

Leveraging Your Research Data | Quantifying Research Collaborations



Robert Pilgrim, Ph.D.
Associate Director of Data Strategy & Insights
Division of Research & Innovation
University of Arkansas
rpilgrim@uark.edu
StreamlyneUp 2022

Introduction

Robert Pilgrim, Division of Research & Innovation,
University of Arkansas

Associate Director for Data Strategy & Insights

Streamlyne Customer since FY2014

Build systems and platforms with institutional data to help to support decision making across campus.

- Steamlyne Reporting Module | Leveraging external visualization/ analytics platforms
- Quantifying Collaborations
 - Visualizing your research data
 - Introduction to social network analytics and why you need them
- Quantifying the impact of a strategic initiative using your research data
- Why you need a data model | Bring together multiple data for cross sources data calculations.
- You have your data model, what's next?

Why is a Strategic Plan Important

- Alignment within and across the units.
- No clearly defined, quantifiable metrics or deliverables == not a serious plan.
- This is where research data plays a large role.

General Example of a Strategic Plan

Campus Strategic Plan

- Increase research collaborations
- Grow research expenditures
- Increase diversity
- Investing in faculty
- Increase research space
- Innovation in teaching & learning
- Improve graduate education
- Student success
- Improving the local community

Research & Innovation/Sponsored Programs

- Improve research infrastructure
- Improve research culture
- Develop our research clusters in areas x, y, and z
- Engage students in research
- Improve research commercialization
- Raise public awareness of our research activities

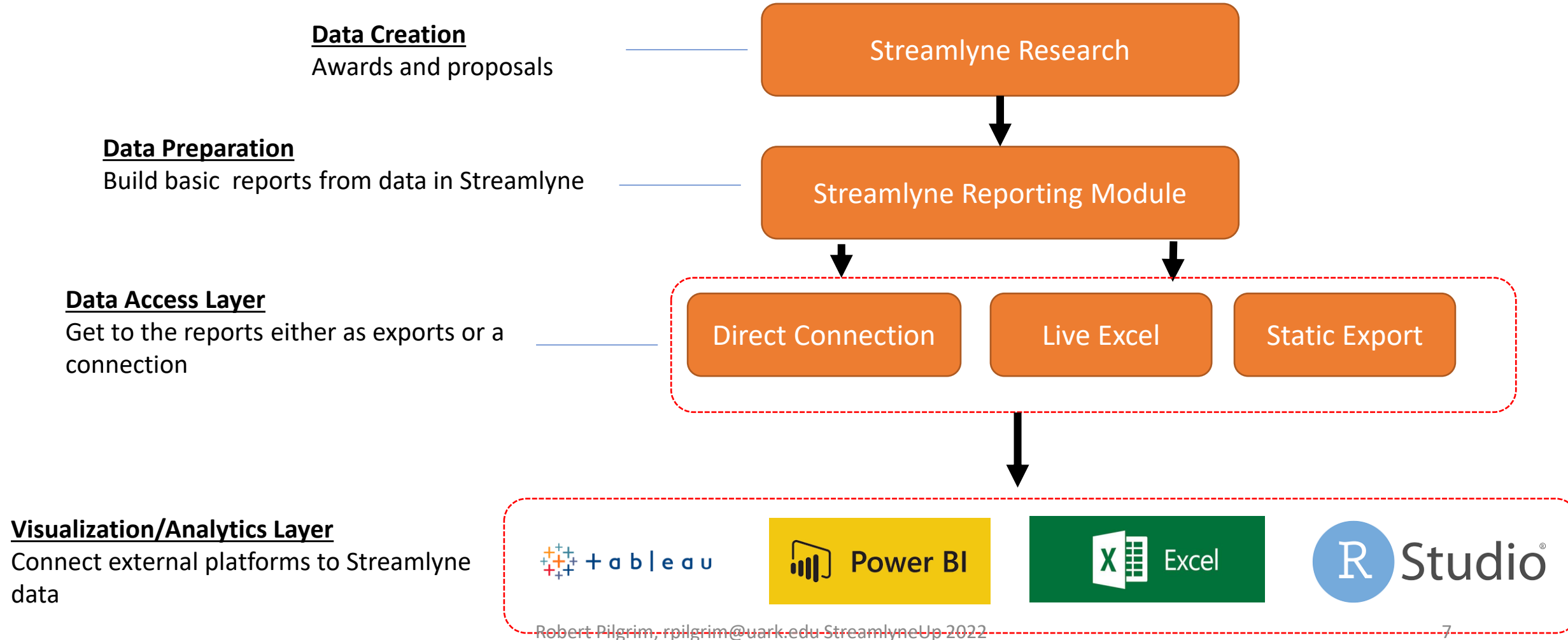


Leveraging the Streamlyne Reporting Module

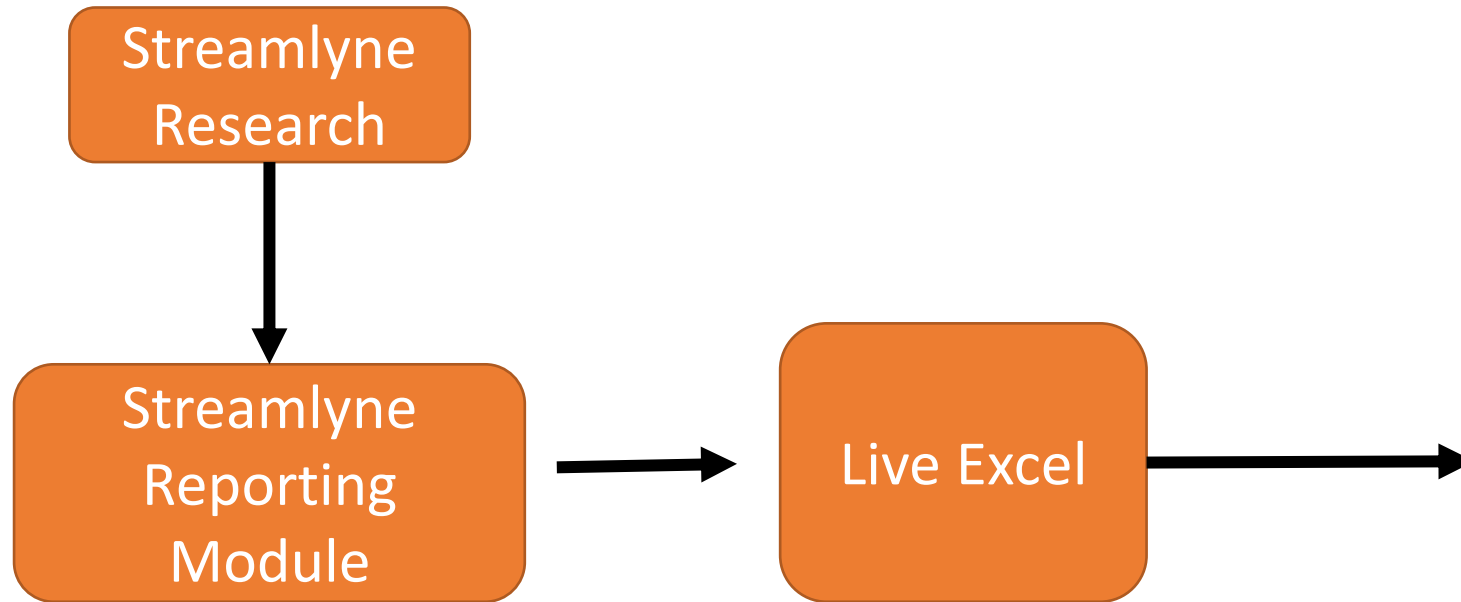


Leveraging External Analytics Platform

Why? If we want to build reports with additional data (e.g. Publications, research expenditures etc.) we need to get data out



Workflow



Live Excel file can be refreshed direct in Excel



Who is using live Excel ?

What other tools are people using ?

Do you have dedicated people for research analytics ?

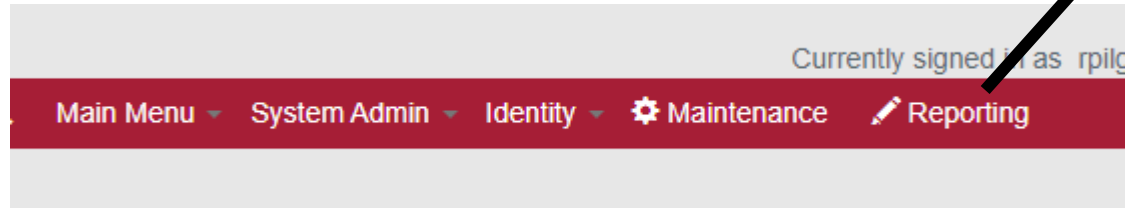
Does your office plan on hiring a dedicated person? Is this a priority ?

Variable List

Campus	Sponsor Originating	Date Last Data Refresh
Flag External Funding	Sponsor Originating Code	Date Obligation
Flag Research Funding	Sponsor Originating (Restricted)	Date Obligation Month
Flag Has Prime Sponsor	Sponsor Originating Funding Type	Date Obligation Day
Flag Group Gifts and Foundation	Sponsor Originating Funding Type Group	Date Obligation Year
Flag UITS RIS	Sponsor Award ID	Fiscal Year
Award Number	Sponsor Direct	Fiscal Month
Parent Unit Code (WD)	Sponsor Direct Code	Date Awarded
Parent Unit Name (WD)	Sponsor Direct Funding Type	Date Project Start
Parent Unit Name	Sponsor Direct Funding Type Group	Date Project End
Parent Unit Code	Investigator Role	Amount
Lead Unit Code (WD)	Key Person Project Role	
Lead Unit Name (WD)	Investigator Name	
Lead Unit Name	Investigator ID	
Lead Unit Code	Investigator Email	
Title	Investigator College Code (WD)	
Title (Restricted)	Investigator College Name (WD)	
	Investigator College Code	
	Investigator College Name	
	Investigator Department Code (WD)	
	Investigator Department Name (WD)	
	Investigator Department Code	
	Investigator Department Name	

How do I get a Live Excel File?

- Create/run your report in Streamlyne reporting module
- Export as “Live Excel Spreadsheet”
- Share this file



Reports Listing

awards amounts vs x

Report Title

Options

Streamlyne Awards Amounts vs Proposed Amounts

→ Launch

All new awards finalized after the date specified in the run control.

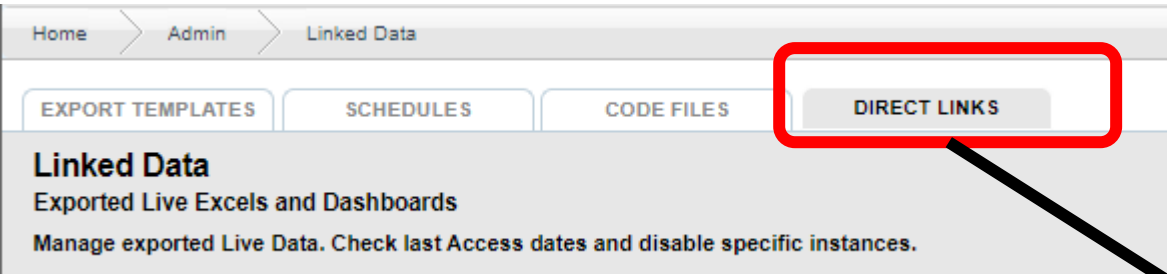
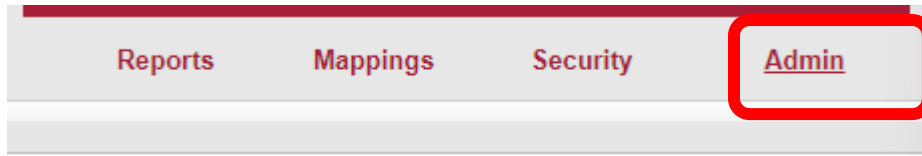
Export Results
Export this result listing

Note: Exporting data may result in the creation of unsecure/unencrypted or permanent files on your computer. Please contact your system administrator for any questions regarding the proper safeguarding of sensitive information.

- Excel Comma-Separated Values (.csv)
- Adobe PDF (.pdf)
- Webpage (.html)
- Customized Delimiter (.txt)
- Tab-Delimited (.txt)
- XML (.xml)
- Fixed Length Columns (.txt)
- Live Excel Spreadsheet

< Back

Managing Live Excel File Access



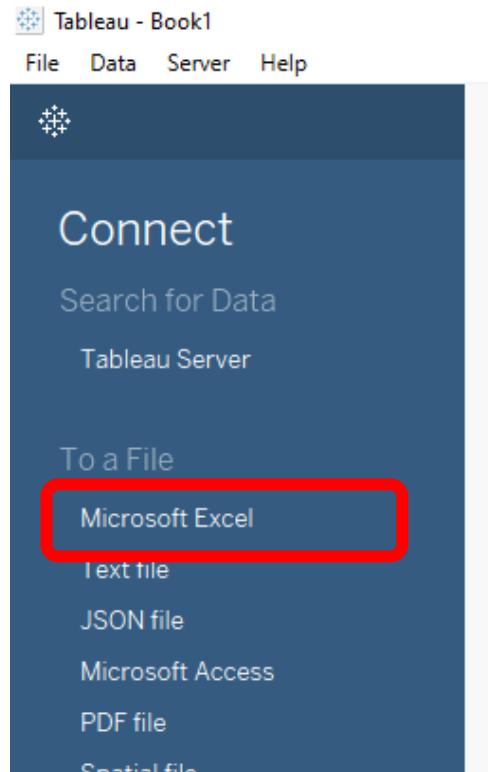
Type	Name	Status	Note	Created by	Last Accessed	IP Address	Options
	_Table_Proposal_Person	enabled					disable remove
	Streamlyne Active Awards	enabled					disable remove
	Streamlyne Award	enabled					disable remove
	Streamlyne Award Report with Persons	enabled					disable remove
	Streamlyne Award Report with Persons	enabled					disable remove

Live Excel files can be managed here

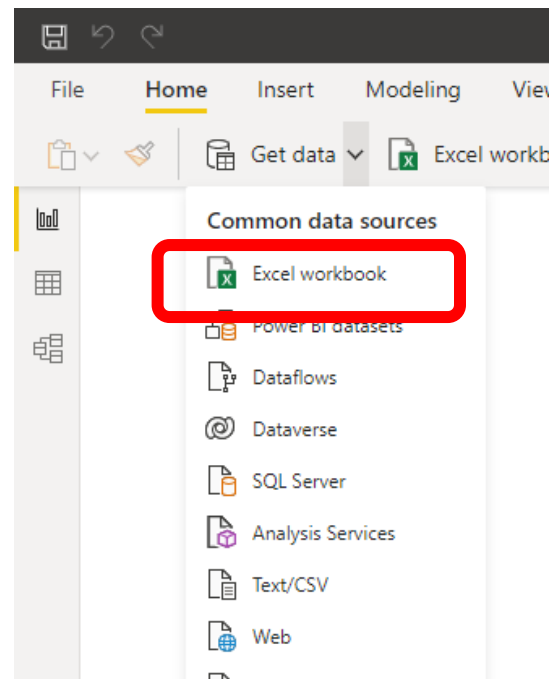
How do I use my data in a report

- Connect your platform of choice to the exported file
 - Powe BI, Tableau, Excel, SAS Visual Analytics, R etc.

