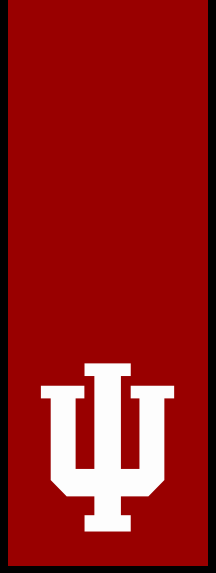


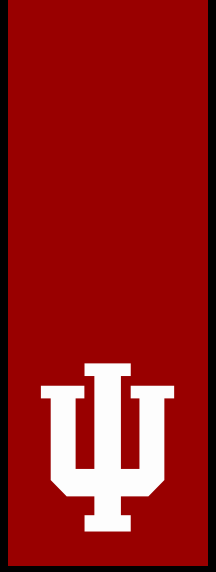
CREST Advisory Board review | February 27, 2018

IVMOOC on Jetstream, Open XD Metrics on Demand Value Analytics & Maps of Science

Katy Börner, School of Informatics, Computing, and Engineering

CNS, IUNI, UITS

INDIANA UNIVERSITY BLOOMINGTON



IVMOOC on Jetstream

CNS, IUNI, UITS

INDIANA UNIVERSITY BLOOMINGTON

Jetstream
A national research and education cloud

Jeremy Fischer – jeremy@iu.edu

Senior Technical Advisor, Collaboration and Engagement Support
UITS Research Technologies

What is Jetstream and why does it exist?

- NSF's first production cloud facility
- Part of the NSF eXtreme Digital (XD) program
- Provides on-demand *interactive* computing and analysis
- Enables *configurable* environments and *programmable cyberinfrastructure*
- User-friendly, widely accessible cloud environment
- User-selectable library of preconfigured virtual machines

Jetstream
http://jetstream-cloud.org/

funded by the National Science Foundation
Award #ACI-1445604

Jetstream - Expanding NSF XD's reach and impact

Around 350,000 researchers, educators, & learners received NSF support in 2015

- Less than 2% completed a computation, data analysis, or visualization task on XD/XSEDE program resources
- Less than 4% had an XSEDE Portal account
- 70% of researchers surveyed* claimed to be resource constrained

Why are the people not using XD/XSEDE systems not using them?

- Perceived ease of access and use
- HPC resources – the traditional view of what XSEDE offers - are often not well-matched to their needs
- They just don't need *that much* capability

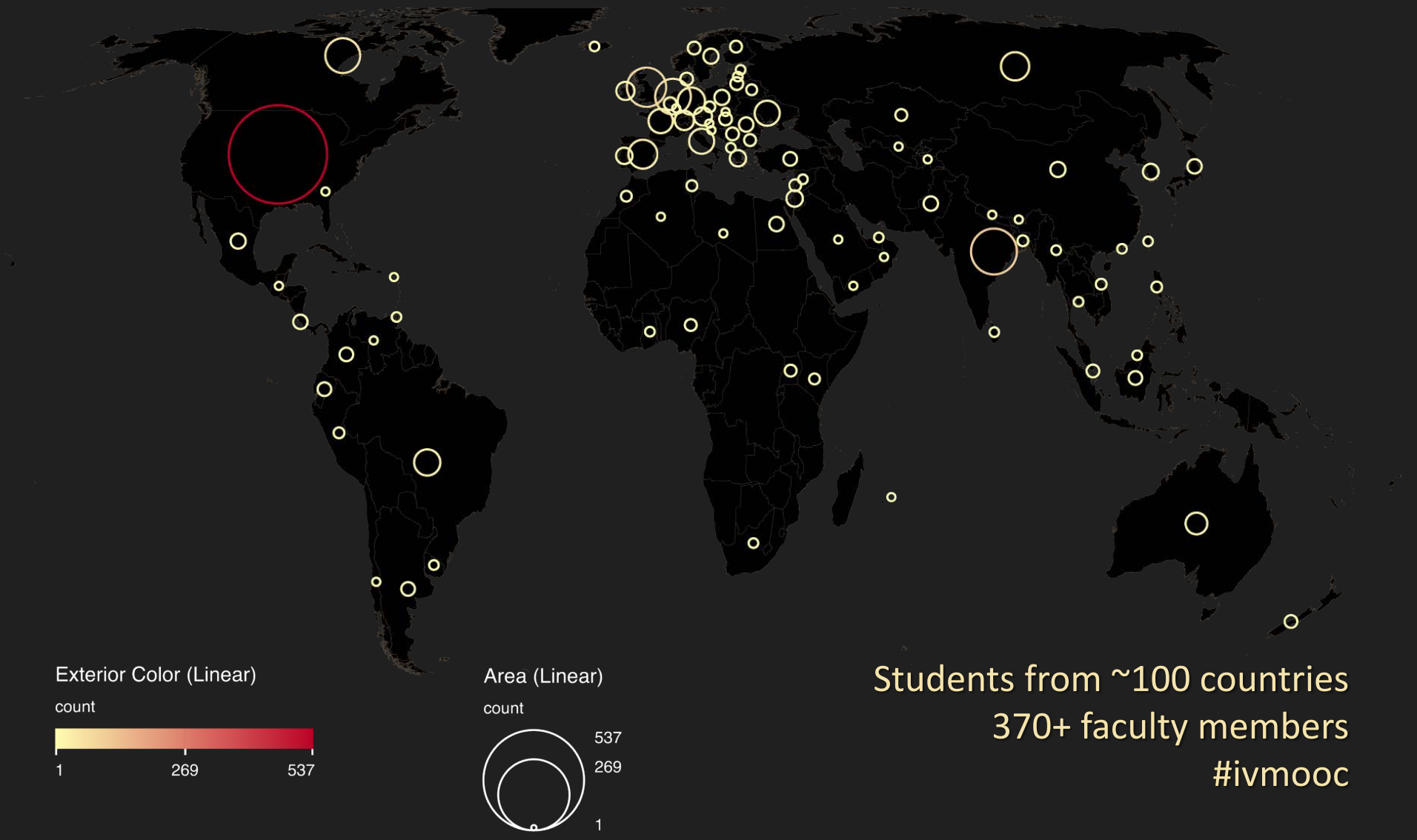
* XSEDE Cloud Survey Report - http://hdl.handle.net/2142/45766

Jetstream
http://jetstream-cloud.org/

funded by the National Science Foundation
Award #ACI-1445604

Register for free: http://ivmooc.cns.iu.edu.

The Information Visualization MOOC
ivmooc.cns.iu.edu



Students from ~100 countries
370+ faculty members
#ivmooc

Course Schedule

Part 1: Theory and Hands-On

- **Session 1** – Workflow Design and Visualization Framework
- **Session 2** – "When:" Temporal Data
- **Session 3** – "Where:" Geospatial Data
- **Session 4** – "What:" Topical Data

Mid-Term

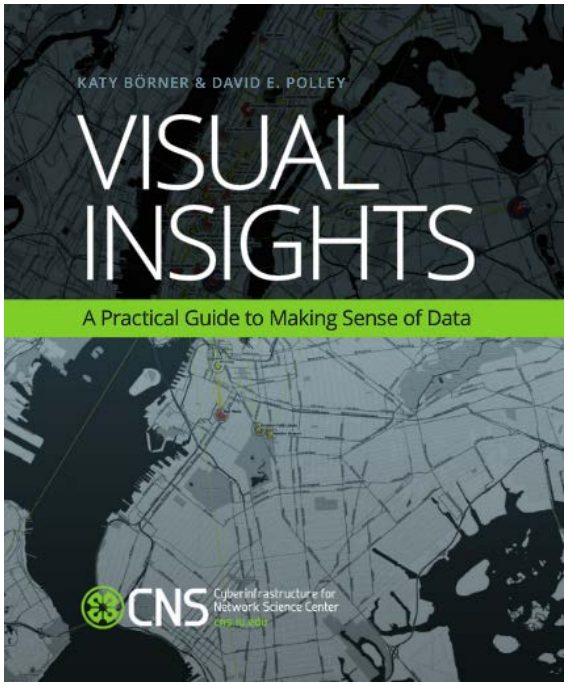
- **Session 5** – "With Whom:" Trees
- **Session 6** – "With Whom:" Networks
- **Session 7** – Dynamic Visualizations and Deployment

