PROJECT: NYU TANDON FACULTY CONNECTIONS

Sanyukta Tuti (st5442), Kaartikeya Panjwani (kp3291), Dhairyasheel Patil (dp3979) *CS-GY 6313 B*

1 Abstract

Our project, "NYU Tandon Faculty Connections," aims to develop an interactive visualization that maps the interdisciplinary network of faculty members at NYU Tandon School of Engineering. By leveraging a chord diagram, we provide an intuitive way to explore faculty members' research interests, teaching areas, and departmental affiliations. This visualization helps students identify potential thesis supervisors, researchers discover collaboration opportunities, and administrators analyze departmental strengths and gaps. Using D3.js for interactivity and a curated dataset of faculty details, our project bridges the gap between static directories and dynamic academic exploration.

2 Introduction

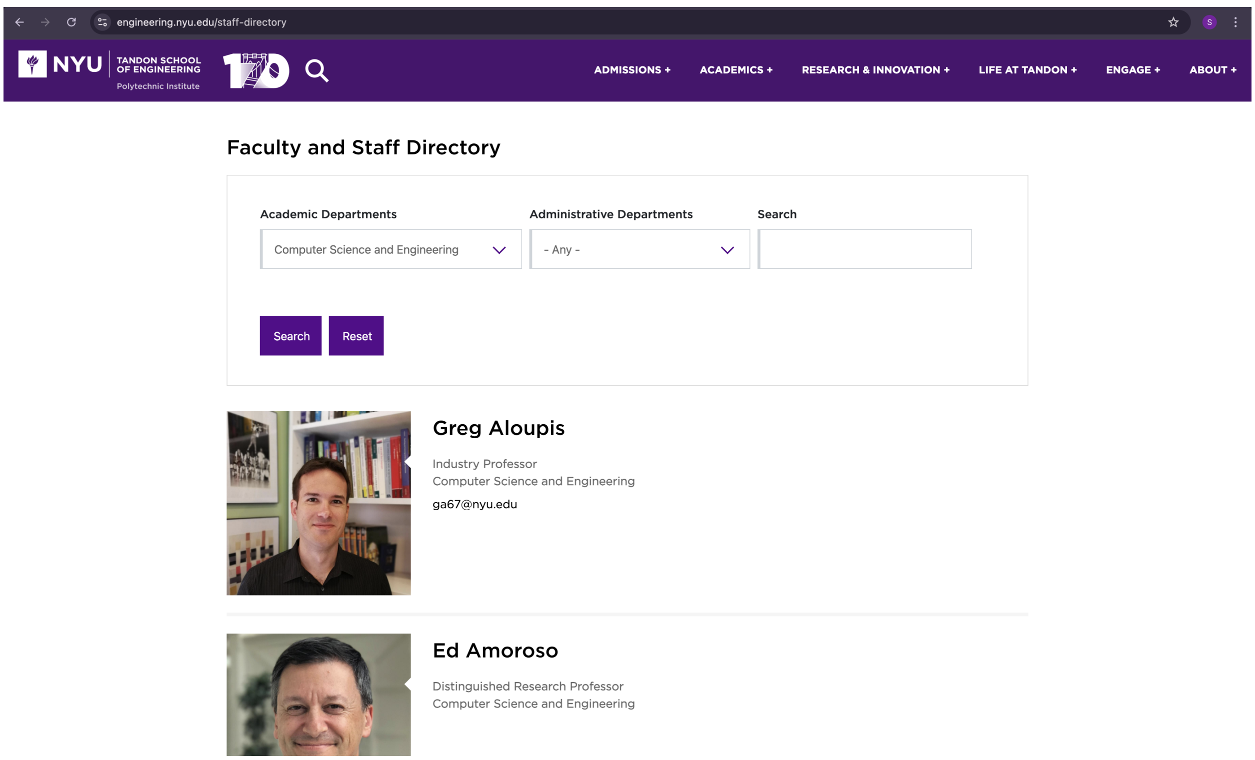

Figure 1: NYU Tandon School Of Engineering Faculty and Staff Directory[3]

Faculty directories are critical resources in academic institutions, but they often lack an intuitive design that enables efficient exploration. Our project addresses this gap by creating a visualization tool that highlights interdisciplinary relationships among NYU Tandon faculty members. This tool caters to diverse audiences:

Students: Simplifies finding faculty aligned with their research or academic goals.

