

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 3 Feedback

- Download powerpoint of demo slides and put them in the github repo with links from the README.md file
- Update README.md. Keep this as a living document.
- It would be nice to actually see your API.
- Concerned a bit that you are having problems in figuring out how to persist volume.
- It feels like you are continuing to make progress, but are blocking on pretty simple things, and it feels like you are not right now on a trajectory to do the super compelling visualization that we were hoping to see.

UI/UX Design Decisions

HARIRI FACULTY AFFILIATE PROGRAM

The Hariri Institute for Computing community consists of 257 Faculty Affiliates from 59 units, across 17 Boston University Schools and Colleges including departments in the College of Arts & Sciences, College of Engineering, Questrom School of Business, College of Communication, College of General Studies, Metropolitan College, School of Medicine, School of Public Health, School of Dental Medicine, School of Social Work, and Sargent College.

The Faculty Affiliate Program is open to all BU faculty members pursuing research projects, or leading teaching or training initiatives in computing or computational sciences. Faculty Affiliates are eligible to tap into Institute resources. To help promote highly-collaborative interdisciplinary research culture, our Faculty Affiliates are connected and interact through more focused activities and programs that are managed through [Institute centers and initiatives](#), each reflecting a distinct intellectual research scope and culture, shared by faculty and students spanning multiple computing disciplines.

In addition to participation in the intellectual endeavors of the Institute, Faculty Affiliates enjoy many tangible benefits, including:

- Eligibility for research incubation awards
- Access to Institute facilities, including event and meeting spaces
- Support for event planning and organization
- Pre/post-award grant administration for Institute-based projects
- Inclusion in Institute directory of faculty profiles and interests
- Dissemination of noteworthy achievements through Institute portals and social media
- Serendipitous networking with BU colleagues, fueled by endless, free espresso at the Institute

Faculty Affiliate Application

Directory of Affiliated Faculty

Search in the bar below to filter affiliates by common interests, department, or name.

Search table...

Full Name	Department	College	BU Email	Interests	Domains
Bin Gu	Information Systems	Questrom School of Business	bgu@bu.edu	Distributed Systems, and Cloud Computing; Big Data Knowledge Discovery, Management, and Mining; Machine Learning and Artificial Intelligence;	Digital Arts and Humanities Digital Health and Rehabilitation Sciences Biostatistics and Public Health Computational Finance and

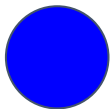
- Reach out to users of the website, so we can ask them their ideal workflow
- Visualization will be embedded in WordPress

- Refining the Visualization
 - Search: user should be able to search their interests or affiliate by name (like how it is now)
 - Filter: possible drop down list of departments; user should be able to add multiple departments

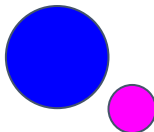


UI/UX Design Decisions

- Each **node** represents an individual person of the Hariri Institute.



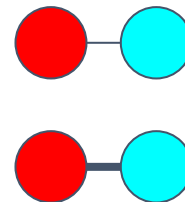
- The **size** of each node represents the number of publications for that author.



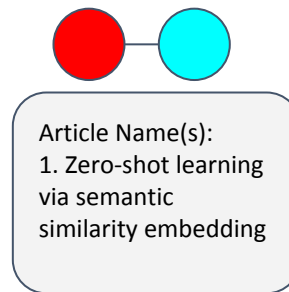
- Hovering** your mouse over the **node** displays that person's name.



- A **link** between two nodes represents a direct collaboration; the **thicker** the link, the greater the number of collaborations between them.



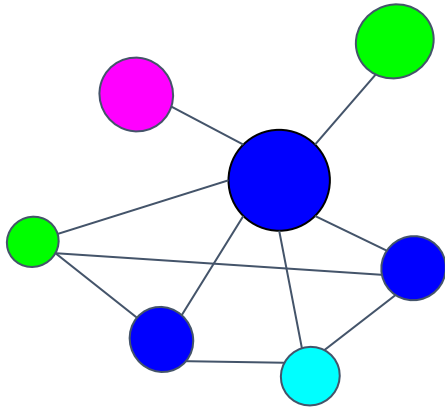
- Clicking** on a **link** will display name of the papers that connect them.