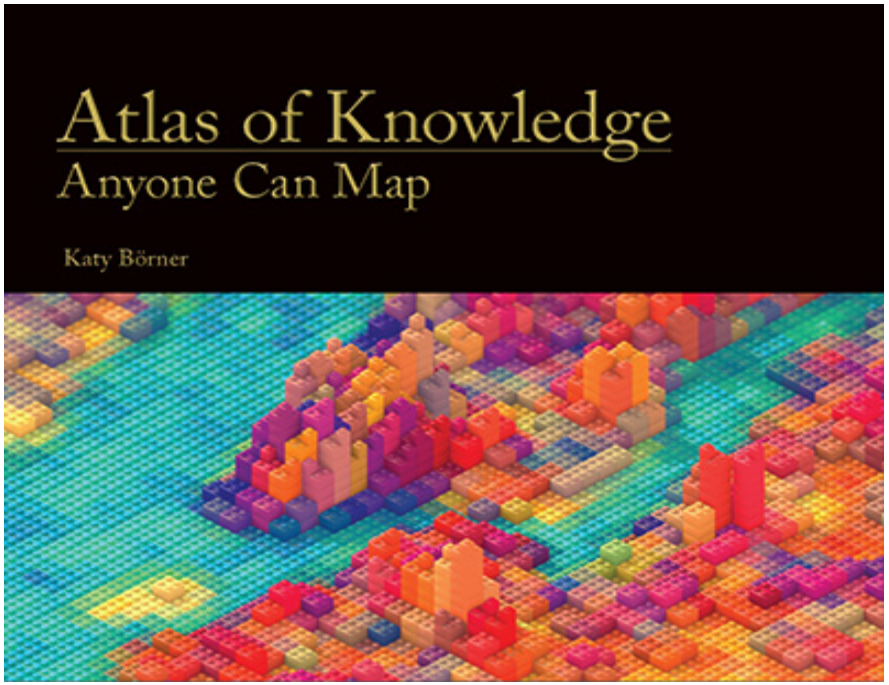
Final Exam

Part 2: Students work in teams on client projects.

Final grade is based on Class Participation (10%), Midterm (30%), Final Exam (30%), and Client Project(30%).

Books Used in the IVMOOC





Teaches timely knowledge:

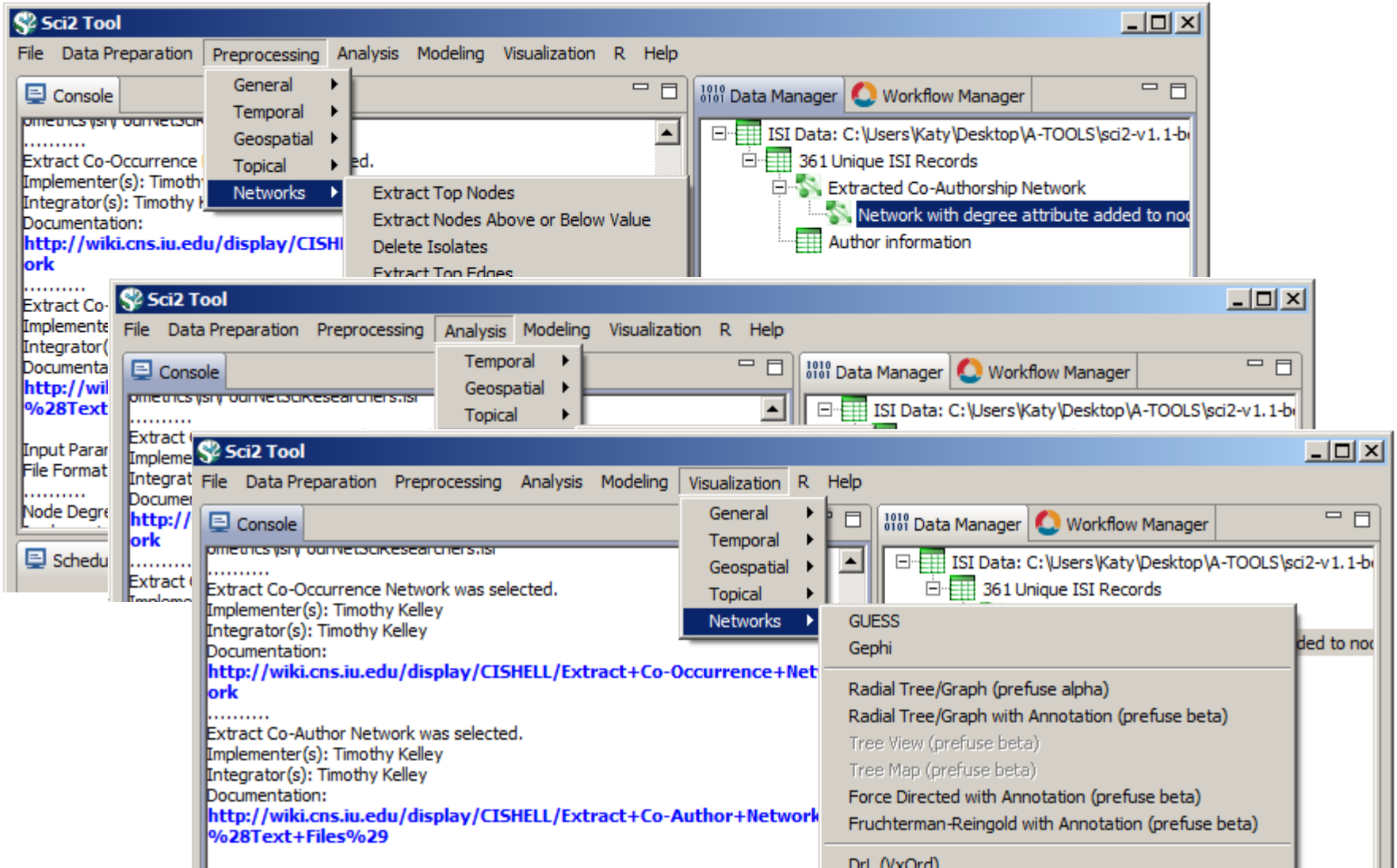
Advanced algorithms, tools, and hands-on workflows.

Teaches timeless knowledge:

Visualization framework—exemplified using generic visualization examples and pioneering visualizations.

Sci2 Tool Interface Components

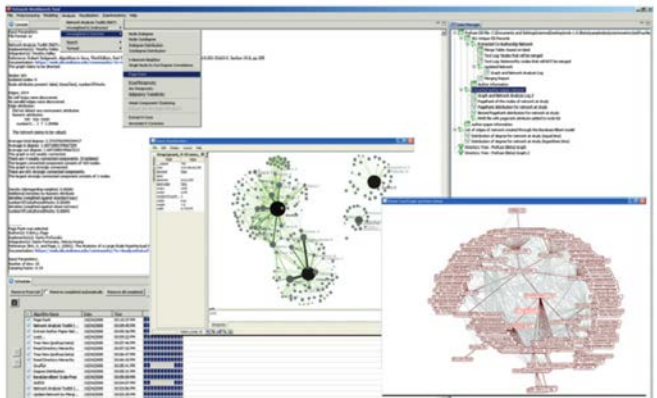
Download tool for free at http://sci2.cns.iu.edu



IVMOOC 2017: Using NSF XSEDE and Jetstream to Run Data Analysis and Visualization Workflows in the Cloud

In Spring 2017, IVMOOC students are invited to beta-test a novel cloud computing setup that supports low- to high-bandwidth users in running more compute intensive data analysis and visualization workflows in the cloud. Specifically, students will create a user account for the Extreme Science and Engineering Discovery Environment (XSEDE) and then use the Jetstream cloud-computing environment to run specific workflows using the Network Workbench Tool.

Summary
Network Workbench: A Large-Scale Network Analysis, Modeling and Visualization Toolkit for Biomedical, Social Science and Physics Research.This project will design, evaluate, and operate a unique distributed, shared resources environment for large-scale network analysis, modeling, and visualization, named Network Workbench (NWB). The envisioned data-code-computing resources environment will provide .. more
How to cite this project



Hi students,

We wanted to let you know that there is a new assignment posted today that students can complete to earn an extra 2% to their final grade. (https://iu.instructure.com/courses/1595331/quizzes/2271360)

To earn the extra credit, students will have to create an account with XSEDE cloud computing program, and then run a compute instance on the Jet Stream platform to test running a workflow in the cloud environment.

To earn the extra credit, you will need to create your XSEDE account and submit your username to us by Monday April 24 at noon, then we'll link your account to Jetstream. Then from April 25-April 27th at noon, you will need to create an instance on the Jetstream, run a workflow, and submit your result to us.