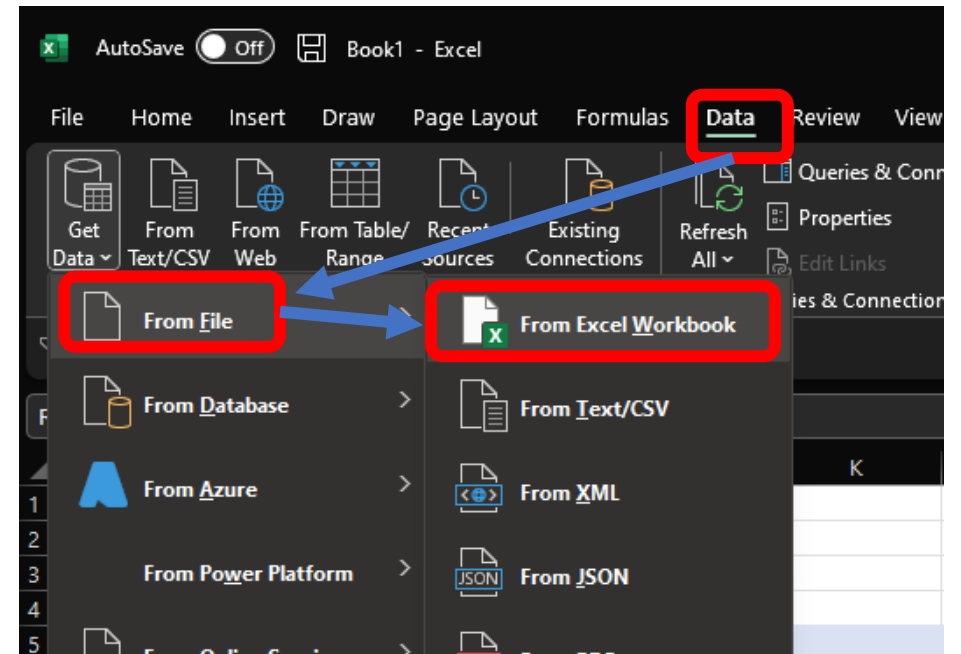
Tableau



Power BI



Excel

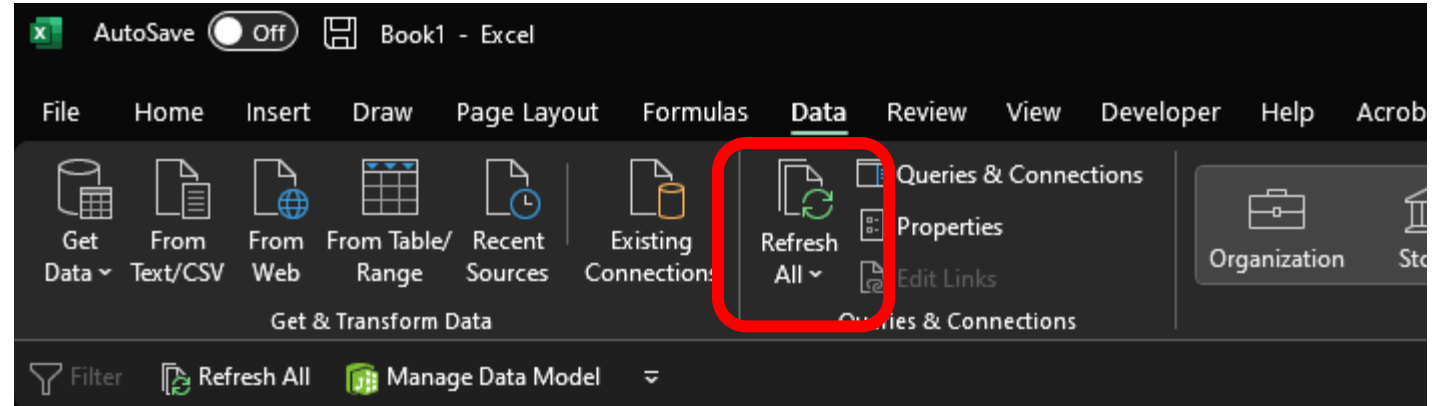


Updating Streamlyne Reports

Step 1

Refresh the “Live Excel” file.

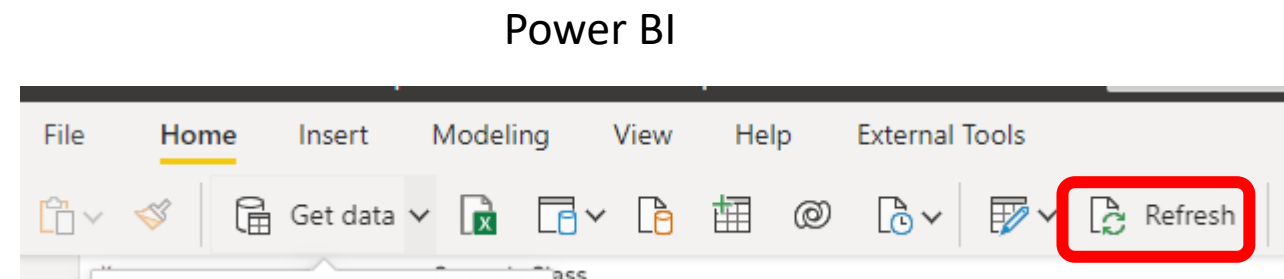
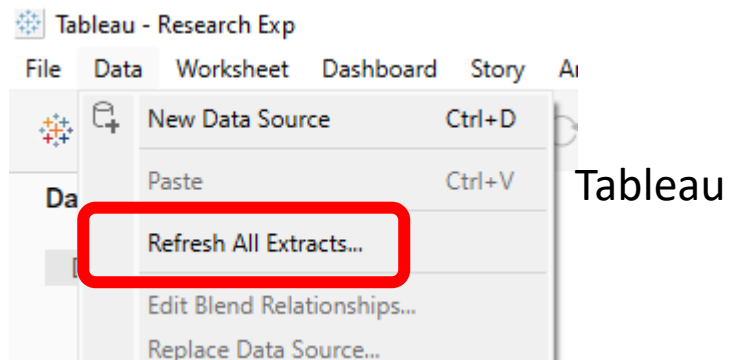
Open file and click “Refresh All”