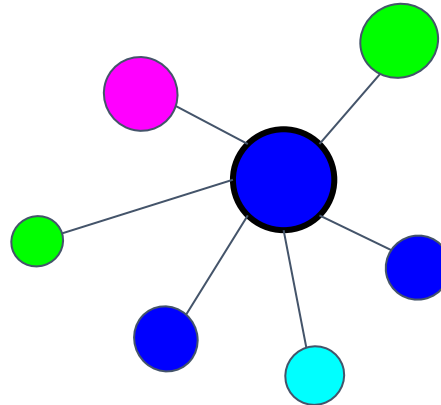
UI/UX Design Decisions

- **Clicking** on a **node** will display the person's name, email, number of publications, and department. It will also isolate direct links.

Before **click**:







After **click**:



Name: Orran Krieger
Email: okrieg@bu.edu
of Pub.: 18
Department: Electrical
& Computer Eng.

Department Key:

-  Biomedical
-  Electrical & Computer Engineering
-  Mathematics & Statistics
-  Computer Science

Current Obstacles

- Connecting to the OpenShift container to use the Postgres database
 - [Port forwarding](#) within OpenShift was the issue
 - However, once setup was unable to rsh into the database itself
 - Evaluating other ways to resolve these issues
- Using the persistent storage in OpenShift if our pod goes down
 - Issue with persistent storage unable to be removed or edited during testing
 - Submitted ticket to MOC and awaiting their response

Current Obstacles

OPENSIFT CONTAINER PLATFORM

Application Console

Interactive, web-based visualization of scientific collaboration networks

Search Catalog

Add to Project

Overview

Applications

Builds

Resources

Storage

Monitoring

Catalog

Storage

Learn More

Create Storage

Filter by label

Add

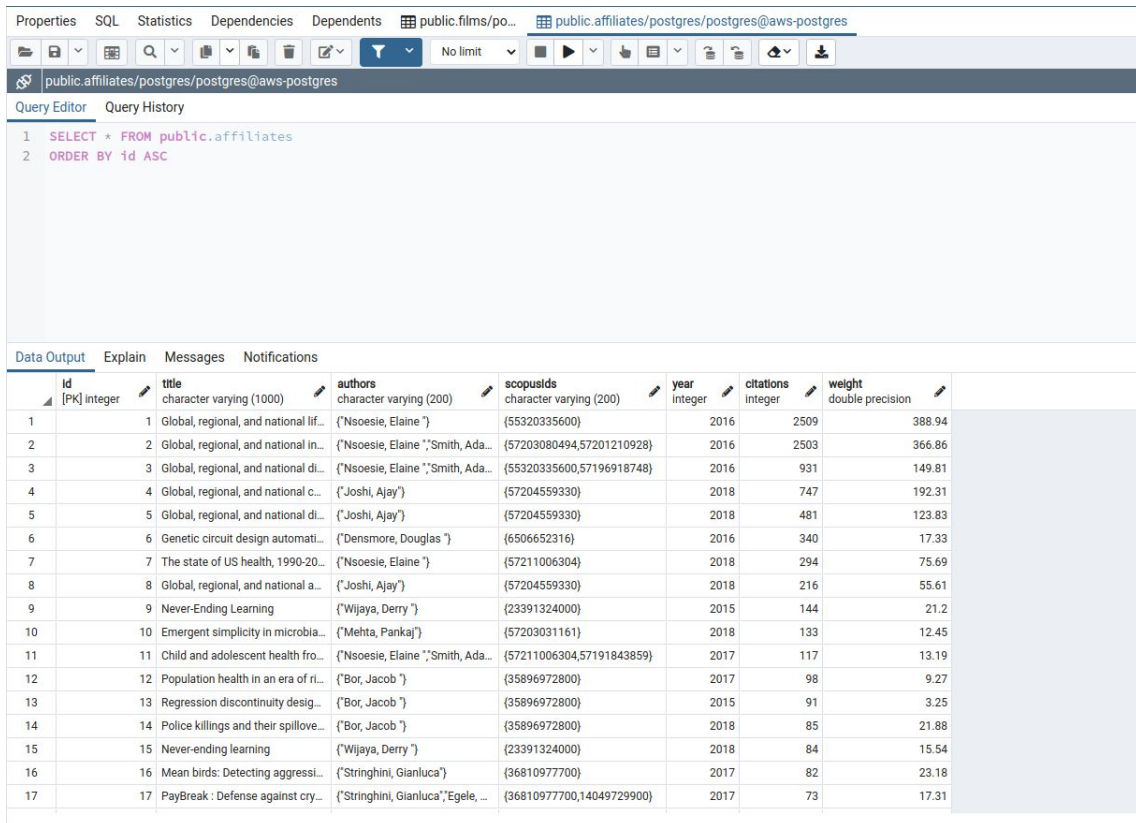
Name	Status	Capacity	Access Modes	Age
postgresql	✓ Bound to volume nfs0003	2 GiB	RWO (Read-Write-Once)	5 days