Best,

Michael, Katy, and Andreas

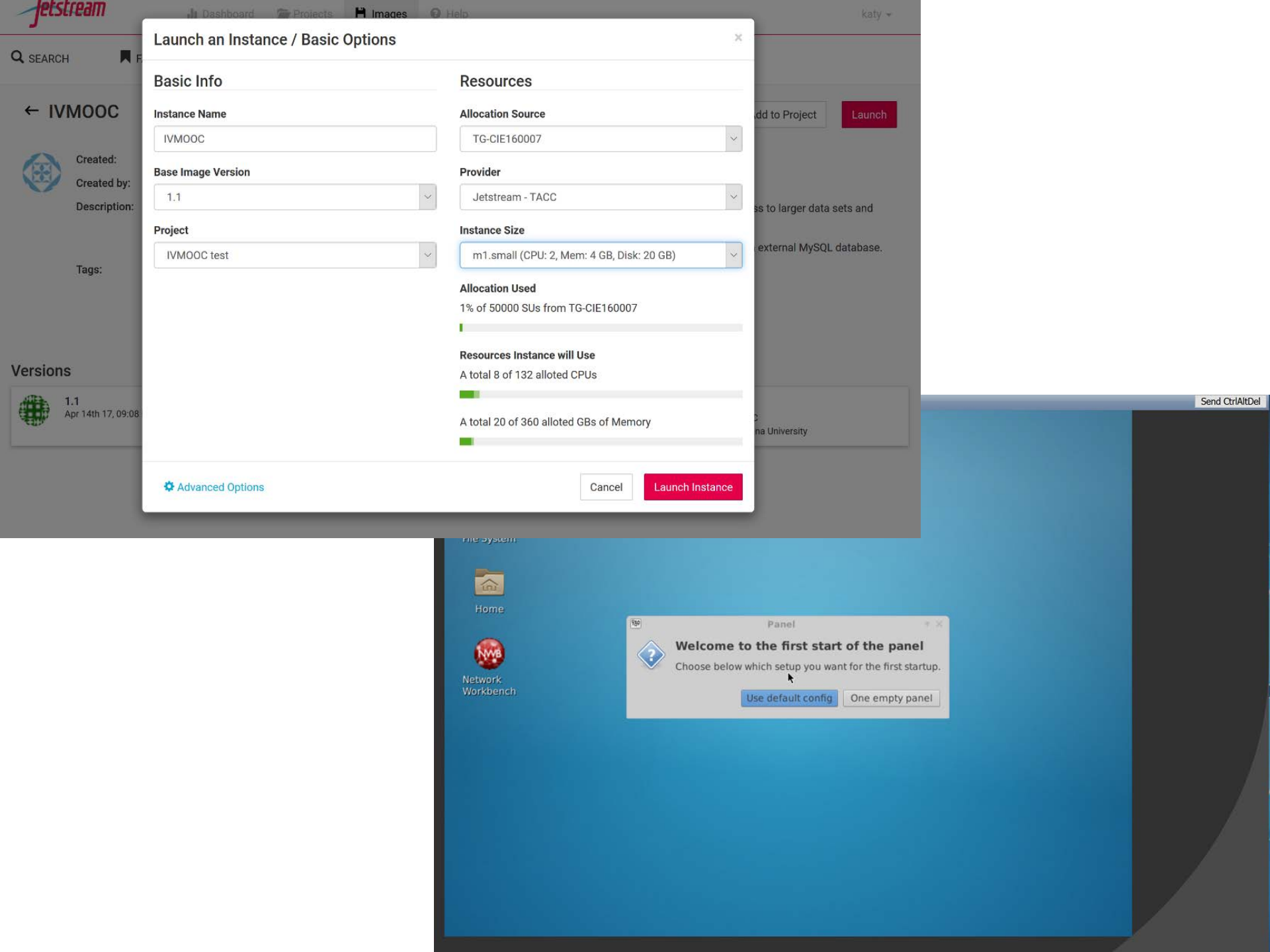
Account Setup

In order to use the IVMOOC Virtual Desktop via the Jetstream cloud environment, you need to get an XSEDE user account and this account needs to be linked to the IVMOOC cloud instance so that you can use this unique resource.

- Create an XSEDE user account via XSEDE User Portal.
- Submit your XSEDE user name via Canvas by 4/24, noon.
  (XSEDE user names were batch-added to IVMOOC VM)
- All IVMOOC students who submitted their XSEDE user name on time gained access to IVMOOC instance on Jetstream by 4/25.

Create to IVMOOC Instance on Jetstream

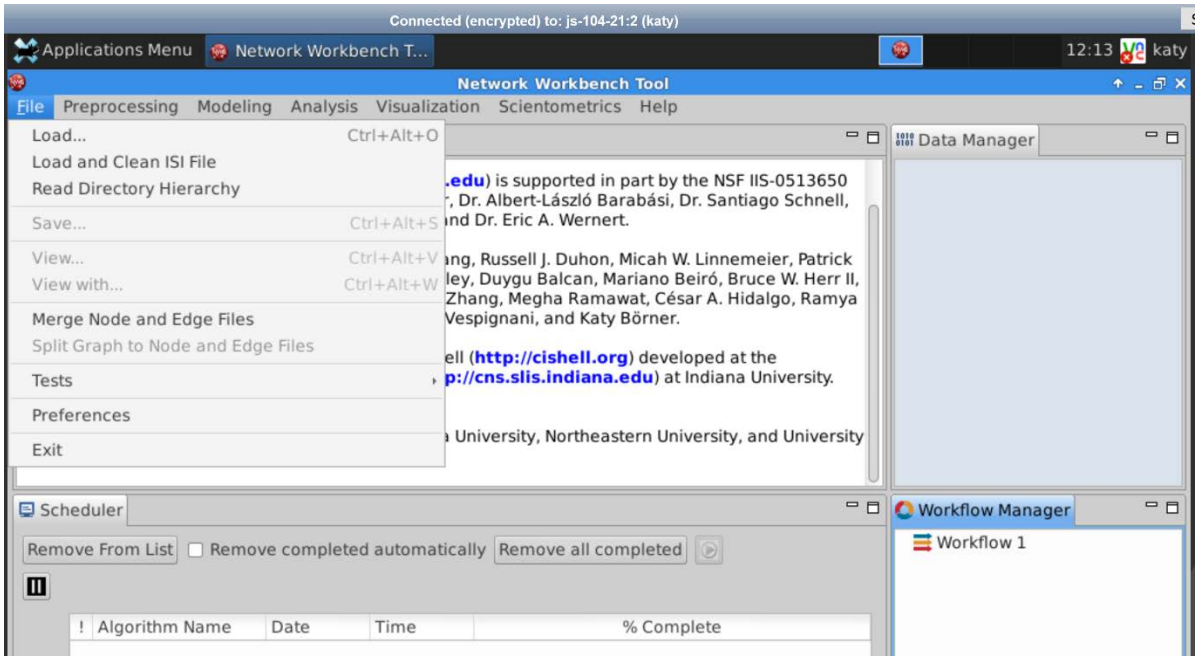
- After 4/25, noon, go to https://use.jetstream-cloud.org/application/images
- Login to use Jetstream using your existing organizational login, select "Indiana University" and use DUO.
- Select "Launch New Instance"
- Search for "IVMOOC" or directly go to https://use.jetstream-cloud.org/application/images/366
- Click on "Launch" in top right.
- Keep default values for the instance but increase "Instance Size" to **m1.small** (2 CPUs, 4096 GB memory, 20 GB disk). Click "Launch Instance" in lower right, see next slide.
- Wait until Status is "Active" then click on Instance with Name "IVMOOC."
- Wait until "Activity" is N/A (about 1-2 mins). The VW is now ready for usage.
- Reload page. "Open Web Desktop" should now be visible in lower right. Click on it to open virtual desktop in web browser.
- Confirm "Use default config."