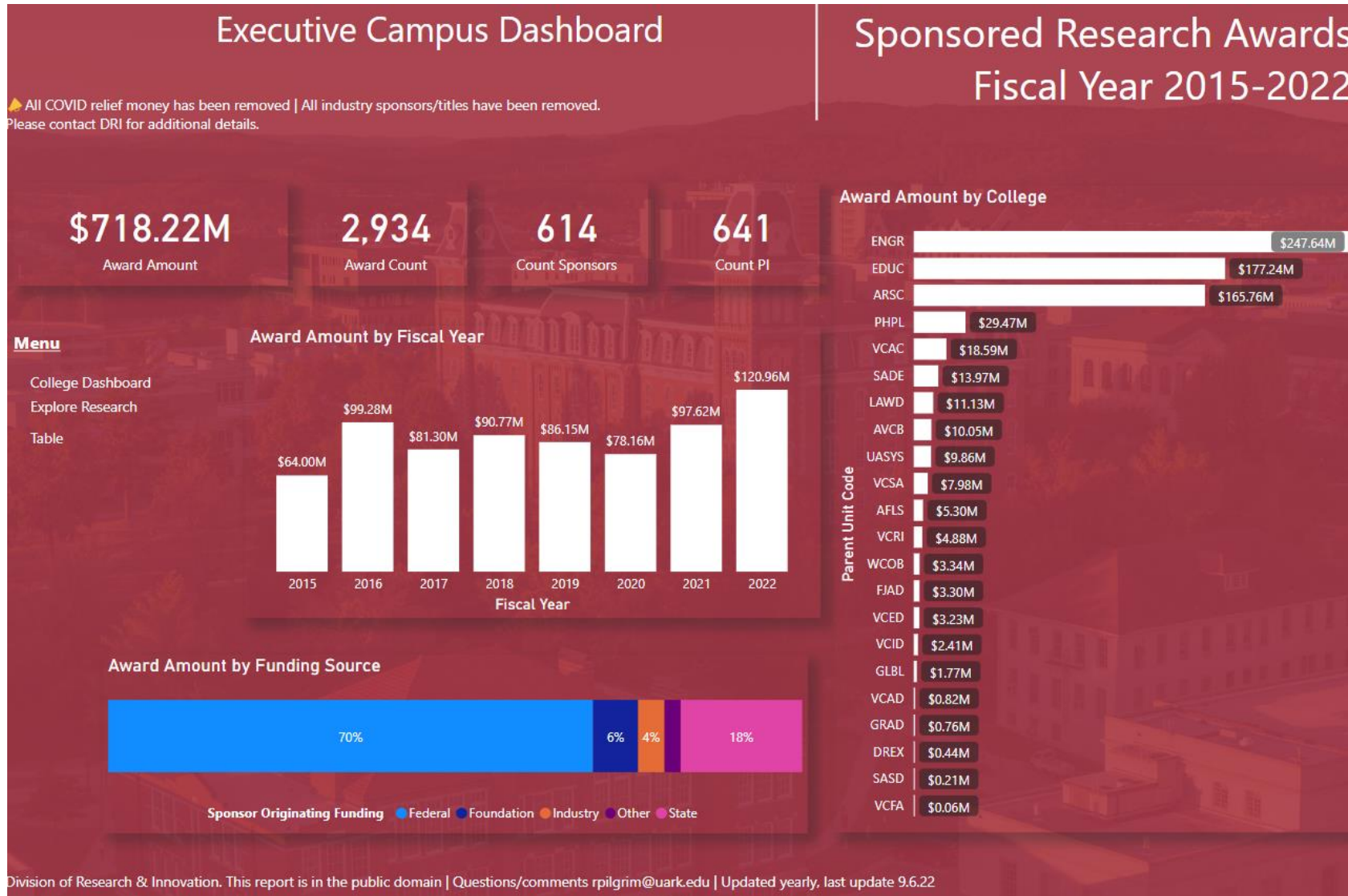
Step 2

Refresh from within your report

These will pull the date from step 1 above into your report



Example Power BI Report



\$718.22M

Award Amount

2,934

Award Count

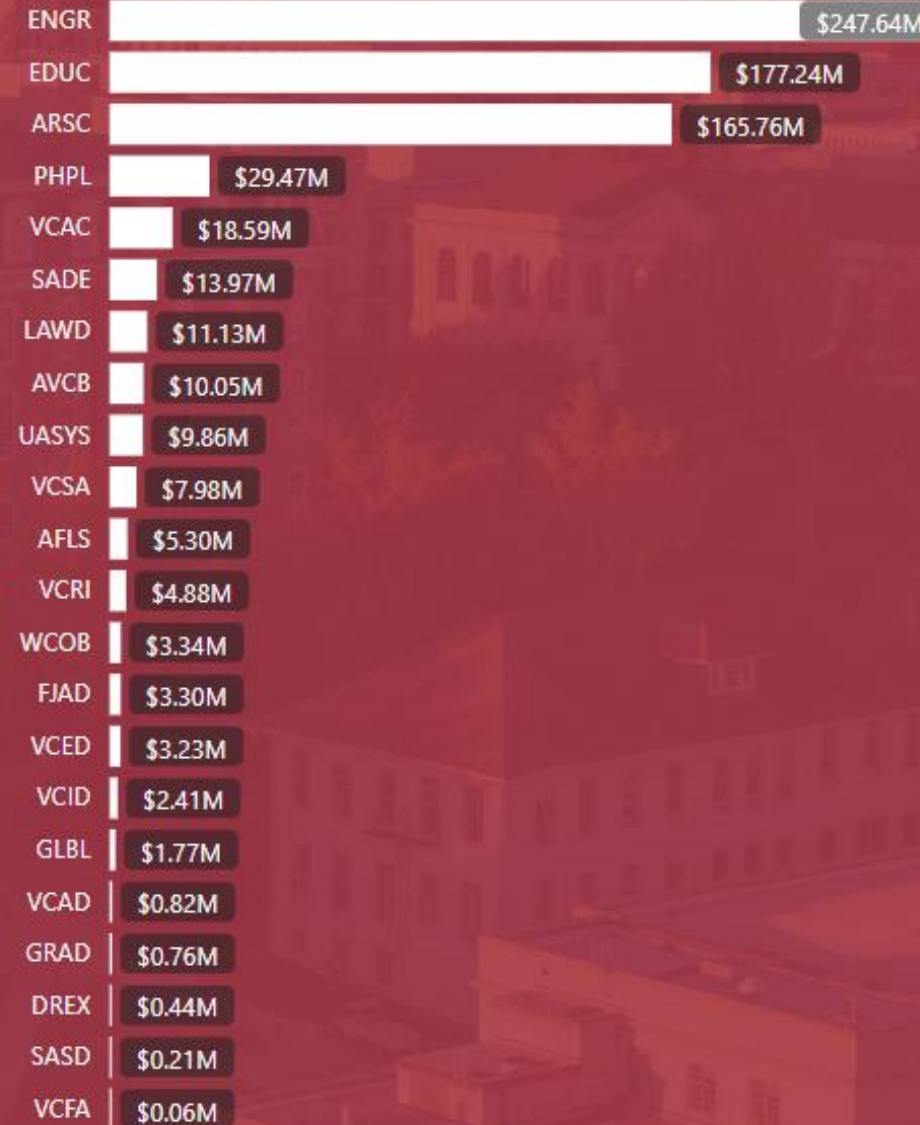
614

Count Sponsors

641

Count PI

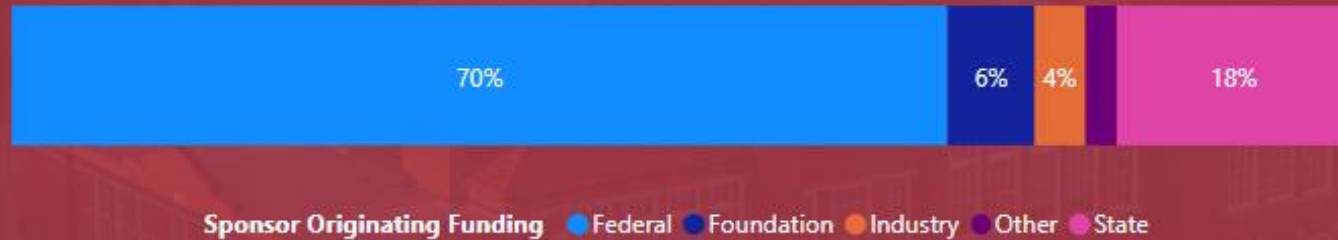
Award Amount by College



Award Amount by Fiscal Year



Award Amount by Funding Source



Menu

College Dashboard

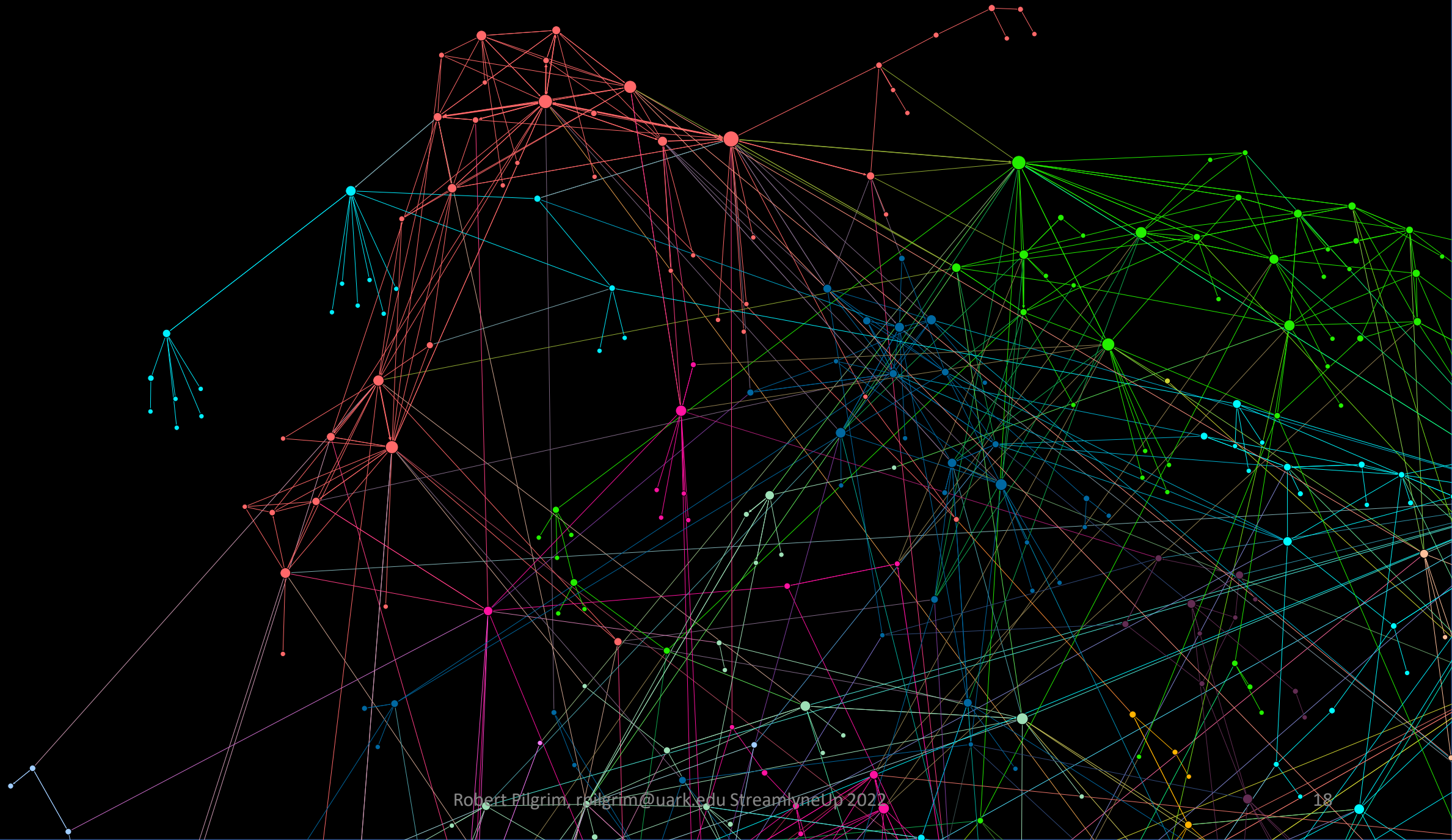
Explore Research

Table

Research Collaboration

“Collaboration is a combined effort to address a scientific challenge that leverages the strengths and expertise of professionals trained in different fields”





Research Collaboration

Overview

- Visualizing your research clusters.
- A few basic definitions.
- How can this help you understand your research growth.
- How can you use this in research development.
- How can research networks be useful to other areas outside of sponsored programs.

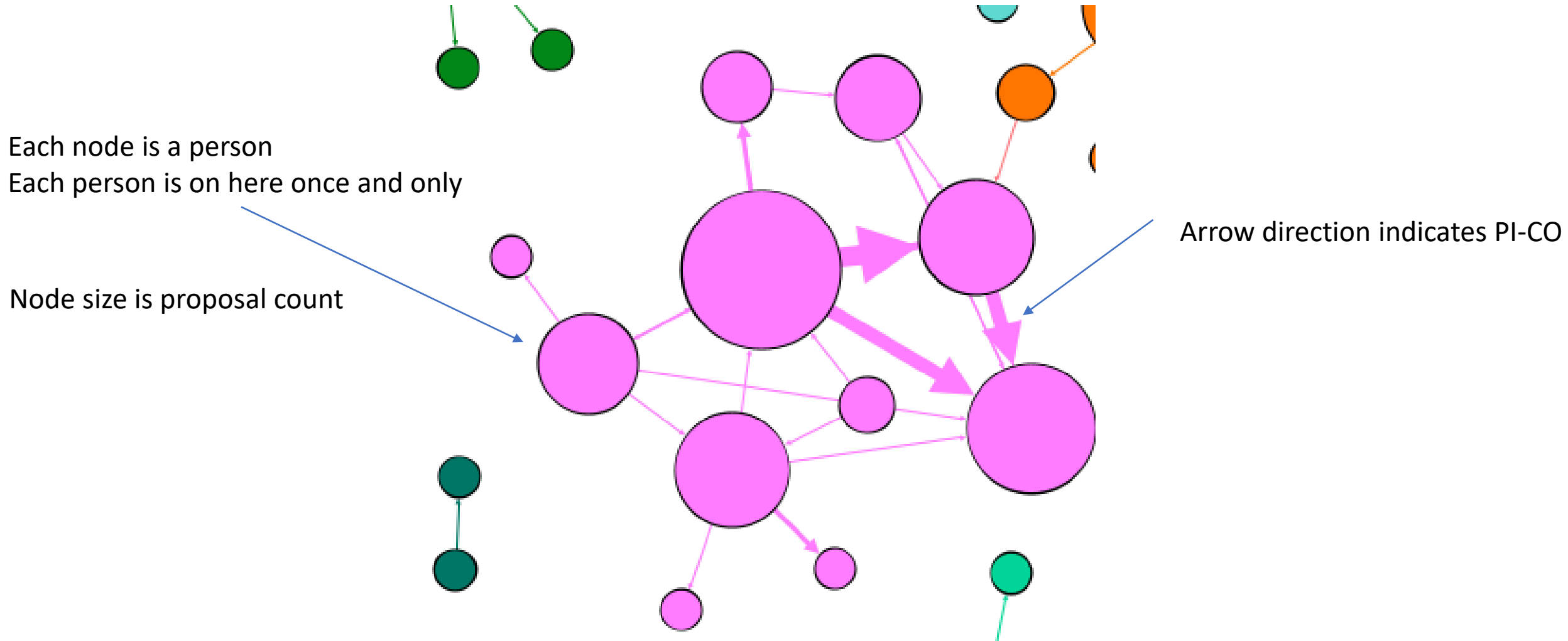
Network Basics

3 simple concepts that you can apply to your job today.

(for more info google 'Social Network Analytics')