Current Obstacles

```
You have one project on this server: "ece-528-interactive-web-visualization".  
Using project "ece-528-interactive-web-visualization".  
benleone@benleone-Z370-HD3P:~$ oc get pods  
No resources found.  
benleone@benleone-Z370-HD3P:~$ oc delete pv postgresql  
Error from server (Forbidden): persistentvolumes "postgresql" is forbidden: User "bleone90@bu.edu" cannot delete persistentvolumes at the cluster scope: no RBAC policy matched  
benleone@benleone-Z370-HD3P:~$ █
```


Current Obstacles

- Using AWS RDS free tier as interim solution
- Slow for data uploads (db.t2.micro)
 - 1 vCPU
 - 1 GiB RAM
 - 20 GiB Disk



The screenshot shows a PostgreSQL query editor interface. The top bar includes tabs for Properties, SQL, Statistics, Dependencies, and Dependents. The main query editor displays the following SQL query:

```
1 SELECT * FROM public.affiliates
2 ORDER BY id ASC
```

Below the query editor, the 'Data Output' tab is active, showing the results of the query. The results are displayed in a table with 17 rows and 10 columns. The columns are: id, title, authors, scopusids, year, citations, weight, and two additional columns with icons. The data is as follows:

id	title	authors	scopusids	year	citations	weight			
1	Global, regional, and national lif...	{'Nsoesie, Elaine'}	{55320335600}	2016	2509	388.94			
2	Global, regional, and national in...	{'Nsoesie, Elaine','Smith, Ada...	{57203080494,57201210928}	2016	2503	366.86			
3	Global, regional, and national di...	{'Nsoesie, Elaine','Smith, Ada...	{55320335600,57196918748}	2016	931	149.81			
4	Global, regional, and national c...	{'Joshi, Ajay'}	{57204559330}	2018	747	192.31			
5	Global, regional, and national di...	{'Joshi, Ajay'}	{57204559330}	2018	481	123.83			
6	Genetic circuit design automati...	{'Densmore, Douglas'}	{6506652316}	2016	340	17.33			
7	The state of US health, 1990-20...	{'Nsoesie, Elaine'}	{57211006304}	2018	294	75.69			
8	Global, regional, and national a...	{'Joshi, Ajay'}	{57204559330}	2018	216	55.61			
9	Never-Ending Learning	{'Wijaya, Derry'}	{23391324000}	2015	144	21.2			
10	Emergent simplicity in microbia...	{'Mehta, Pankaj'}	{57203031161}	2018	133	12.45			
11	Child and adolescent health fro...	{'Nsoesie, Elaine','Smith, Ada...	{57211006304,57191843859}	2017	117	13.19			
12	Population health in an era of ri...	{'Bor, Jacob'}	{35896972800}	2017	98	9.27			
13	Regression discontinuity desig...	{'Bor, Jacob'}	{35896972800}	2015	91	3.25			
14	Police killings and their spillove...	{'Bor, Jacob'}	{35896972800}	2018	85	21.88			
15	Never-ending learning	{'Wijaya, Derry'}	{23391324000}	2018	84	15.54			
16	Mean birds: Detecting aggressi...	{'Stringhini, Gianluca'}	{36810977700}	2017	82	23.18			
17	PayBreak : Defense against cry...	{'Stringhini, Gianluca','Egele, ...	{36810977700,14049729900}	2017	73	17.31			

Current Database Content

- Title
 - Authors (multiple)
 - Scopus IDs (multiple)
 - Year
 - Citations
 - Field-weighted citation impact
-
- Name matching algorithm to find BU affiliates

How the Algorithm works

Spreadsheet from Hariri Institute

181	Lori, Noora	Internati
182	Gursky, Olga	Physiolo
183	Krieger, Orran	Electrica
184	Kotiuga, P. Robert	Electrica
185	Mehta, Pankaj	Physics

Krieger, Orran

Spreadsheet from SciVal

545	A survey b	Xu, X.	Hua	571890272	2016	Journal of t	6	0.59	10.1016/j
546	Seeing into	Turk, A.	Chen, H.	Tuncer, O.	Li, H.	Li, Q.	Krieger, O.	Coskun, A.K.	
547	Event excit	Khazaeni,	366187569	2016	2016 13th		6	2.11	10.1109/

Problem with Current Name Matching Algorithm

ECE Department

Li, Wenchao



Li, W

Astronomy Department

Li, Wen



Li, W

Algorithm gets “confused” and ends up listing both authors

Solution to the Current Algorithm

- Match by Scopus IDs instead of names
 - ◆ Have to manually search up each Prof's Scopus IDs...?