Launch an Instance / Basic Options
Basic Info
Instance Name
IVMOOC
Base Image Version
1.1
Project
IVMOOC test
Resources
Allocation Source
TG-CIE160007
Provider
Jetstream - TACC
Instance Size
m1.small (CPU: 2, Mem: 4 GB, Disk: 20 GB)
Allocation Used
1% of 50000 SUs from TG-CIE160007
Resources Instance will Use
A total 8 of 132 alloted CPUs
A total 20 of 360 alloted GBs of Memory
Advanced Options
Cancel
Launch Instance
Send CtrlAltDel
Home
Network Workbench
Panel
Welcome to the first start of the panel
Choose below which setup you want for the first startup.
Use default config
One empty panel

Run NWB Tool

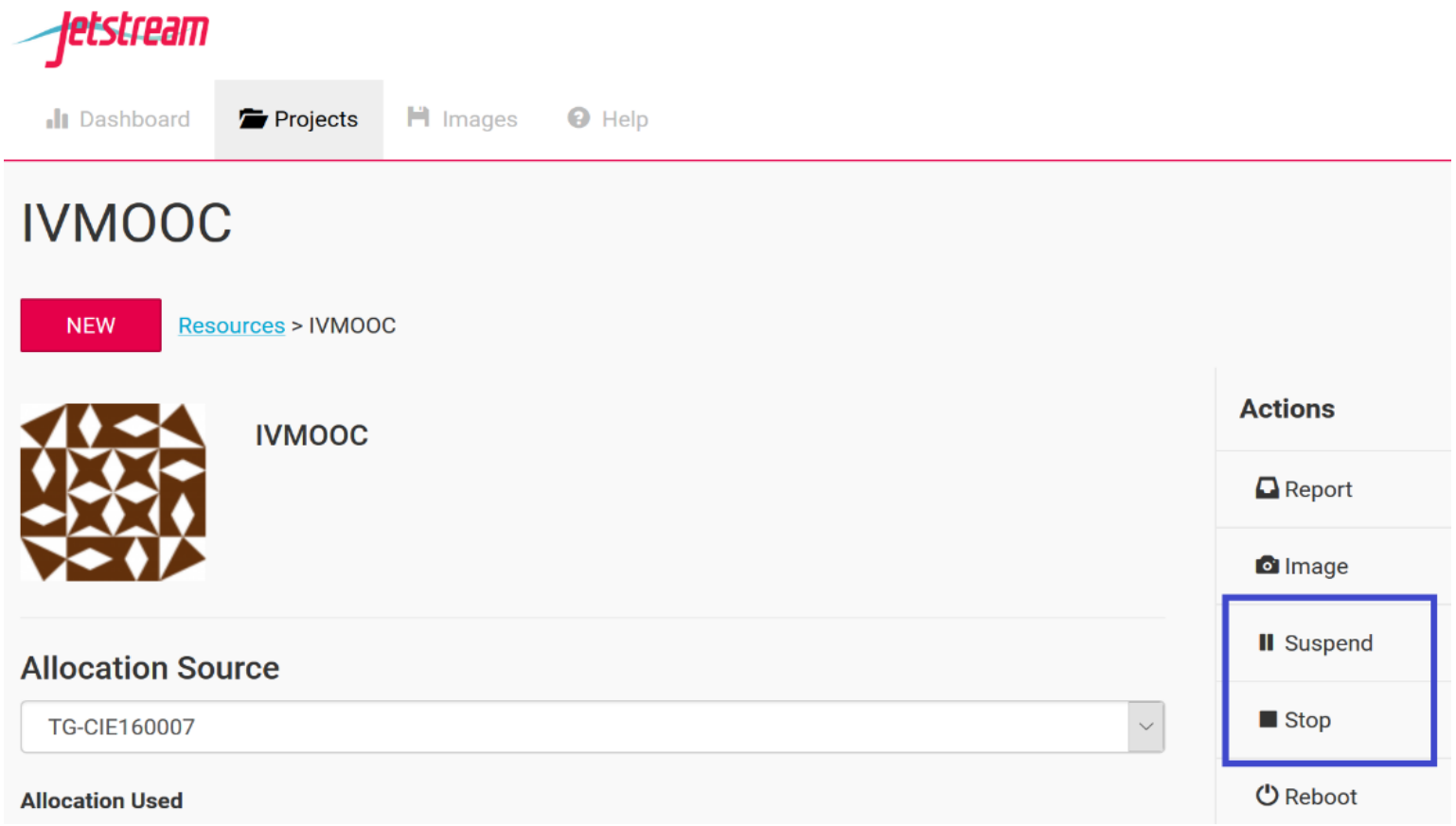
- On Web Desktop, double click Network Workbench (NWB) to run the tool.
- Run “Model > Hypergrid” and generate a network with 10,000 nodes and a Maximum degree of each node: 4.
- With the Hypergrid Network Model being selected in the Data Manager, run “Analysis > Network Analysis Toolkit (NAT).”
- Record the “Average degree” for nodes and submit value via Canvas.
- Feel free to try out other workflows.

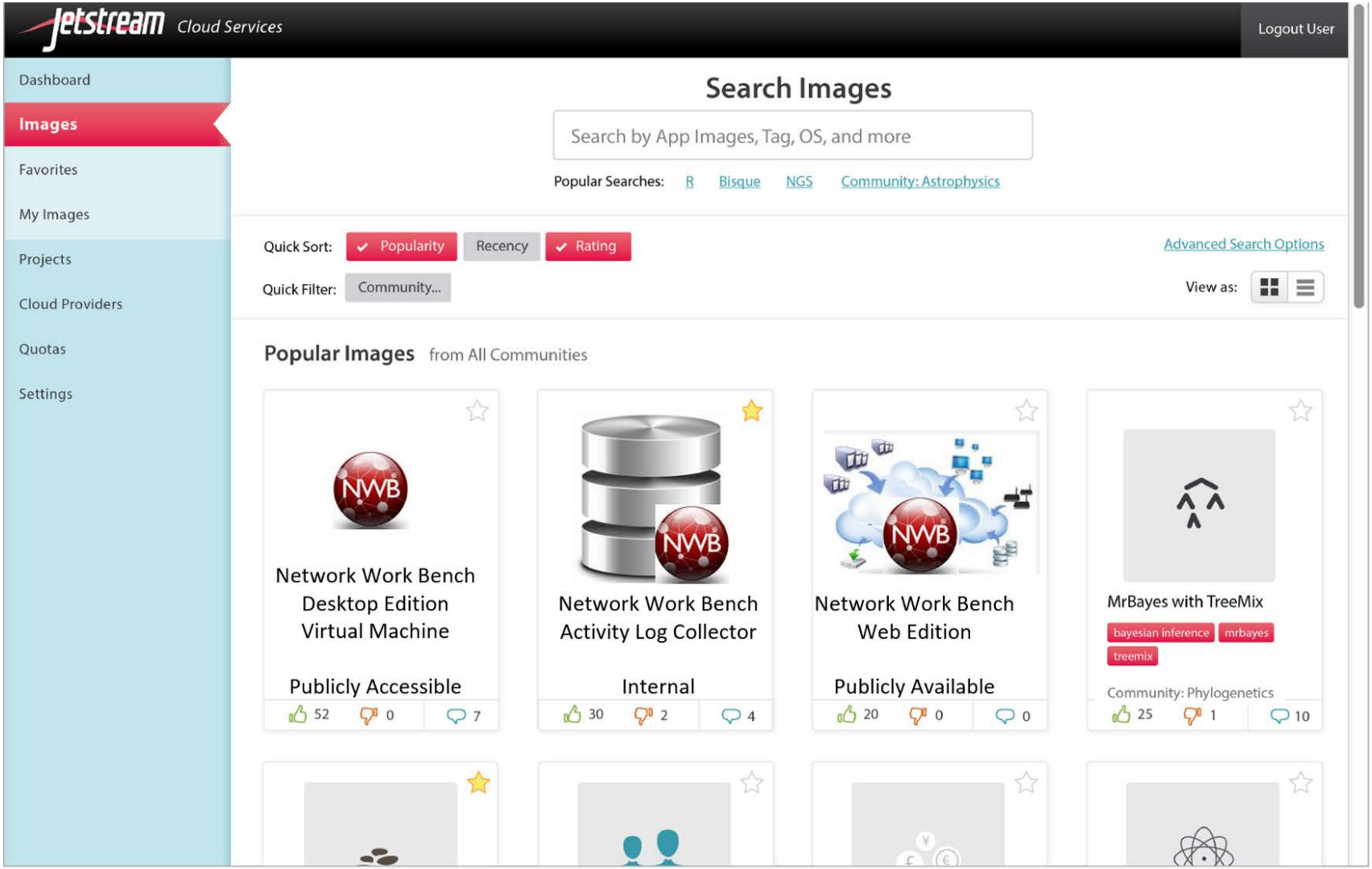
Fletcher, George, Hardik Sheth, and Katy Börner. 2005. "Unstructured Peer-to-Peer Networks: Topological Properties and Search Performance". *Agents and Peer-to-Peer Computing: Third International Workshop, AP2PC 2004. New York, NY* 14-27.