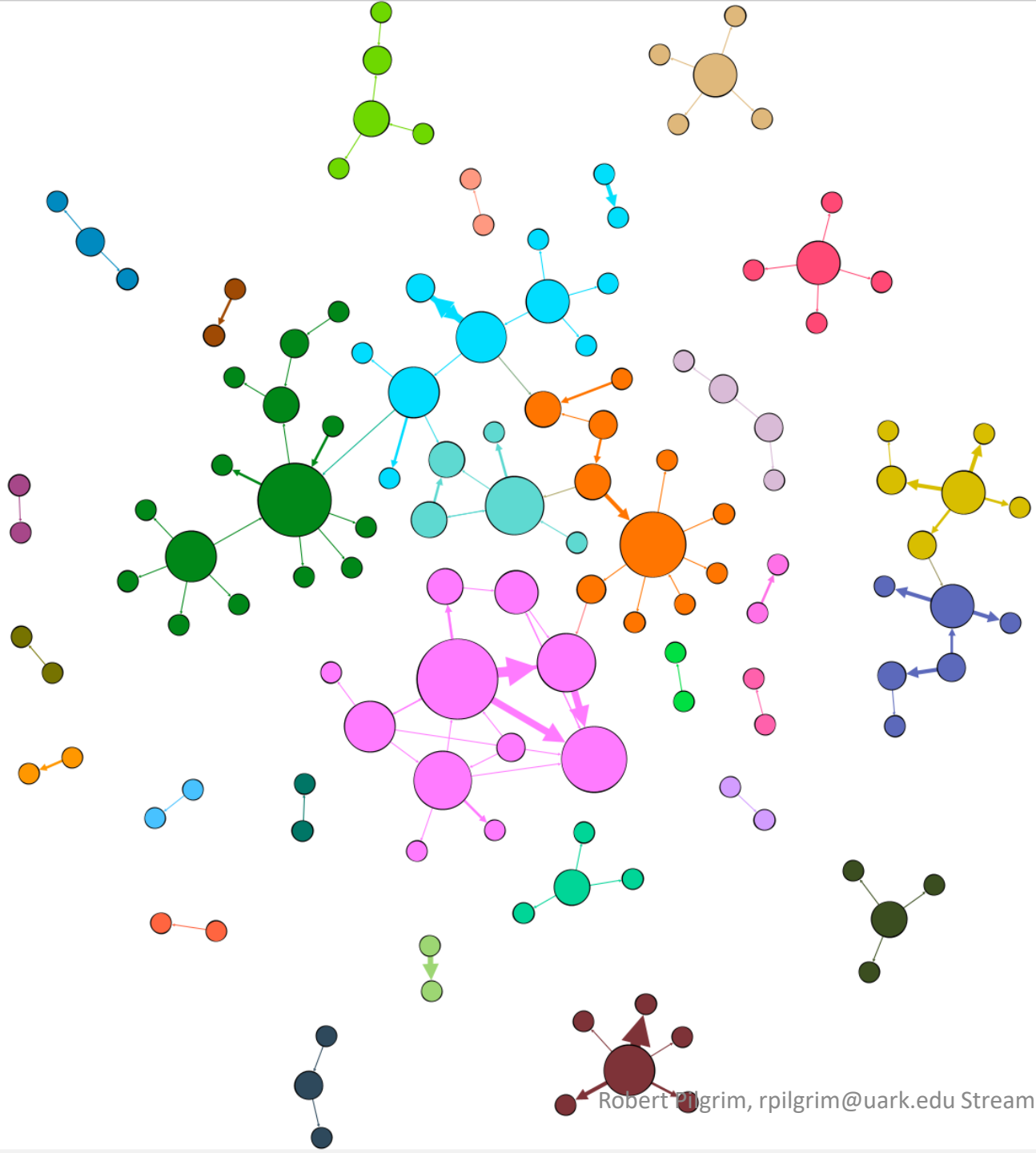
Network Basics



Color represents probable research cluster detected by the algorithm

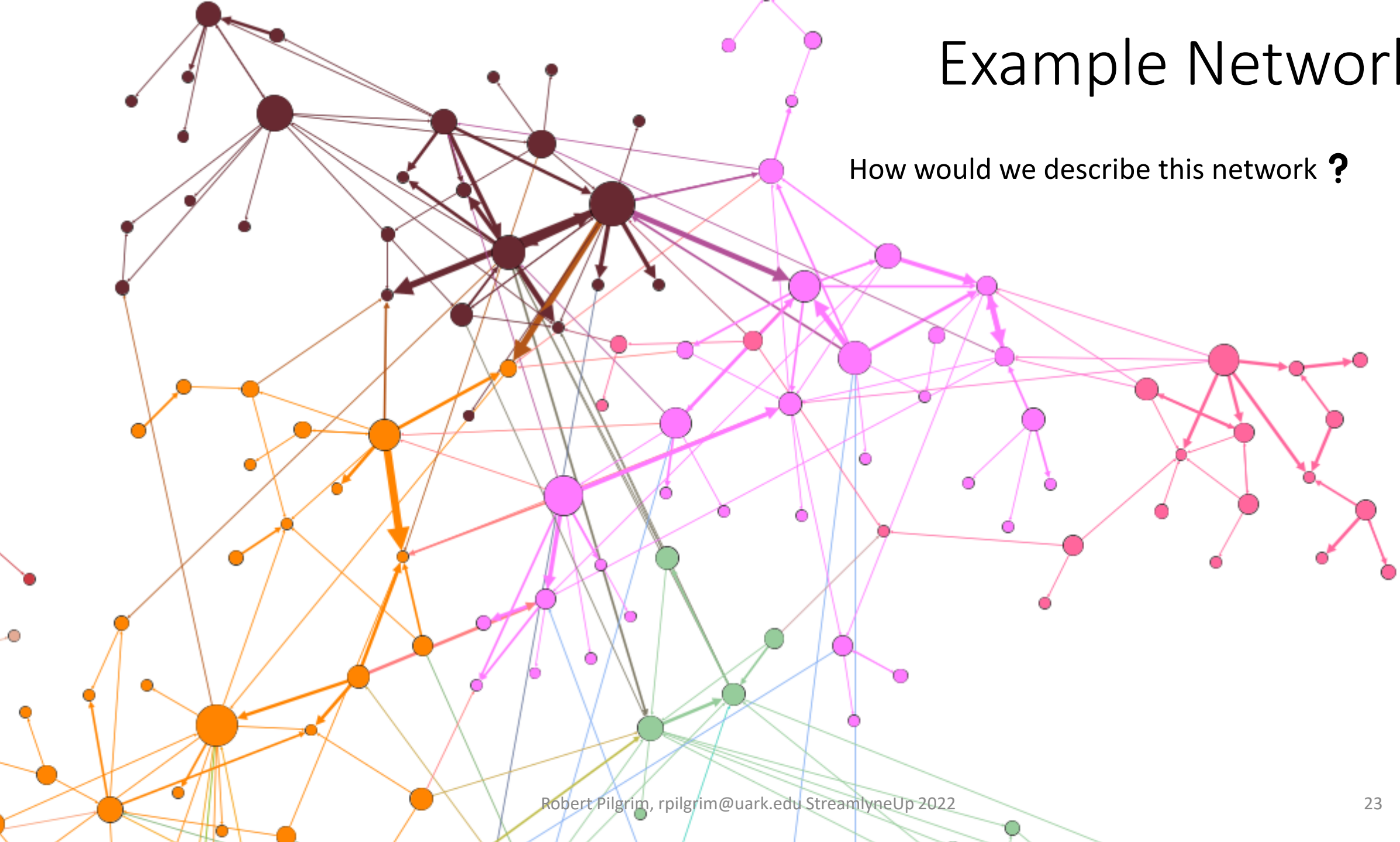
Example Network 1

How would we describe this network ?

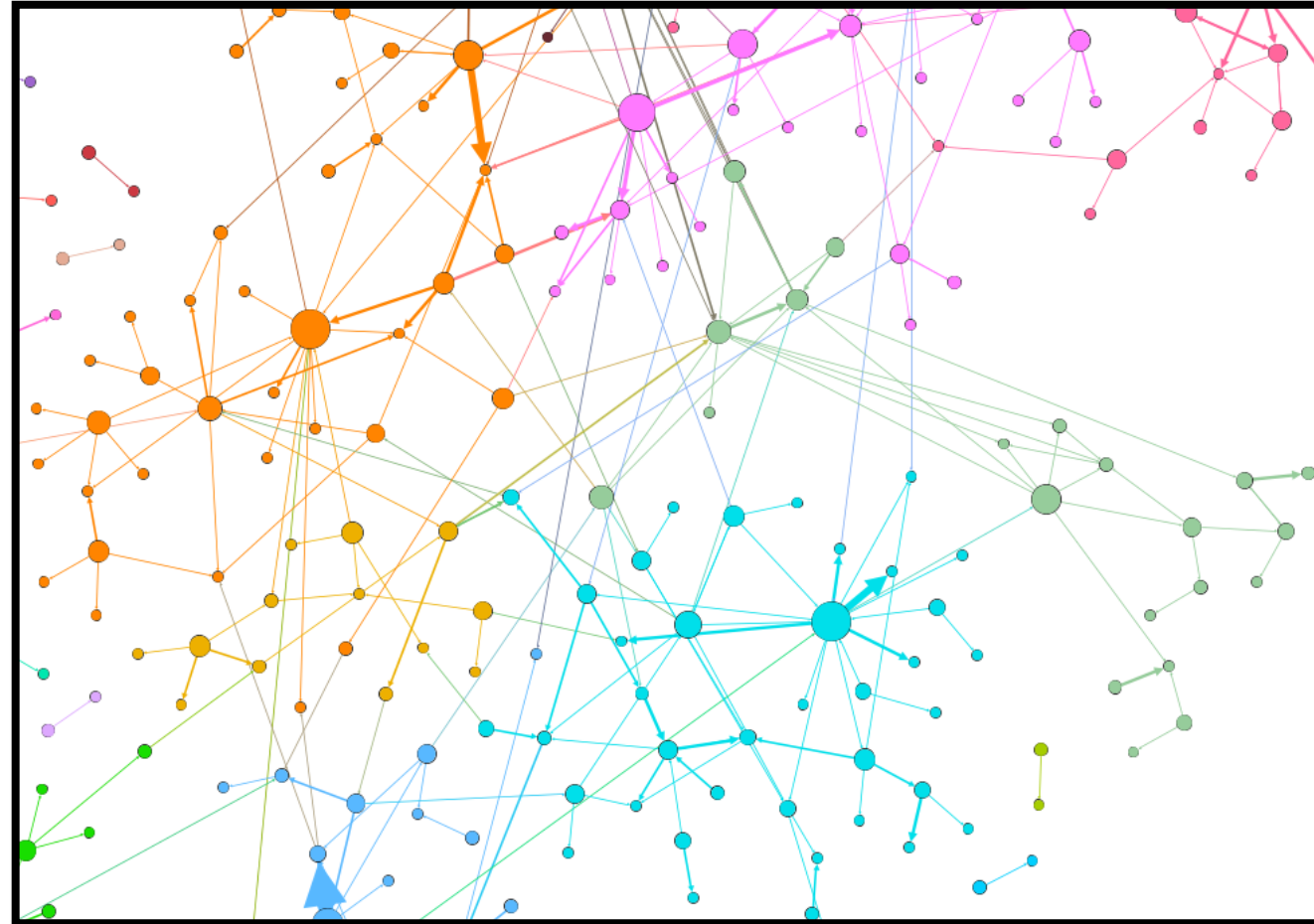
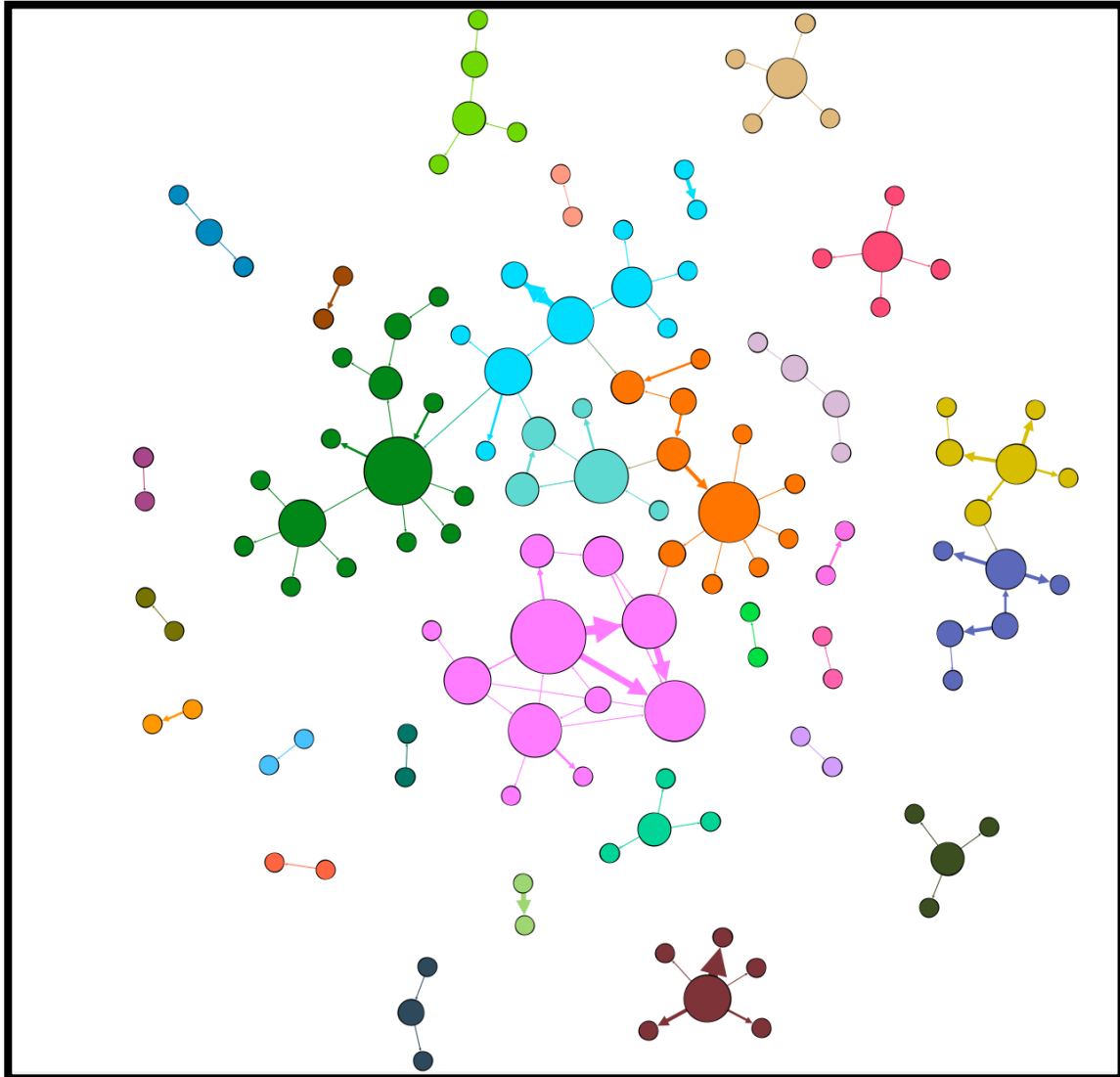


Example Network 2

How would we describe this network ?



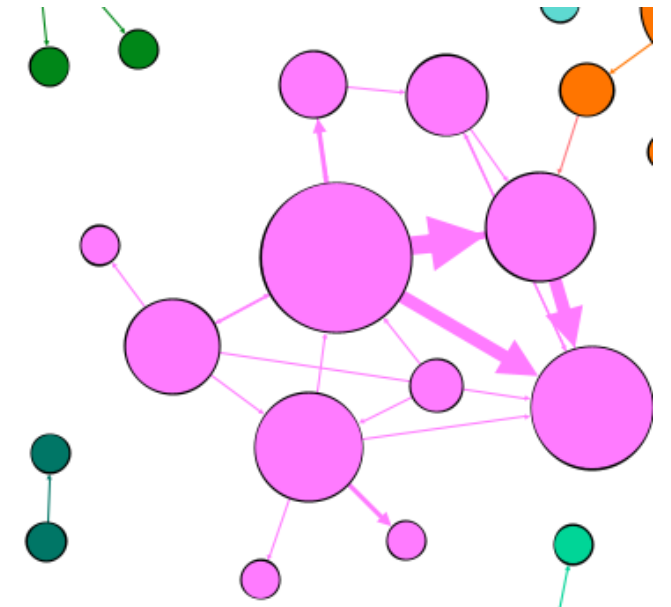
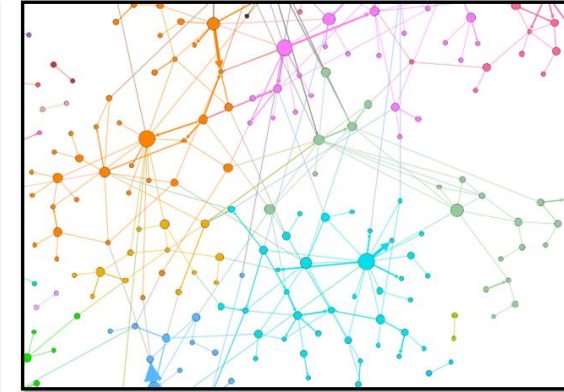
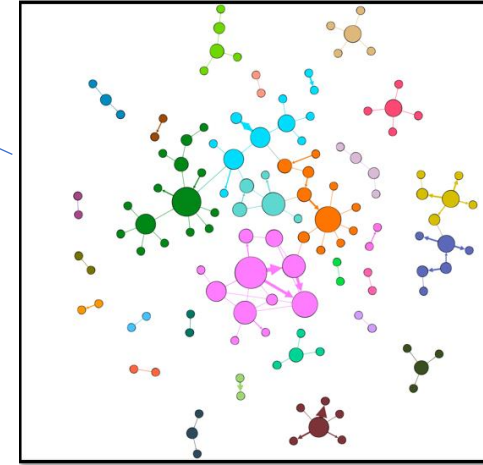
Which Graphic is more Collaborative



Questions?

