).

In this project, we created data visualizations to explain HSD to users and to help them incorporate it into in their research.

All papers, maps, tools, talks, press are linked from http://cns.iu.edu
These slides will soon be at http://cns.iu.edu/presentations.html

CNS Facebook: http://www.facebook.com/cnscenter
Mapping Science Exhibit Facebook: http://www.facebook.com/mappingscience