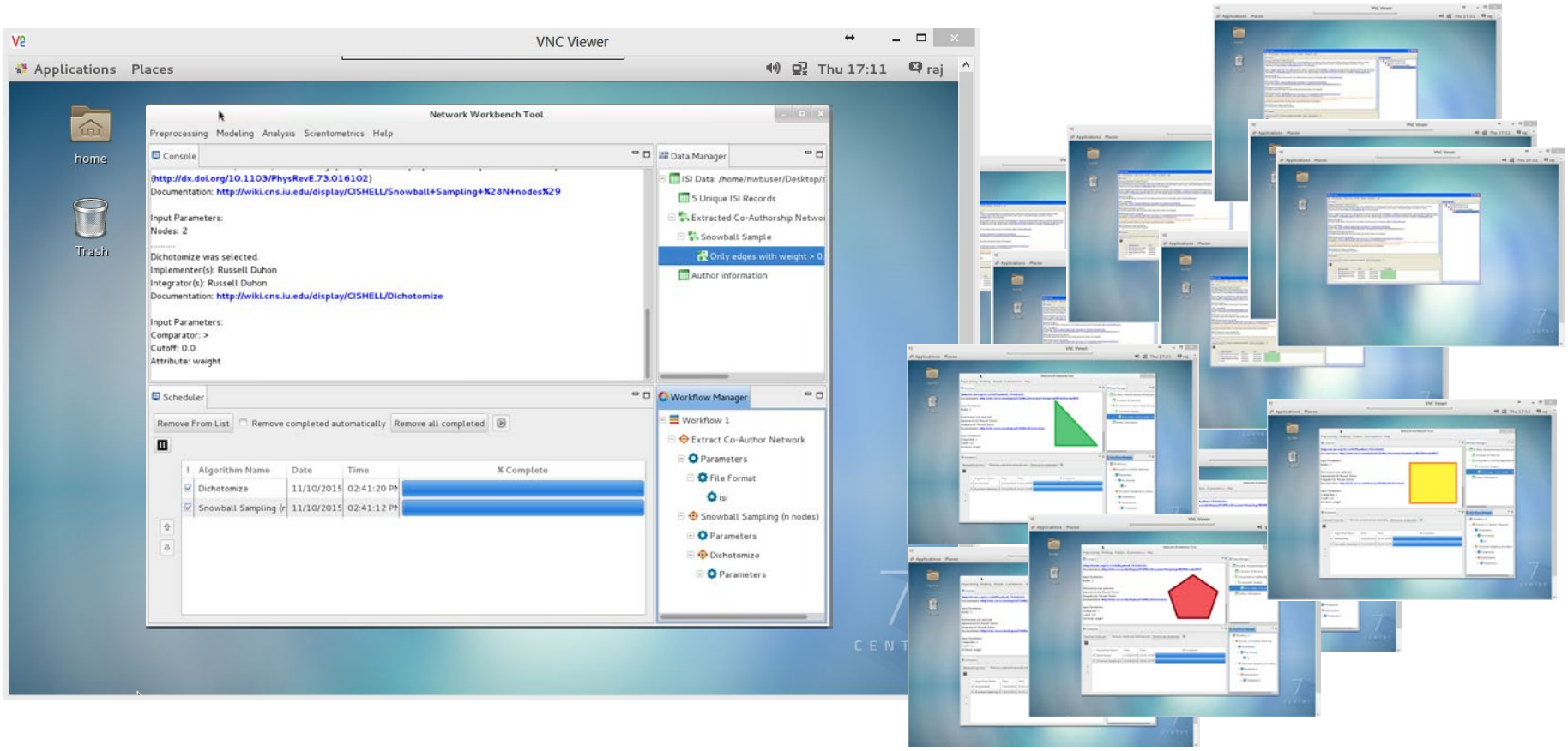


Suspend/Stop IVMOOC Instance on Jetstream

- When done, go back to Jetstream web interface and click on “Suspend” or “Stop” in Actions list on right:



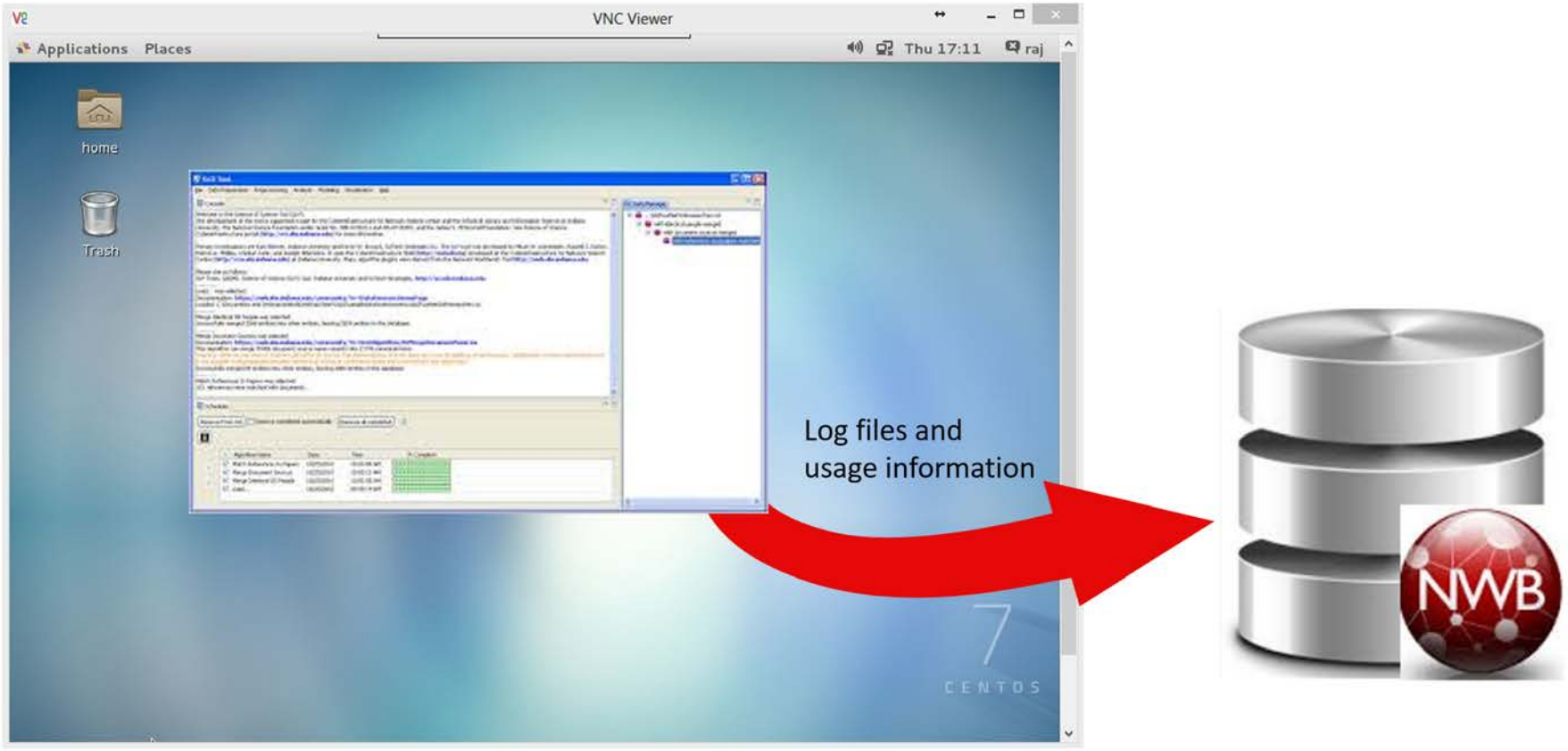

Jetstream Cloud Services
Logout User
Dashboard
Images
Favorites
My Images
Projects
Cloud Providers
Quotas
Settings
Search Images
Search by App Images, Tag, OS, and more
Popular Searches: R Bisque NGS Community: Astrophysics
Quick Sort: Popularity Recency Rating
Advanced Search Options
Quick Filter: Community...
View as:
Popular Images from All Communities
Network Work Bench Desktop Edition Virtual Machine
Publicly Accessible
52 0 7
Network Work Bench Activity Log Collector
Internal
30 2 4
Network Work Bench Web Edition
Publicly Available
20 0 0
MrBayes with TreeMix
bayesian inference
mrbayes
treemix
Community: Phylogenetics
25 1 10



One Instance: Single VM ready to run. Always the latest and stable version.

Multiple instances: Ability to spin up and down as desired. Resources on demand.