Some Metrics

- **Graph Density** | How connected are the various clusters (See previous slide)
- **Degree** | How many people connect to a given person
- **Betweenness Centrality** (See next slide)



Betweenness centrality - Wikipedia

In graph theory, **betweenness centrality** (or "**betweenness centrality**") is a measure of centrality in a graph based on shortest paths.

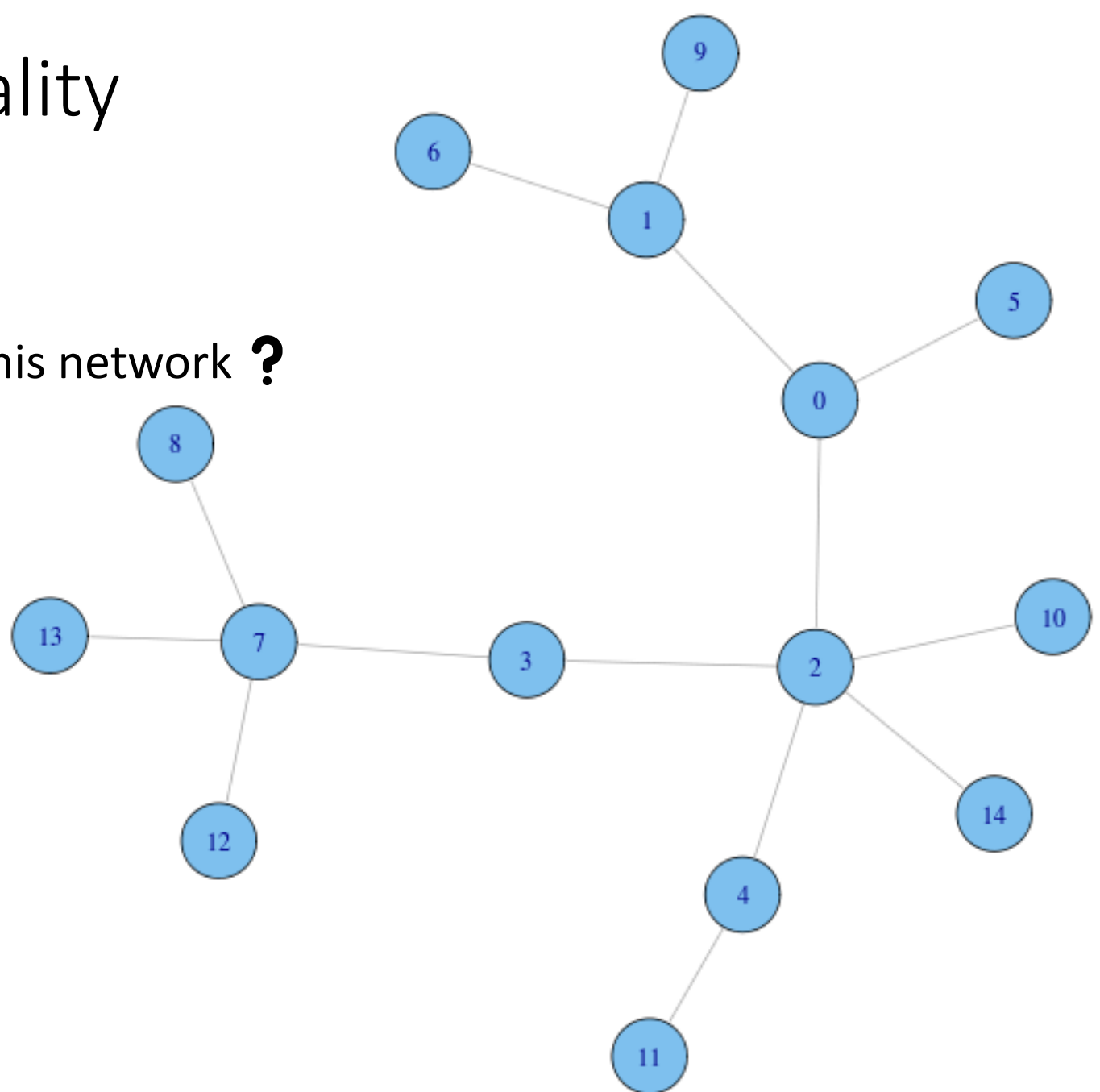
Betweenness Centrality

Simplified Example

Who is the most “Important” person in this network ?

Each node represents a researcher

Lines represent collaboration



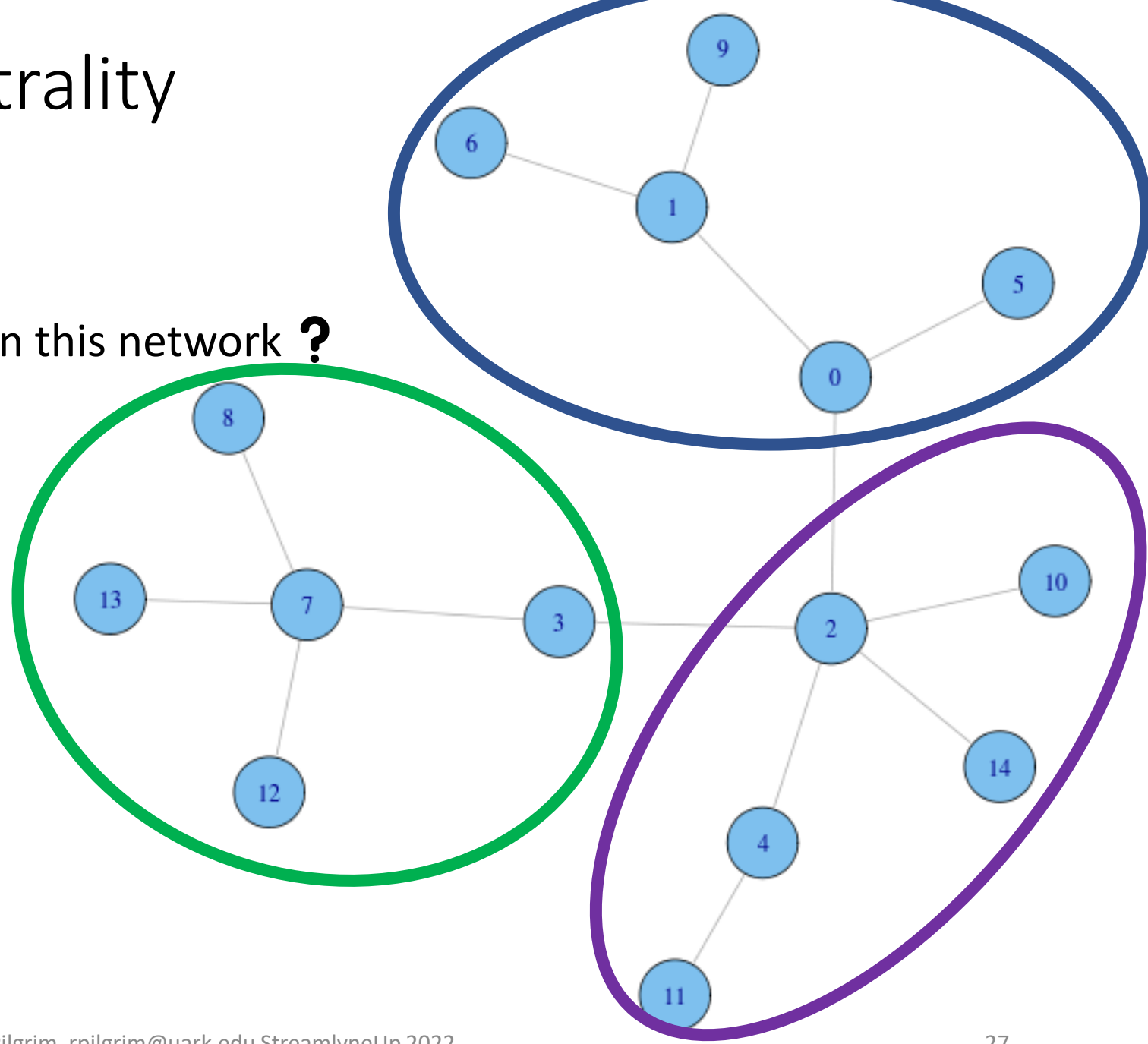
Betweenness Centrality

Simplified Example

Who is the most “Important” person in this network ?

