

# Interactive, web-based visualization of scientific collaboration networks

Ben Leone, Ryuichi Ohhata, Lukas Rosario

Jinyu Tian, Angela Vellante

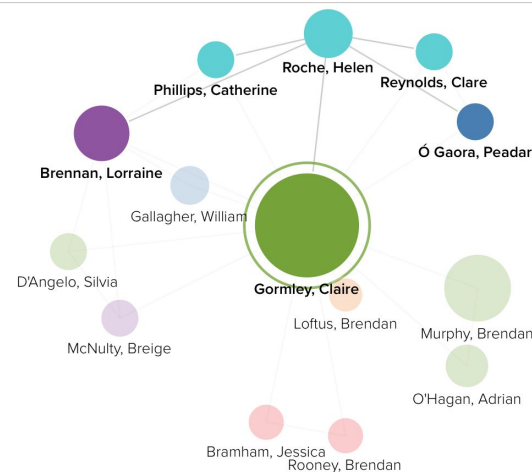
Mentors:

- Eric Kolaczyk (Hariri Institute, Director)
- Arezoo Sadeghi (SAIL)
- Margot Menestrot from Red Hat UX team
- Joe Farmer (BU Office of Research, Program Manager & Data Analyst)
- Jonathan Chamberlin



# Demo 1 Feedback and Improvements

- What is the MVP for this project?
  - A clean and intuitive visualization of the collaboration network between affiliates within the Hariri Institute
  - The size of the network will be ~250 affiliates
  - Admin accessibility to the database
- What cloud technologies will you be exploiting?
  - Cloud storage for hosting database (possibly SCC)
  - Docker for CI/CD and building database
- What database is this information coming from?
  - SciVal
  - Table on faculty affiliates page is hardcoded and manually updated by Institute staff
- What database will we use to store this information?
  - PostgreSQL

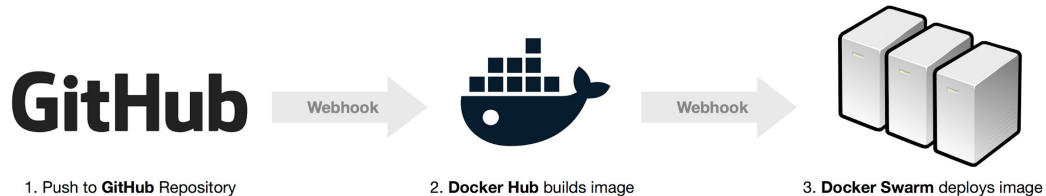
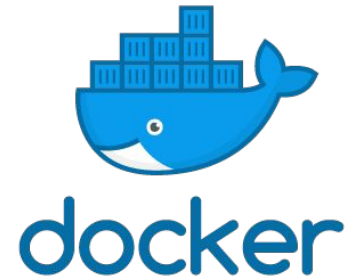


# SciVal

- Provides access to the research performance of over 14,000 institutions
- Perform a static pull of publications from SciVal yearly
  - Why didn't we choose to use a SciVal api to make requests?
    - Not amenable to hook up api
    - Project is a competitor
- What kind of metrics are we using?
  - Citations
  - Field weighted citation impacts
  - Possibly Scopus ID

# Docker

- Efficient Deployment with CI/CD
- Easier to manage in the future
- Scalability



# Mapping Science

- Mapping Science
- Use authors of articles as units
  - Mapping the patterns of co-authorship among individuals in a given scientific discipline
- Measures based on citations
  - The citation of one paper by the authors of another is an explicit and universally accepted manner of acknowledging a connection between the two
- To measure the strength of the citation between the two journals, we will use inter-citation frequencies
  - That is, for each journal pair  $(i,j)$ , the strength of their relationship will be quantified by the number of times  $C_{ij}$  that journal  $i$  in a given time period (one year) cites journal  $j$  in any time period
- Partial sampling: Databases may not include all journals published in all disciplines
  - Computer science publishing is done in the form of refereed conference proceedings papers, which are excluded from many of the main citation databases
- Network Graph Construction

# Sprint 2 User Stories, Objectives and Progress

## User

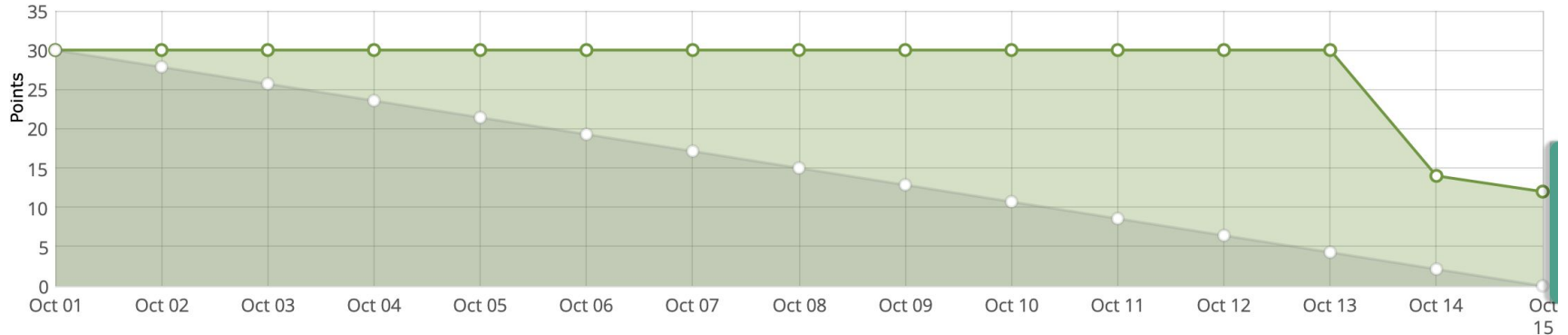
- See (sample) data that is from the current affiliates list in a web page served from the database
- Work towards using that in a visualization

## Developer

- Understand more clearly how we can integrate what we're creating into BU's WordPress environment
- Create a simple “drag-and-drop” flow so that admins of the page can simply upload a CSV to the site and the rest will be handled by the application

# Demo

# Taiga Burndown Chart



## Explanation:

- Uncertainty with how/where points were allocated and were burnt down
- Work was completed on several parts before current sprint



# Going Forward/Next Steps and Obstacles

## Sprint 3:

- Create a visualization from our API into our front-end and eventually into WordPress
- Implement analysis from SciVal into the network to show meaningful connections between nodes (researchers)
- Find hosting solution for the data and a way for Institute admins to access it
- Update project management with more concise user stories and specific tasks

# Going Forward/Next Steps and Obstacles

## Obstacles:

- Integrating our front-end into WordPress without using native plug-ins
- Connecting our API with a remotely hosted Postgres database that is low-cost/free
- Creating a streamlined workflow for submitting the SciVal data to our database

Thank you!  
Questions?