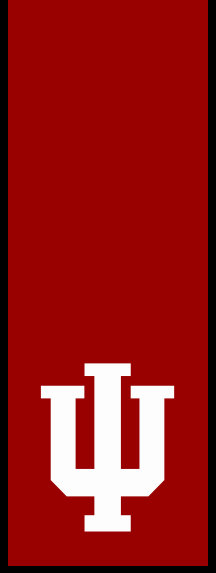
Data Replicability: VM Snapshots and templates allow instant replicability & reuse



Logging allows for capturing all user actions.

Supports student evaluation and guidance.

Helps identify frequency of dataset/alg usage per branch of science in support of future tool development.



Open XD Metrics on Demand Value Analytics

CNS & UITS

INDIANA UNIVERSITY BLOOMINGTON

XDMoD

METRICS ON DEMAND

Value Analytics Module

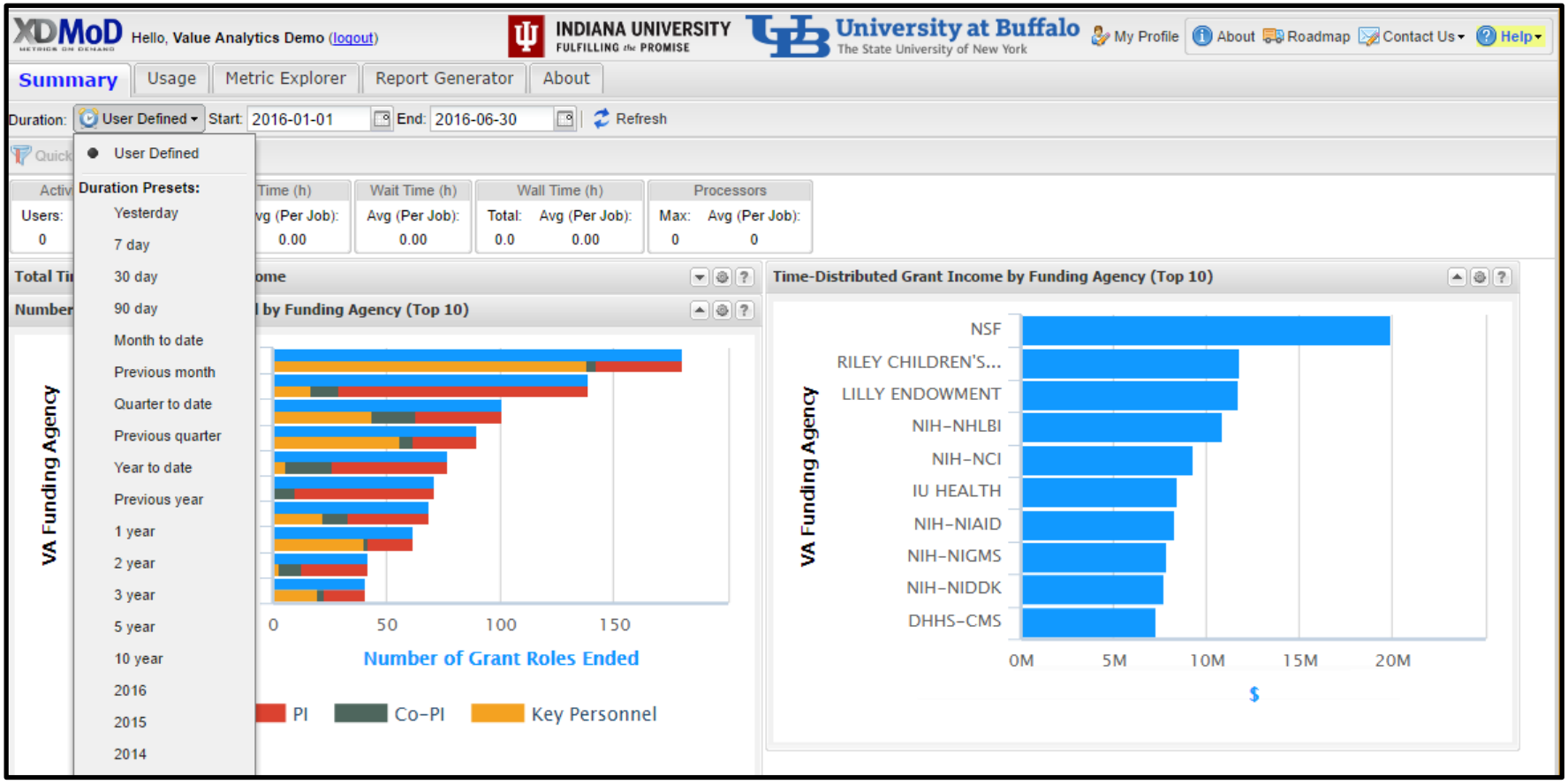
Matthew Link

February 24, 2017

Supported by the National Science Foundation

- Comprehensive resource management for HPC systems
- Provide detailed operational and usage data
- Support optimization of HPC resource utilization
- Facilitate planning and analysis
- Improve user experience through quality assurance
  - Application Kernels
  - Job-level Performance Analysis (SUPReMM)

Standard XDMoD Filtering Available



New XDMoD Filtering & Visualizations

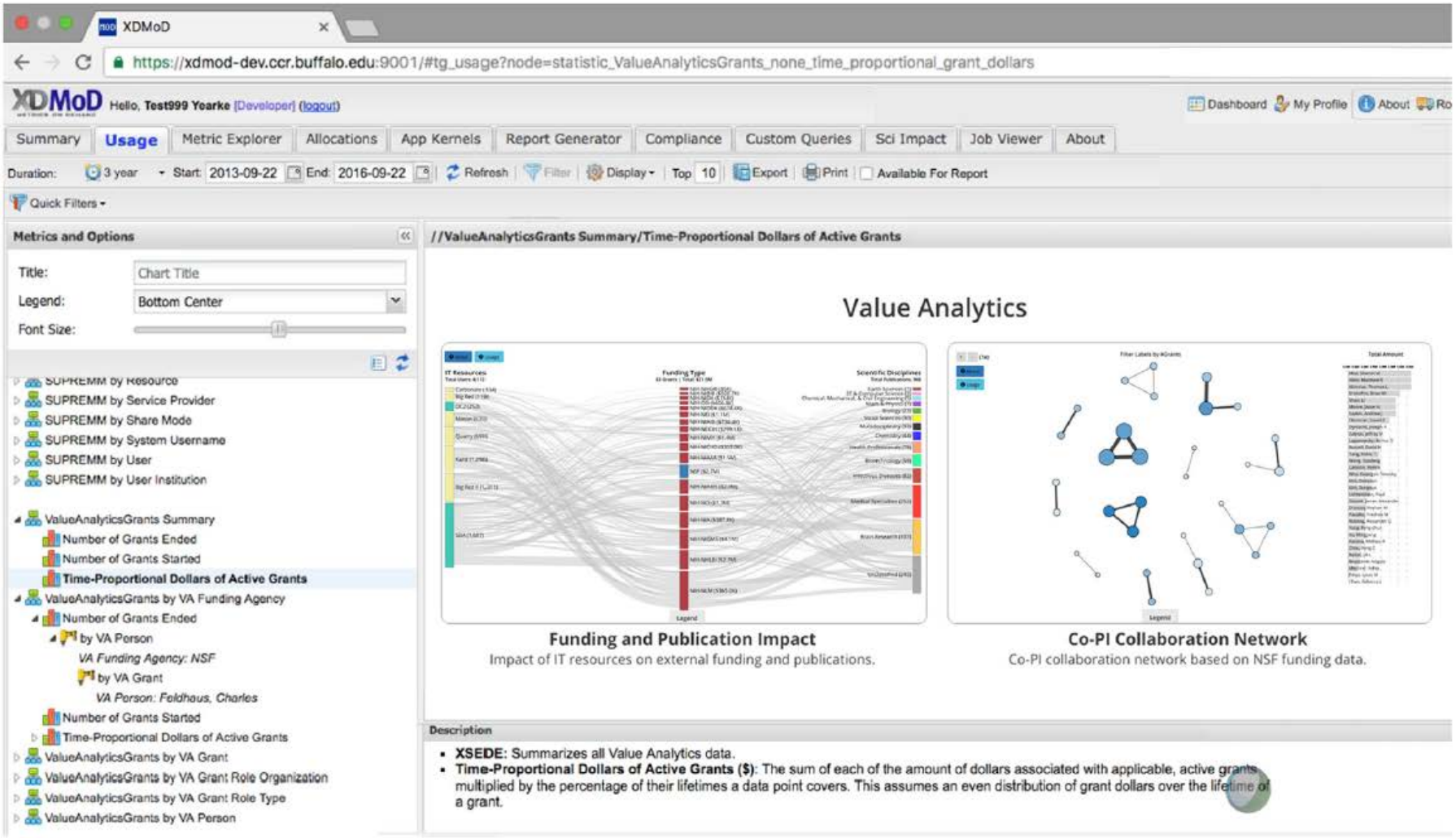


FIGURE 5 | XD metrics on demand value analytics Portal, see interactive version at http://demo.cns.iu.edu/xdmod-p/portal.html.