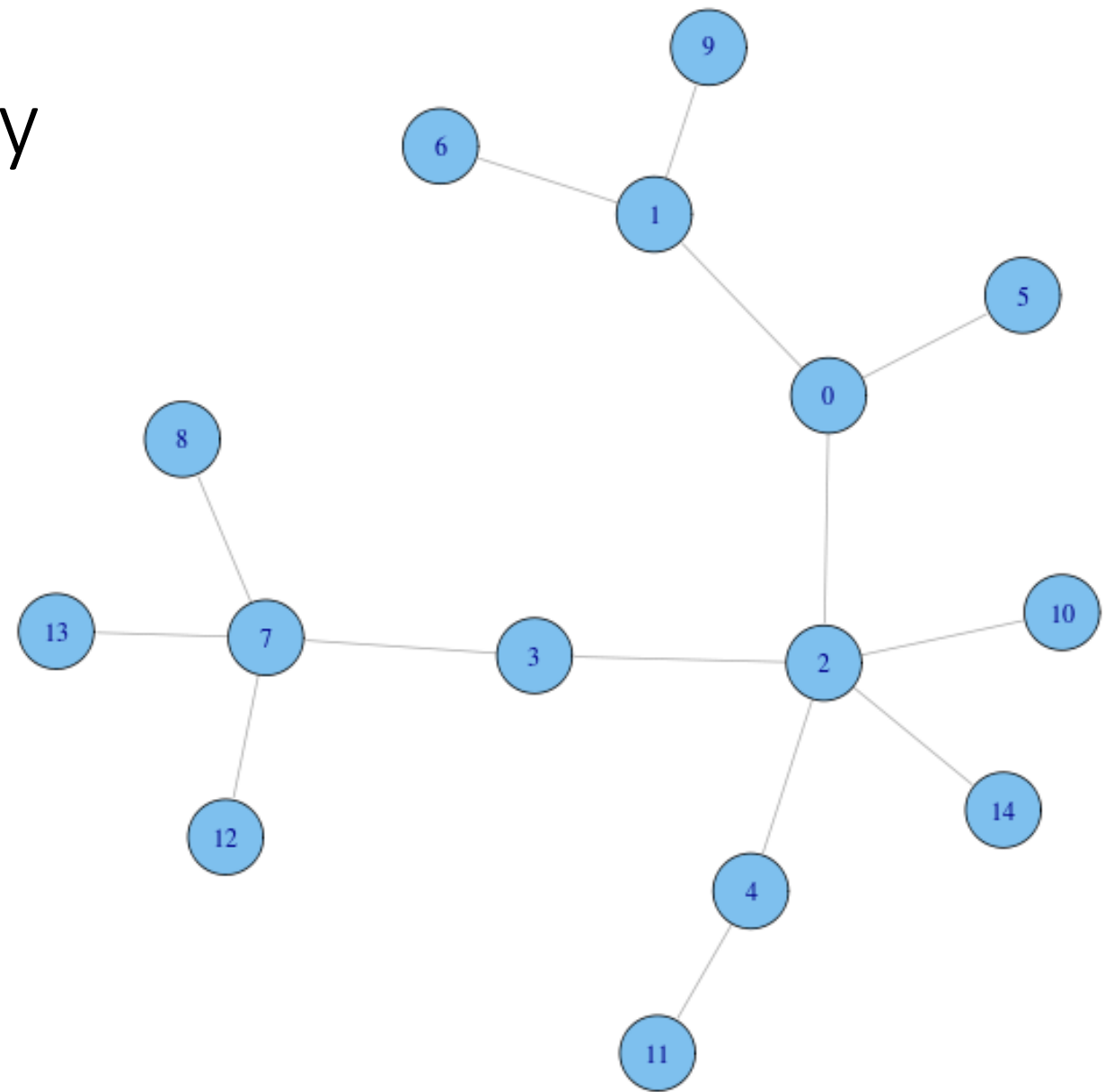
Each node represents a researcher

Lines represent collaboration



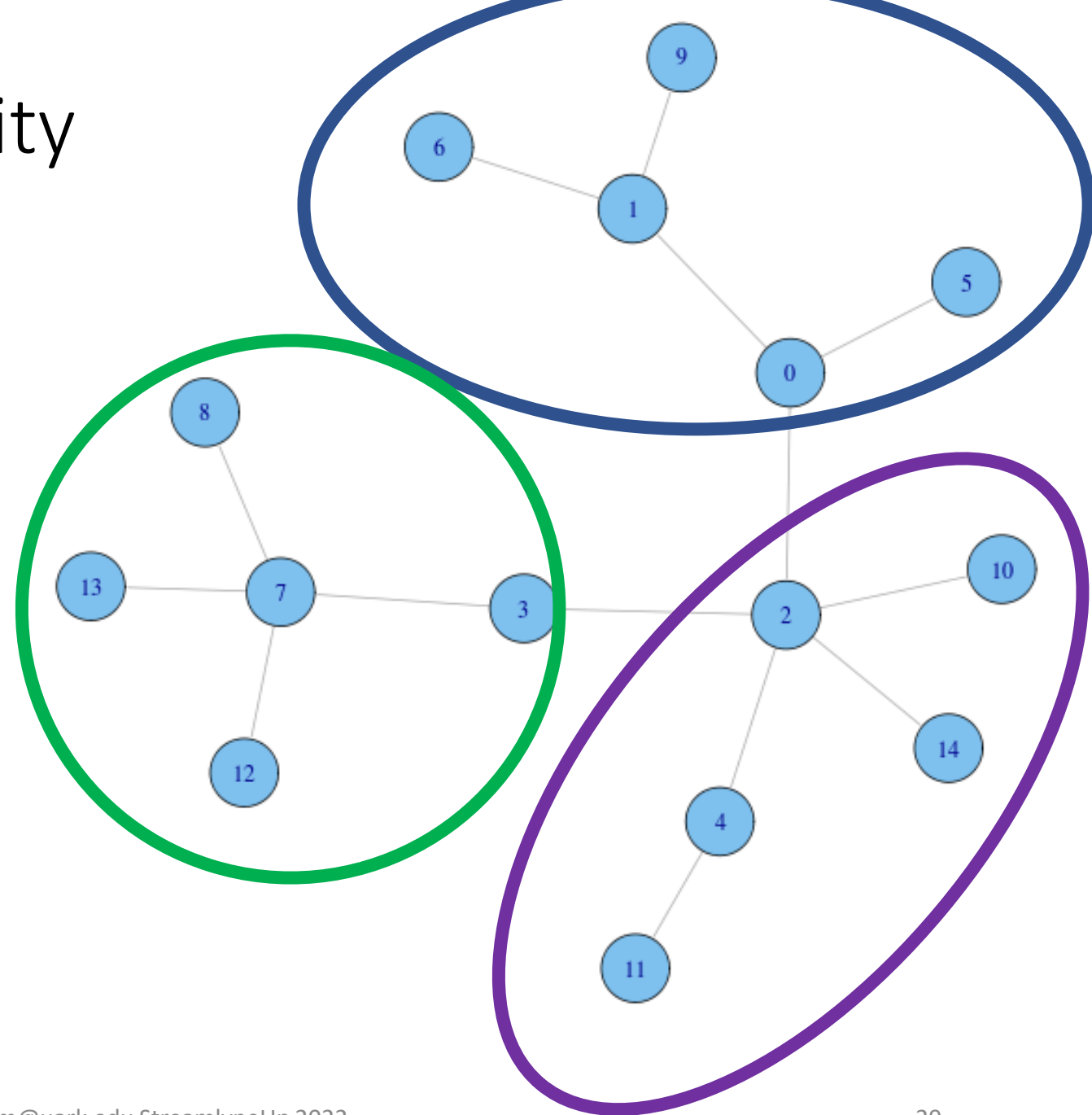
Betweenness Centrality

- How many research clusters are here ?
- Based on the shortest path between all the nodes in a network
- How important is a node (Person) to the flow of information.
- Why should you care about this metric ?
- Experience sponsored programs staff intuitively know the “important” researchers or those who are very collaborative.



Betweenness Centrality

- How many research clusters are here ?
- Based on the shortest path between all the nodes in a network
- How important is a node (Person) to the flow of information.
- Why should you care about this metric ?
- Experience sponsored programs staff intuitively know the “important” researchers or those who are very collaborative.



Betweenness Centrality

The higher the value the more “Important” the node

Definition [\[edit\]](#)

The betweenness centrality of a node v is given by the expression:

$$g(v) = \sum_{s \neq v \neq t} \frac{\sigma_{st}(v)}{\sigma_{st}}$$

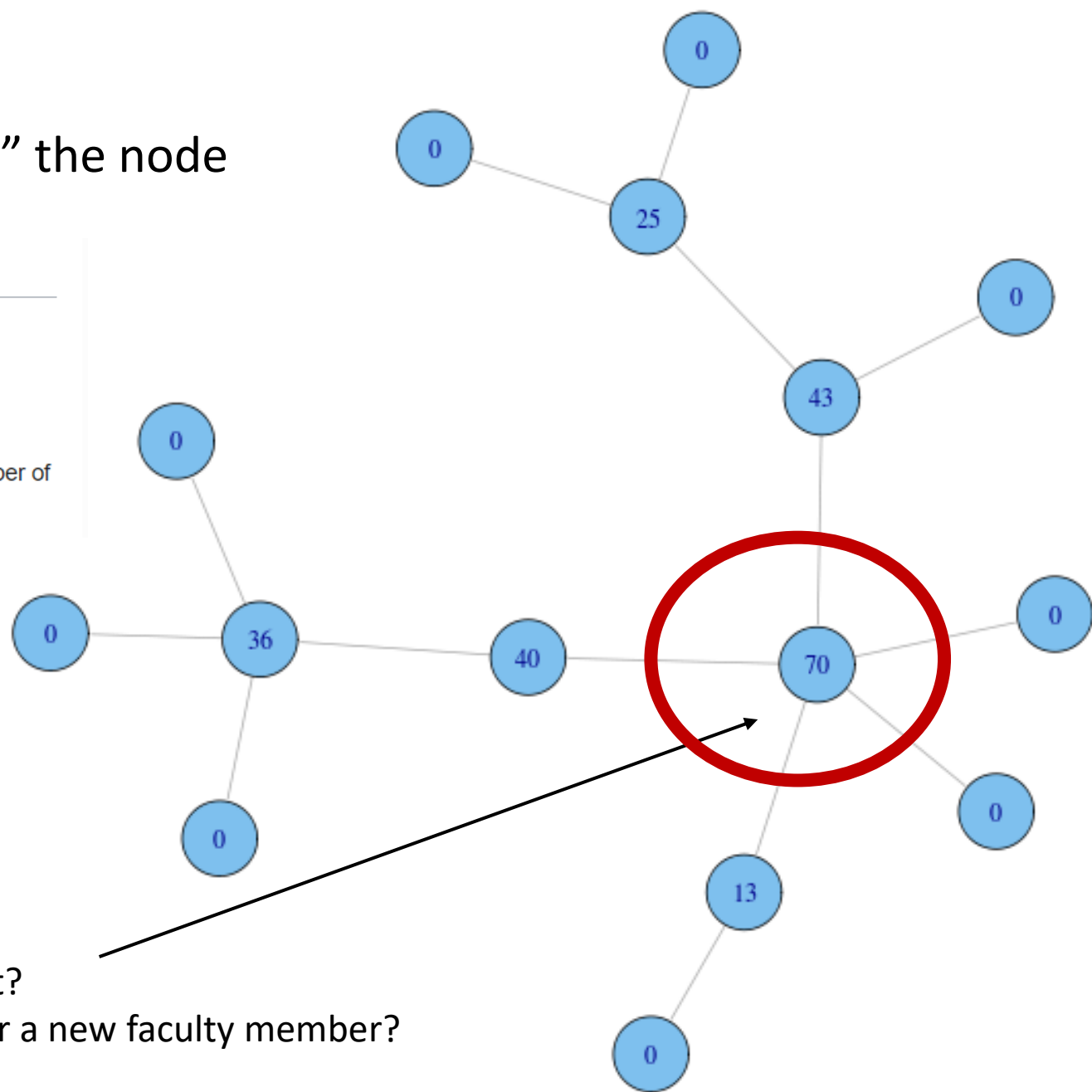
where σ_{st} is the total number of shortest paths from node s to node t and $\sigma_{st}(v)$ is the number of those paths that pass through v (not where v is an end point).^[2]

Detailed statistics is outside the scope of this talk

What could we say about this researcher?

What would happen to our teams if they left?

Who would be a good person to be a mentor a new faculty member?



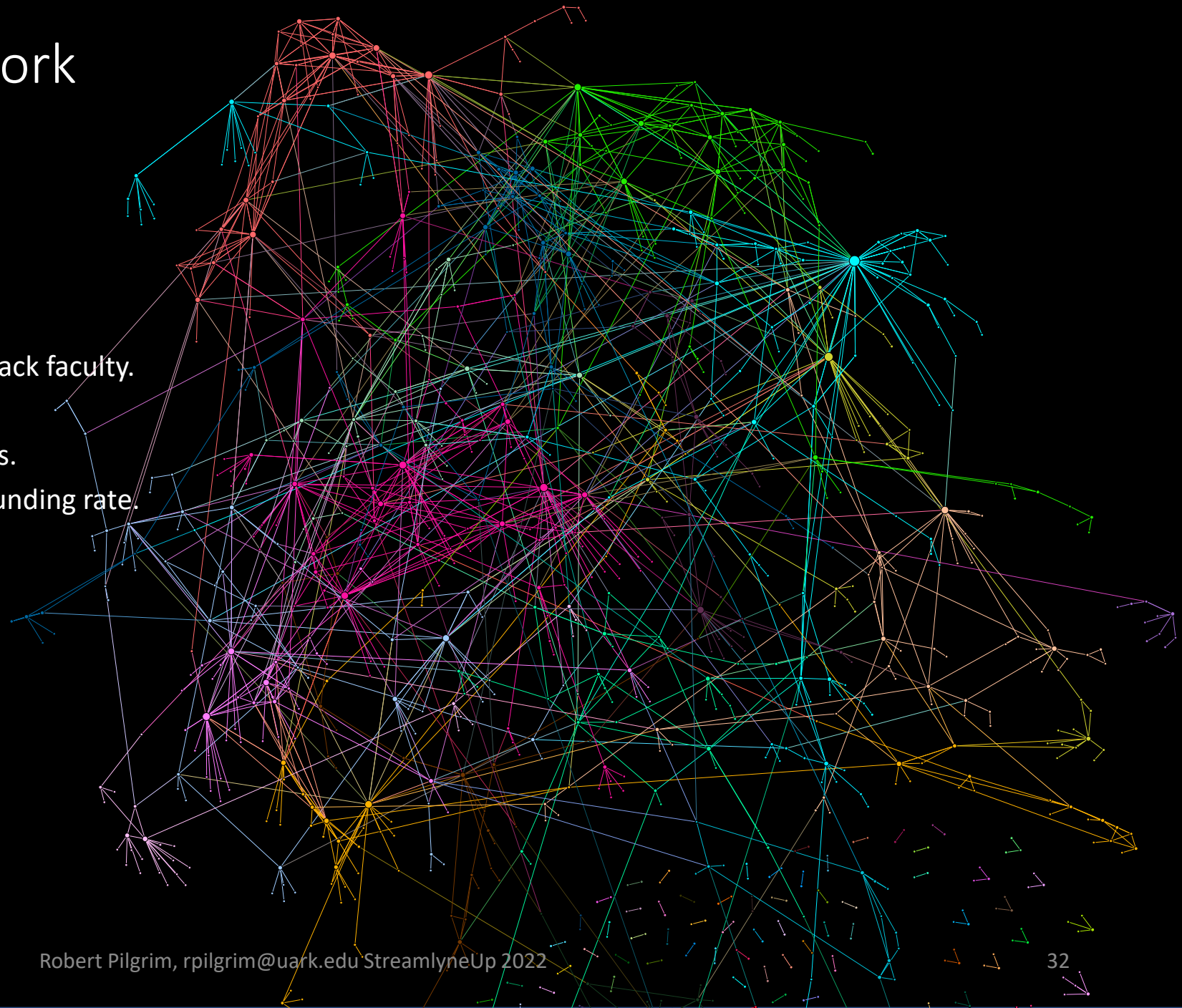
In this room who has the Highest Betweenness Centrality

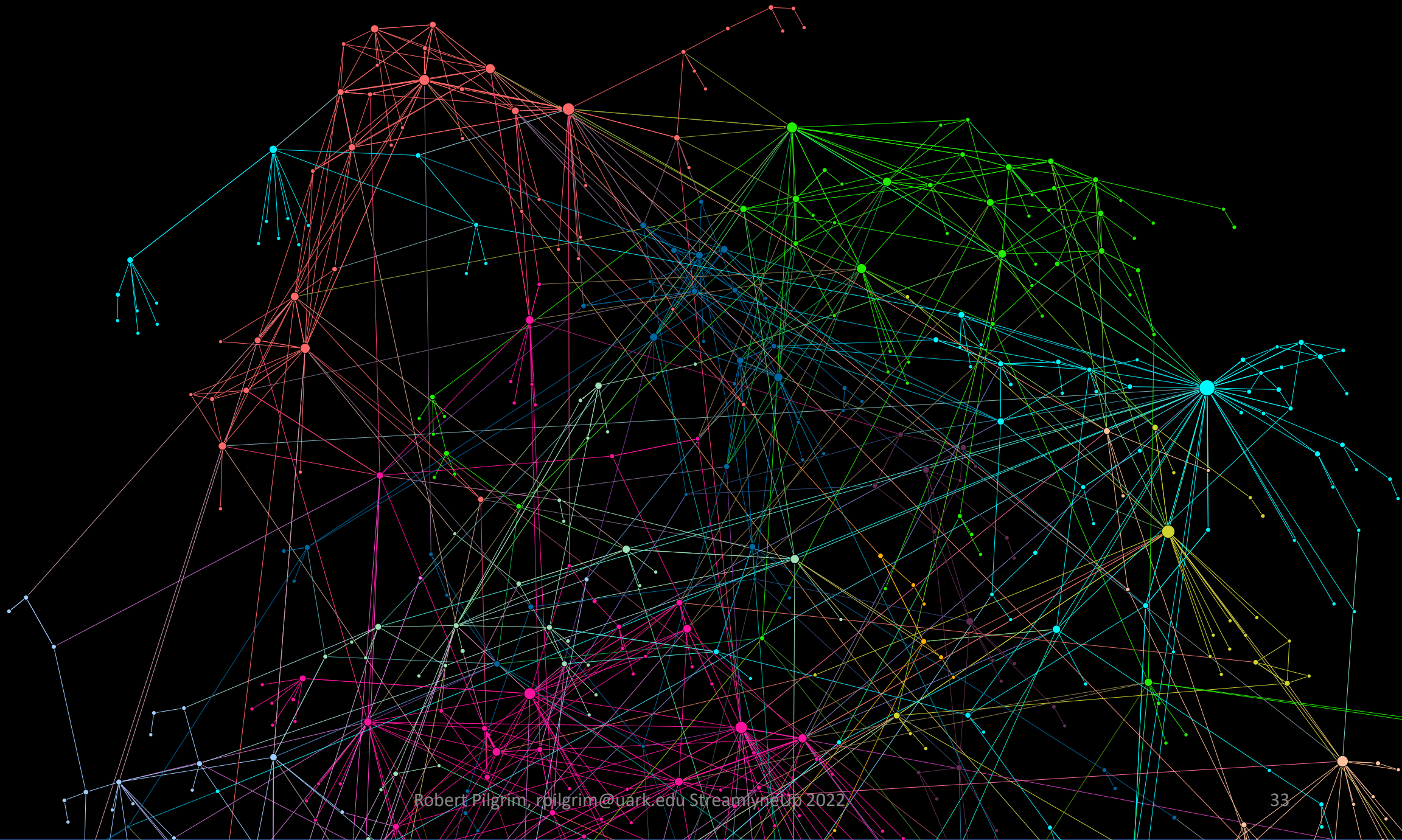
If we were to plot a network of who knows who in this room, who would have the highest betweenness centrality ?