API

```
db_string =         
db = create_engine(db_string)
base = declarative_base()

class MemberAPI(Resource):
    # with open(os.path.dirname(__file__) + '/demo.json') as j:

    from models.memberitem import MemberItem
    Session = sessionmaker(db)
    session = Session()
    # memberlist = Affiliate_Member

    base.metadata.create_all(db)
    members = session.query(MemberItem)
    memberlist = []
    for member in members:
        memberlist.append({
            "id": member.id,
            "title": member.title,
            "authors": member.authors,
            "scopusIds": member.scopusIds,
            "year": member.year,
            "citations": member.citations,
            "weight": member.weight
        })

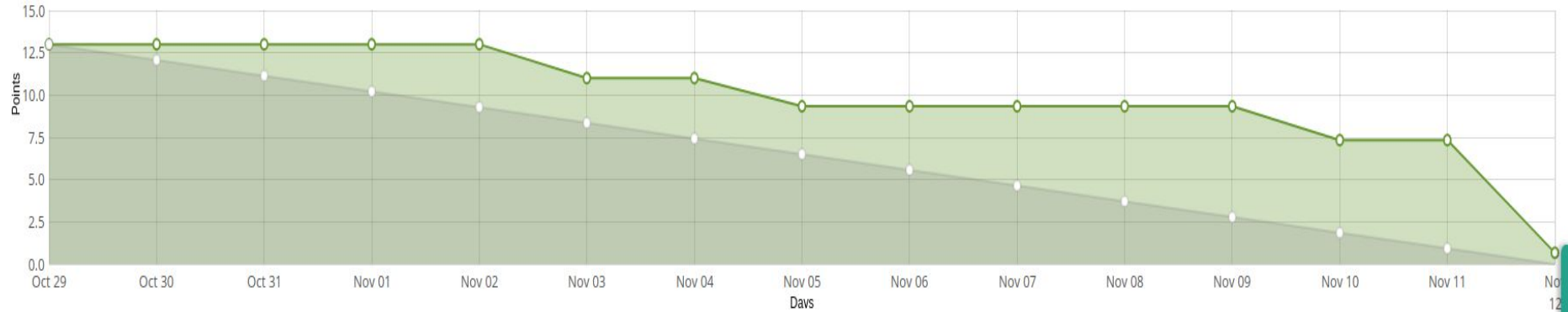
    def get(self):
        # print(self.memberlist)
        return self.memberlist
```

Demo

API

```
▼ 3059:
  id: 3059
  ▼ title: "Correction for \"Ligand deconstruction: Why some fragment binding positions are conserved and others are not,\""
  authors: "{\"Vajda, Sandor\"}"
  scopusIds: "{7005156142}"
  year: 2015
  citations: 2
  weight: 2.38
▼ 3060:
  id: 3060
  ▼ title: "Predicting active facial expressivity in people with Parkinson's disease"
  authors: "{\"Joshi, Ajay\", \"Ellis, Terry\", \"Betke, Margrit\"}"
  scopusIds: "{56915233900, 8315997800, 7004214113}"
  year: 2016
  citations: 2
  weight: 0.18
▼ 3061:
  id: 3061
  ▼ title: "Classifying android malware with dynamic behavior dependency graphs"
  authors: "{\"Zhang, Shengzhi\"}"
```


Taiga Burndown Chart



- Work to keep burndown linear and on time for Sprint 5 due to the longer sprint

Going Forward/Next Steps

Sprint 5 (Dec 3):

- Resolve remote database issues with OpenShift and connect to API
- Fix matching algorithm with affiliates and their works
- Implement visualization into WordPress environment
- Refine visualization for best UI/UX
- Finalize documentation on how to upload data to our database after the end of the project

Thank you!
Questions?