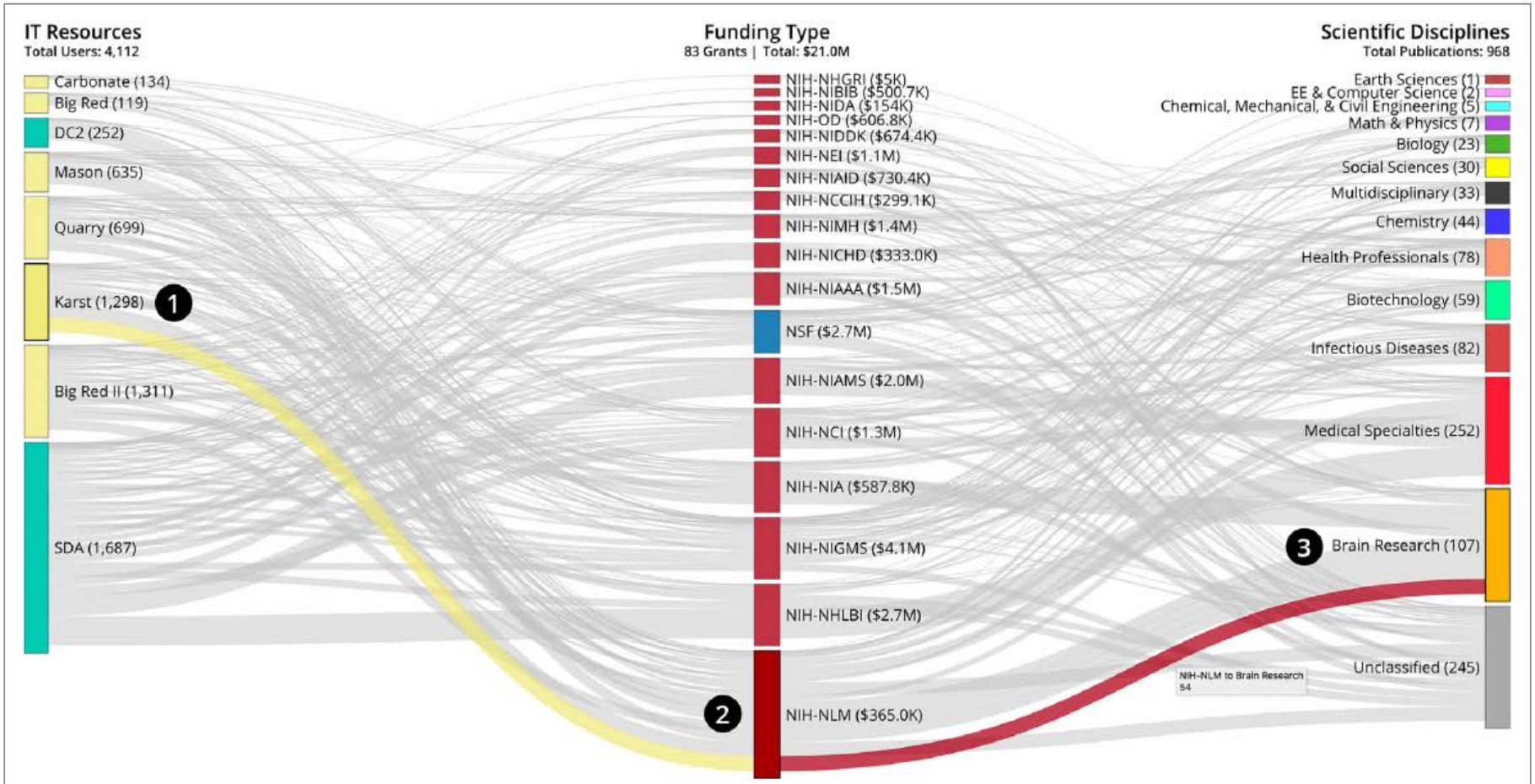


FIGURE 6 | XD metrics on demand value analytics *Funding and Publication Impact Visualization*, see interactive version at http://demo.cns.iu.edu/xdmod-p/impact.html.

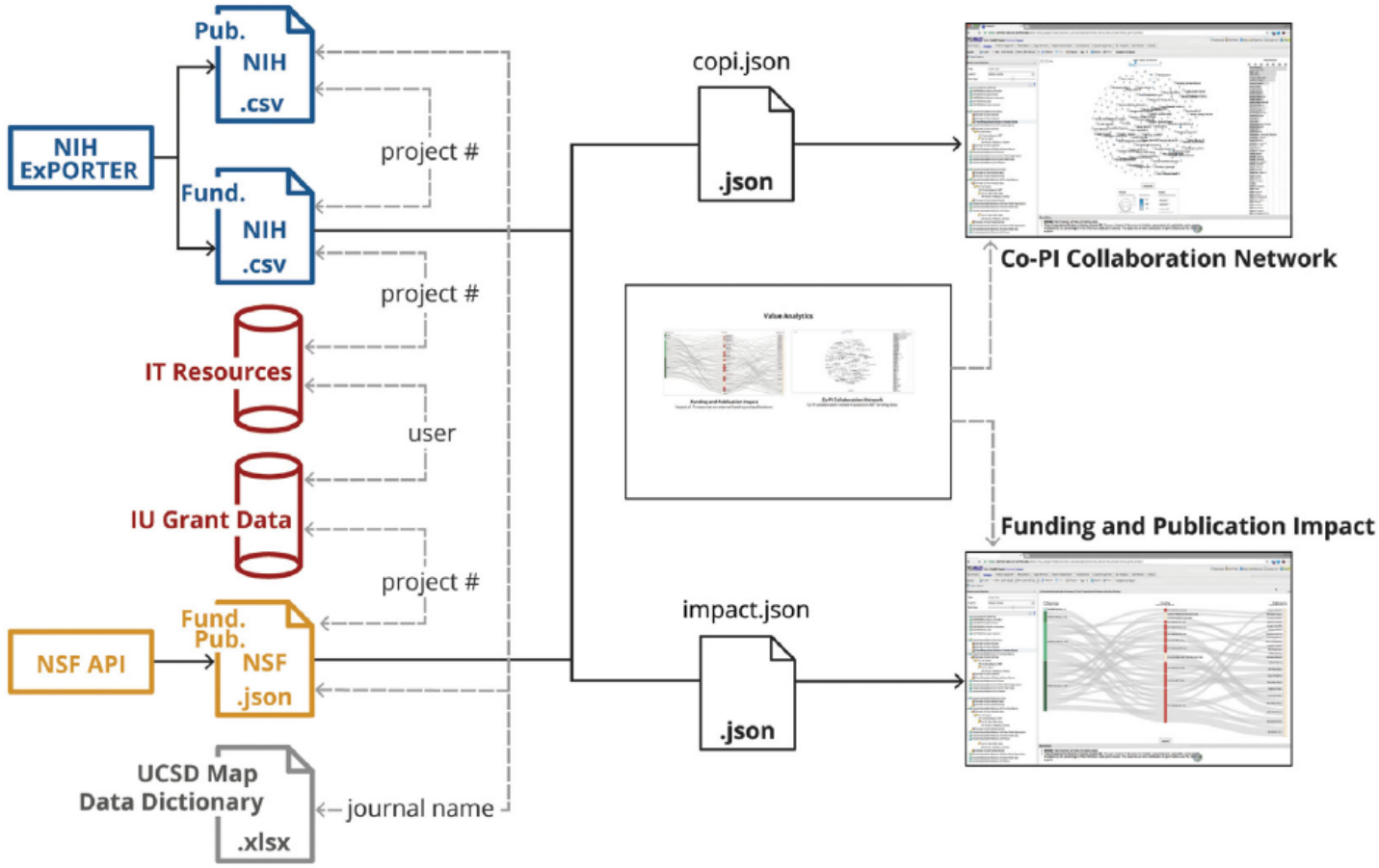


FIGURE 1 | Data sources, data linkage, data preparation for XD metrics on demand value analytics visualizations.