Example Campus Network

- Visualization of your research landscape
- Additional perspectives (Filters)
 - Sponsors | Show me NSF network.
 - College/Unit | Show me Engineering.
 - Tenure Status | Show me my Tenure Track faculty.
 - Custom Groups
 - Those who received startup funds.
 - Those with a very low proposal funding rate.
 - Those in a signature area.
 - Expenditures on large equipment
 - International Travel
 - Funding Graduate Students





Back to the Strategic Plan

You can now quantify collaborative

Take yearly snapshots and see the change over time

- Is collaboration increasing at your institution?
- Which researchers are driving the increase in collaborations?
- How are your cluster hires/ interdisciplinary programs changing the collaboration dynamic on your campus?
- A few other applications



Research Networks | Diversity Perspective

- Use your research networks and break them down by gender.
- Is there any statistical difference in the network metrics?
- Can we explain these differences?
- Research data, from a diversity perspective, allows sponsored programs to branch out into other units on campus such as the Office of Diversity and Inclusion .



Research Networks | Tenure Perspective

- Research shows the more collaborative a tenure track faculty member the greater the chance that they will make it to full tenure.

Research Publications

- This data is many orders of magnitude larger.
- Visualization is not very useful.
- You must parse the data and/or rely on statistical analyses

Where can I get publication data

- Academic Analytics (Download Portal)
- Digital Measures
- Web of Science bulk export
- Your campus “Faculty Experts” database
- “Vos Viewer” software (free) will visualize the data for you



Quantifying a Strategic Initiative to improve Collaborations

Leveraging your research data.



Quantifying the Research Impact of New Research Building

How would *YOU* quantify the impact on research and collaborations across your campus as a result of a new building ?

- Designed to increase collaborations in your signature areas.
- Proposed new multimillion-dollar research building.
- Attract top tier researchers.

Signature Areas

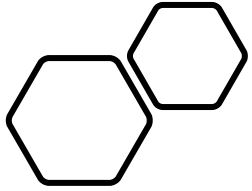
- Food and Technology
 - Data Science
- Materials Science and engineering, Bioscience and Bioengineering
- Research in metabolism Integrative systems neuroscience.



A New Research Building

- \$194.7 million grant from the Walton Family some of which is to build this.
- Institute for Integrative and Innovative Research (**I3R**) building.
- Target completion in 2024





Leveraging Research Data to Help Build Alignment

Case Study

- Leadership has a list of researchers they want to include in the initial discussion on this new initiative.
- They want to understand how these people are involved across the various research groups to ensure cross campus representation.
- How would you do this using your research data ?



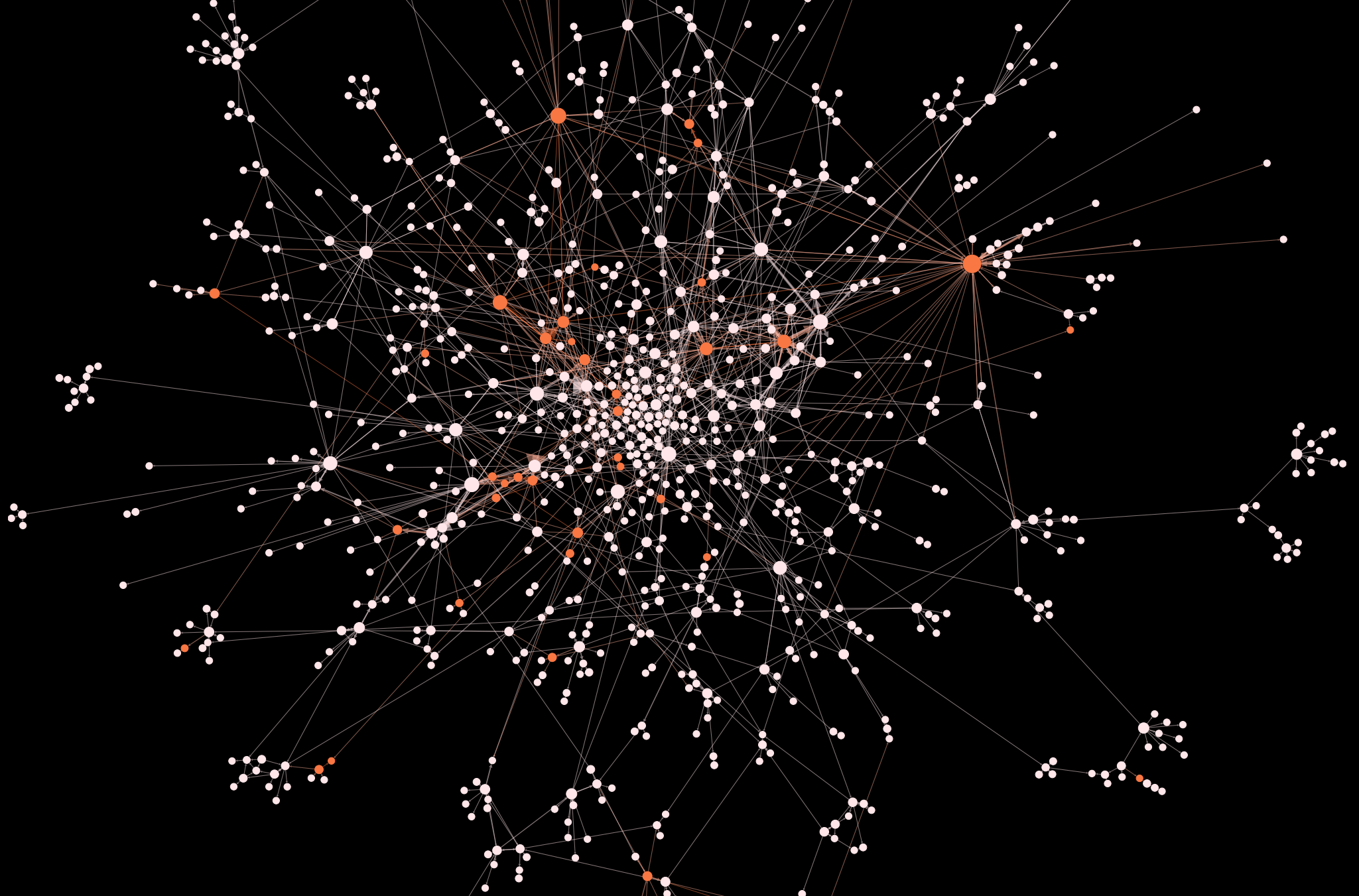
Visualizing the key Researchers in each Signature Area



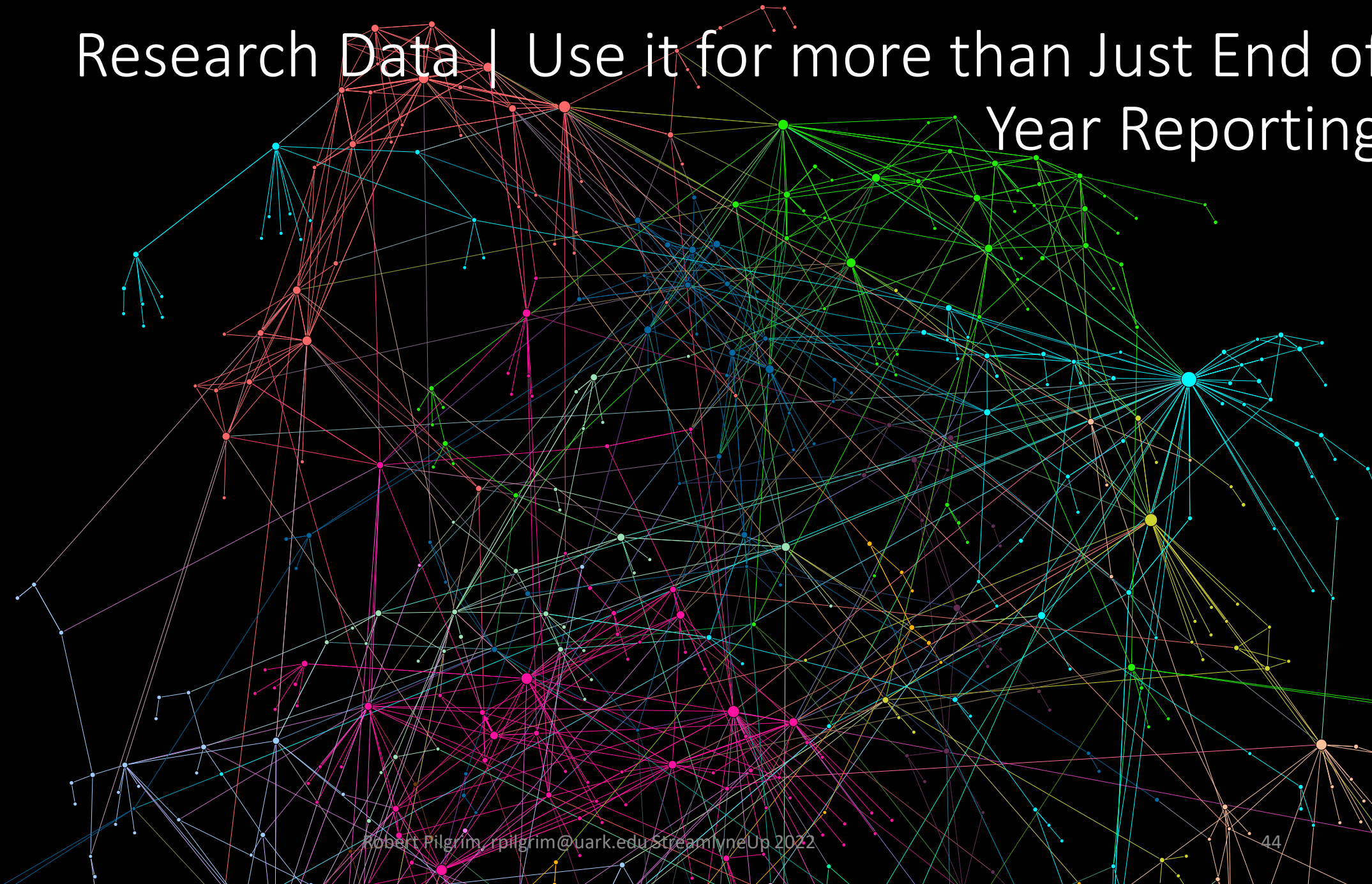
- Use historic **research data** to find researchers that are in the research areas of interest.
- Set the node (faculty) color to be based on the researcher's membership in the research areas
- This gives a visual representation of the distribution of these researchers.

Next slide is a larger version

Previous slides color represented research cluster, now it represents if a faculty is in a signature area.
Here we just pick the top 5 faculty



Research Data | Use it for more than Just End of Year Reporting



Interactive Research Networks

Research Networks

All non-research related COVID relief funding has been removed.

Awards Updated 9/21/2022
Proposals Updated 9/21/2022

2014 2023



Parent Unit Name
Engineering

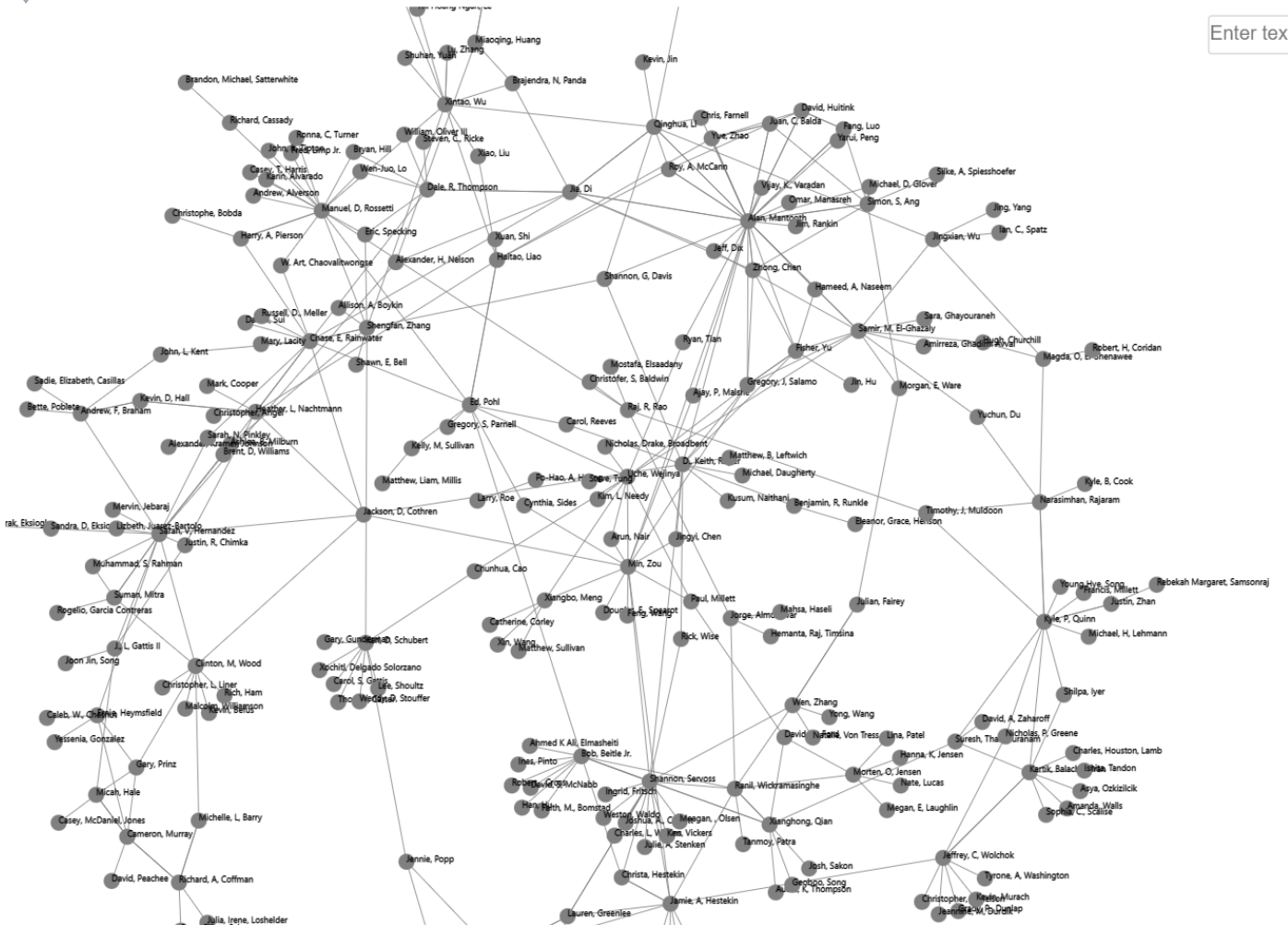
Lead Unit Name
All

Sponsor Originating Funding Type Group
All

Sponsor Originating
All

PI Name
All

PI -> COI



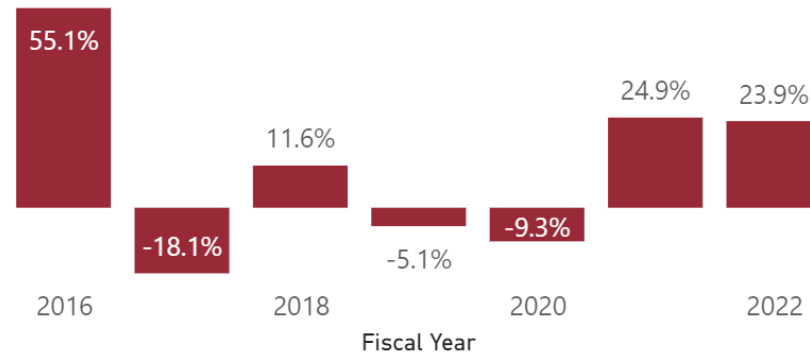
Enter text filter

Time Intelligence Calculations

Year-over-year Comparisons for a department

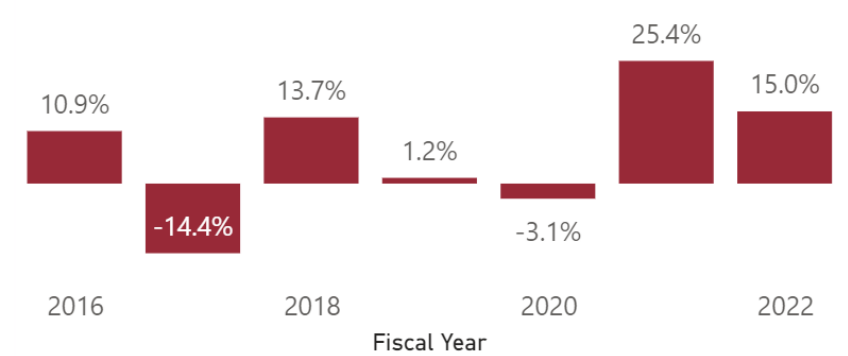
Award Amount

● pct Change from Previous Year



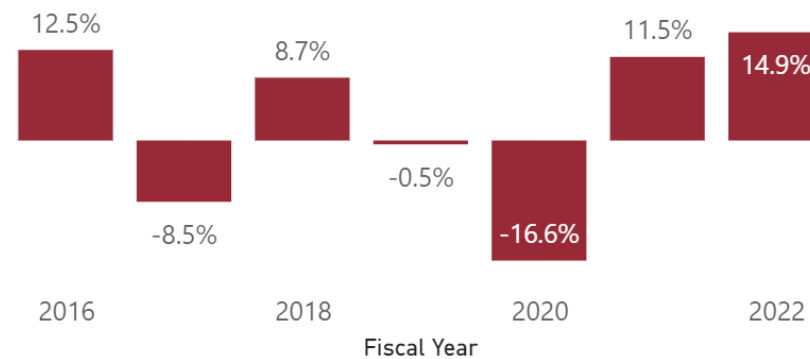
Award Count

● pct Change from Previous Year



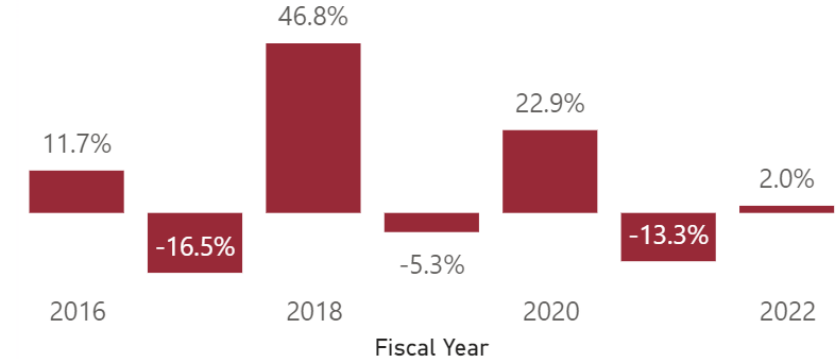
Sponsor Count

● pct Change from Previous Year



% 2 or more Investigators

● pct Change from Previous Year



Example Rolling Averages

Rolling Averages help “stabilize” data
Proposal submission counts

Proposal Count by Date and Time Intel

Time Intel ● Current Value



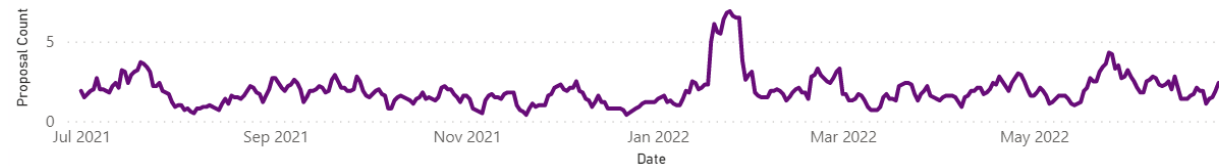
Proposal Count by Date and Time Intel

Time Intel ● Rolling avg 5 Day



Proposal Count by Date and Time Intel

Time Intel ● Rolling avg 10 Day



Proposal Count by Date and Time Intel

Time Intel ● Rolling avg 20 Day



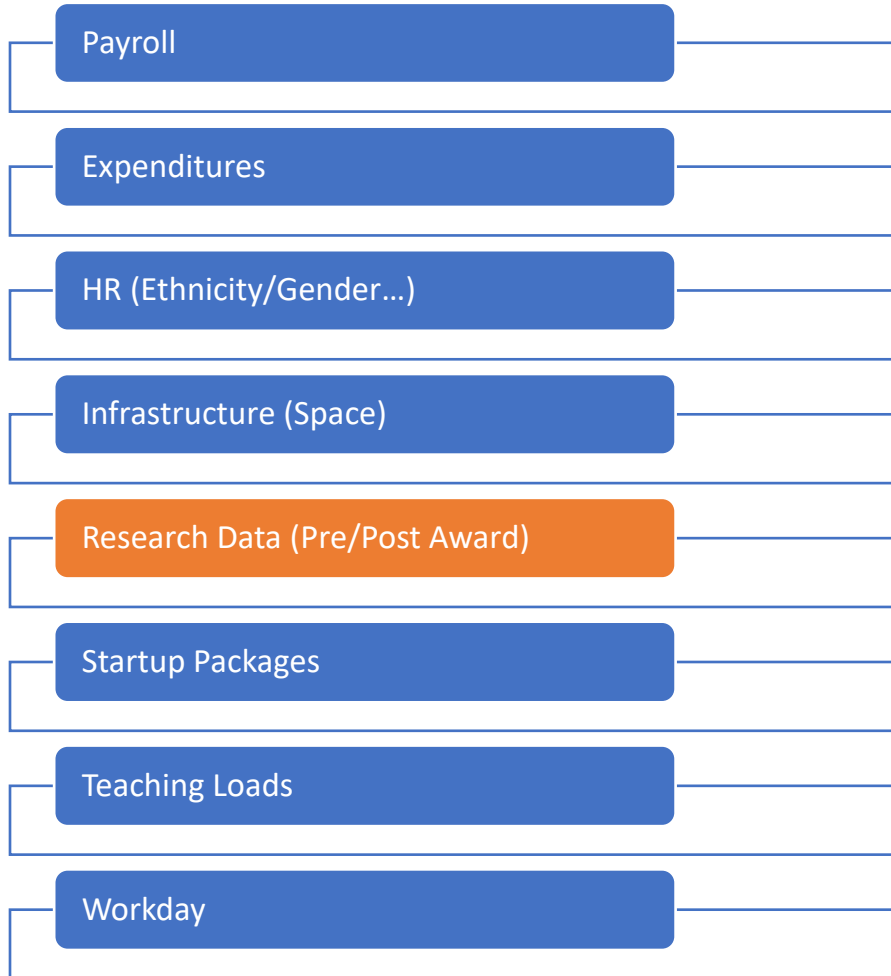
Research Analytics Data Model

- What is a data model?
 - Bring multiple data sources together into a single system
 - **Cross data source calculations**
 - Campus, college, department and faculty level

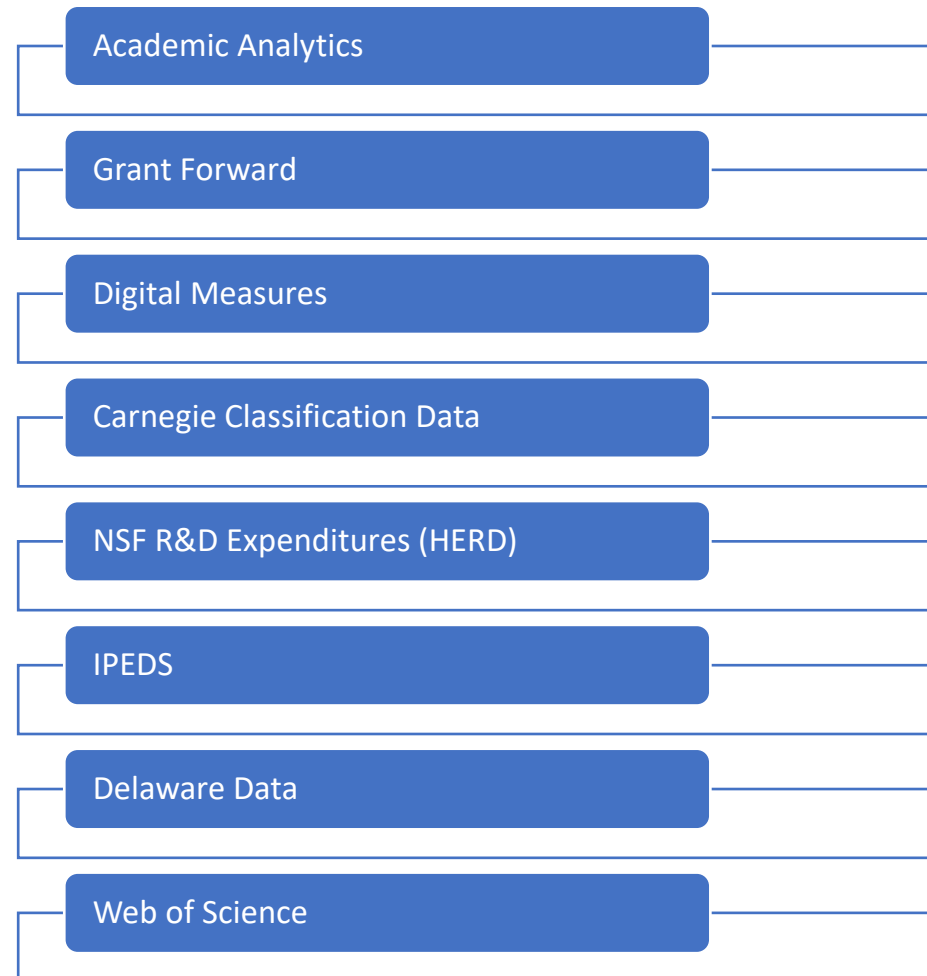


Data Model | Cross Data Analytics

Internal Data



External Data



You have your data model, what next?

- Descriptive Analytics
 - **What happened last Fiscal Year.**
 - Too late to do anything about it, let's hope it is better next year.

- Prescriptive
 - **What caused the changes you are measuring.**
- Predictive
 - **What is likely to happen.**
 - Are we doing enough right now to ensure targets are met.

Machine Learning



Leveraging Your Research Data | Quantifying Research Collaborations

Alice: "Would you tell me, please, which way I ought to go from here?"

Cat: "That depends a good deal on where you want to get to."

Alice: "I don't much care where."

Cat: "Then it doesn't much matter which way you go."

Alice: "...so long as I get somewhere."

Cat: "Oh, you're sure to do that, if only you walk long enough."

Robert Pilgrim
rpilgrim@uark.edu

Associate Director of Data Strategy & Insights
Division of Research & Innovation
University of Arkansas
StreamlyneUp 2022



- END SLIDES

-