frontiers
in Research Metrics and Analytics

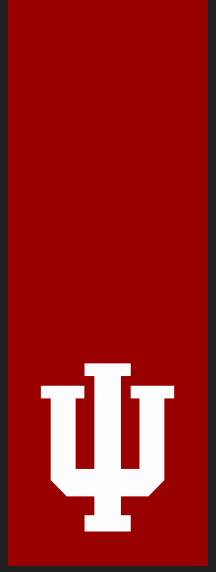
ORIGINAL RESEARCH
published: 29 January 2018
doi: 10.3389/frma.2017.00010

Check for updates

XD Metrics on Demand Value Analytics: Visualizing the Impact of Internal Information Technology Investments on External Funding, Publications, and Collaboration Networks

Olga Scrivner[1], Gagandeep Singh[1], Sara E. Bouchard[1], Scott C. Hutcheson[1], Ben Fulton[2], Matthew R. Link[2] and Katy Börner[1]*

[1] Department of Intelligent Systems Engineering, School of Informatics, Computing and Engineering, Indiana University Bloomington, Bloomington, IN, United States, [2] Pervasive Technology Institute, Indiana University Bloomington, Bloomington, IN, United States



Outlook: Multi-Level Maps of S&T

CNS & UITS

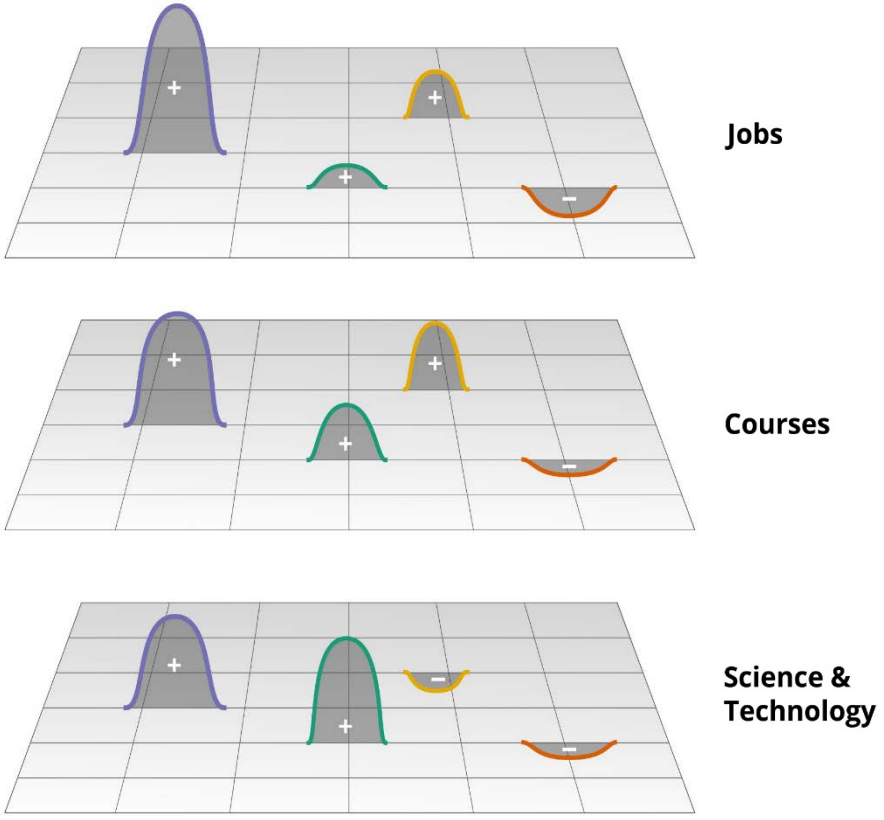
INDIANA UNIVERSITY BLOOMINGTON

Science & Technology vs. Education/Training vs. Jobs

Katy Börner, Olga B. Scrivner, Xiaozhong Liu, Indiana University

Study results are needed by:

- **Students:** What jobs will exist in 1-4 years? What program/learning trajectory is best to get/keep my dream job?
- **Teachers:** What course updates are needed? What curriculum design is best? What is my competition doing? How much timely knowledge (to get a job) vs. forever knowledge (to be prepared for 80 productive years) should I teach? How to innovate in teaching and get tenure?
- **Employers:** What skills are needed next year, in 5 years? Who trains the best? What skills does my competition list in job advertisements? How to hire/train productive teams?



What is ROI of my time, money, compassion?

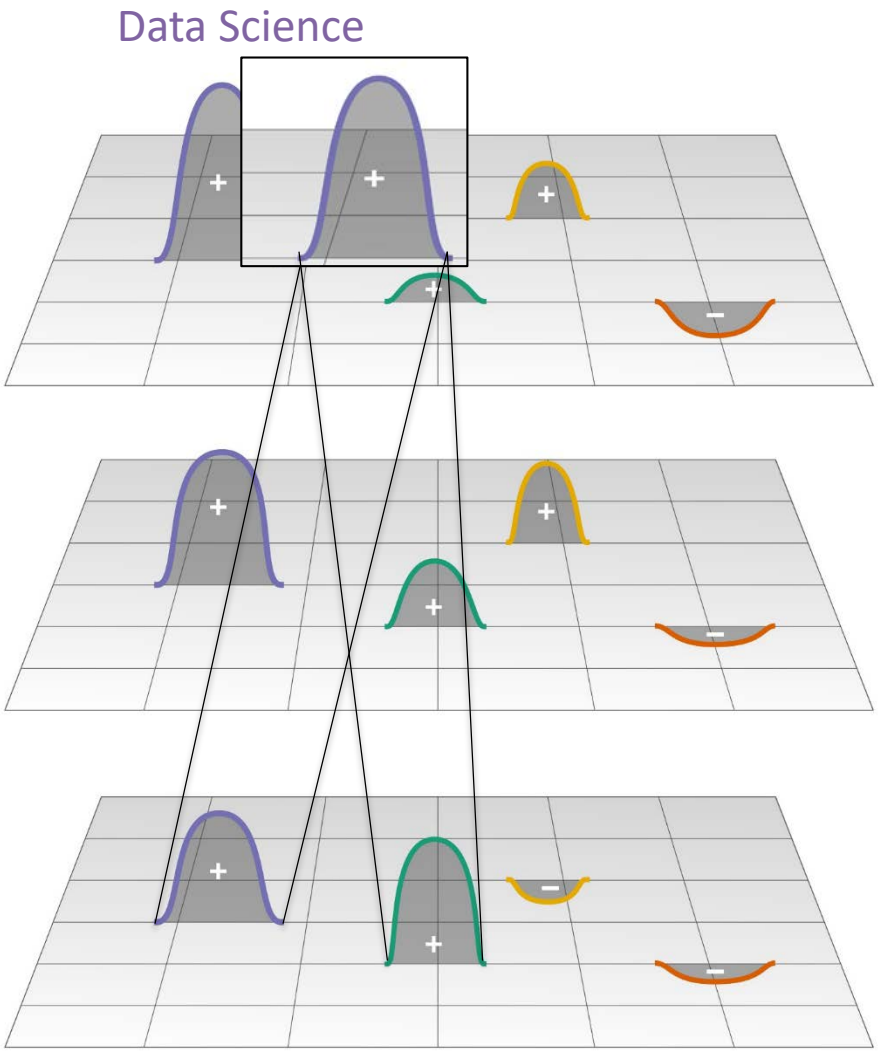
Science & Technology vs. Education/Training vs. Jobs

Katy Börner, Olga B. Scrivner, Xiaozhong Liu, Indiana University

Need to study the **(mis)match** and **temporal dynamics** of S&T progress, education and workforce development options, and job requirements.

Challenges:

- Rapid change of STEM knowledge
- Increase in tools, AI
- Social skills (project management, team leadership)
- Increasing team size



BC Analysis and Visualization of 50 Million Node Bimodal Networks

Katy Börner, Michael Ginda (CNS) and Thomas Sterling, Bo Zhang (CREST)

IU's IUNI is the home for 150 network science researchers. Several of them are interested to study network graphs (Wikipedia, Twitter, scholarly, brain data) that are too large to process.

The NWB tool (http://nwb.cns.iu.edu) is used by 100,000+ users around the globe cannot read/process graph larger than 100,000 nodes; most network layouts do not scale beyond 10,000 nodes. NWB tool on Jetstream (http://jetstream-cloud.org) does not yet support code parallelization.

This project combines CNS and CREST expertise parallelize code needed to analyze and map BIG networks to get a "x 100" advantage.

Initial dataset comprises 23,353,176 papers tagged with 27,624 unique MeSH terms. The undirected bimodal network has 23,353,176 + 27,624 nodes and 243,752,797 million edges.

Figure 2: Multi-level map of the world (left) and a sketch of the envisioned multi-level, topical STH map (right).