Faculty and Researchers: Encourages interdisciplinary collaborations by visualizing shared interests.

University Administrators: Aids in strategic planning by exposing strengths and gaps in departmental research.

Our key questions included:

- How can we make faculty connections more accessible?

- How can we identify overlaps in research interests across departments?
- What visualization best represents the diversity and interconnection of research at NYU Tandon?

By addressing these, we created an interactive tool that replaces the traditional, static faculty directory with a dynamic visualization [1] [4] that fosters academic collaboration and exploration.

3 Background

Academic institutions like NYU Tandon are hubs of innovation, housing a diverse range of faculty members, research interests, and teaching domains. However, navigating these academic networks can be challenging. Current resources, such as departmental directories or scattered web pages, provide fragmented information, making it difficult for students, faculty, and administrators to uncover meaningful connections or collaboration opportunities.

The primary goal of this project is to bridge these gaps by developing an **interactive visualization tool** that maps the relationships between NYU Tandon faculty members and their research interests. Inspired by existing visualizations, such as Harvard's Faculty Connections tool [2], this project aims to create a **chord diagram** tailored to NYU's unique academic environment.

This project addresses the need for a centralized platform that not only simplifies the process of discovering faculty expertise but also fosters collaboration and interdisciplinary research. Unlike static or overly complex tools, this interactive visualization uses dynamic elements to allow users to explore relationships in a meaningful and accessible way.

By focusing on faculty research areas and their connections, this tool aims to support various stakeholders:

- **Students:** To find potential mentors or thesis advisors aligned with their interests.
- **Faculty:** To identify collaborative opportunities with colleagues from other departments.
- **Administrators:** To gain insights into the academic strengths and gaps within NYU Tandon's ecosystem, aiding in strategic decision-making.

The visualization builds upon the lessons learned from prior works, particularly Harvard's tool, which effectively used a circular chord diagram to depict interdisciplinary ties. However, this project addresses limitations seen in existing tools, such as the lack of interactivity, non-scalable designs, and insufficient representation of broader academic ecosystems. By leveraging modern web technologies, such as **D3.js**, this project ensures scalability, interactivity, and visual clarity while aligning with NYU Tandon's academic structure and branding.

Ultimately, this tool will empower users to better navigate the rich academic landscape at NYU Tandon, facilitating discovery, collaboration, and innovation across disciplines.

4 Technical Details

4.1 Technical Implementation

Frontend: HTML, CSS, and JavaScript for rendering and interactivity.
Data Handling: D3.js for dynamic visualization and interaction.
Data Scraping: Python, BeautifulSoup library

4.2 Data WebScraping

The data collection process involved developing a comprehensive Python-based web scraping system to gather faculty information from Google Scholar profiles. We implemented multiple scraping functions using BeautifulSoup4 and Selenium to extract various data points from faculty profiles.

Our scraping architecture consisted of several key components:

- Profile information extraction (name, affiliation)

- Research interests and expertise areas collection
- Publication records and citation metrics gathering
- Paper details including titles, authors, and venues

To ensure data quality and prevent server overload, we implemented:

- Random time delays between requests (2-5 seconds)
- Robust error handling for failed requests
- Data validation and cleaning procedures
- Individual JSON file storage for each professor

Our data processing pipeline followed these steps:

1. Initial collection of faculty Google Scholar URLs
2. Profile information extraction
3. Research interests categorization
4. Publication data aggregation
5. Data cleaning and standardization
6. JSON/CSV export for visualization

This systematic approach to data collection enabled us to create a comprehensive dataset that powers our interactive visualization system. The scraping process collected faculty profiles from NYU Tandon, resulting in a structured dataset containing research interests and publications that formed the foundation for our chord diagram and matrix visualizations.

4.3 Chord diagram Visualization

The chord diagram[5] is the focal point of this project, designed to visualize relationships between NYU faculty members and their research interests. This section outlines the visual elements, interactivity tools, algorithmic details, and design decisions that contribute to the functionality and effectiveness of the visualization.

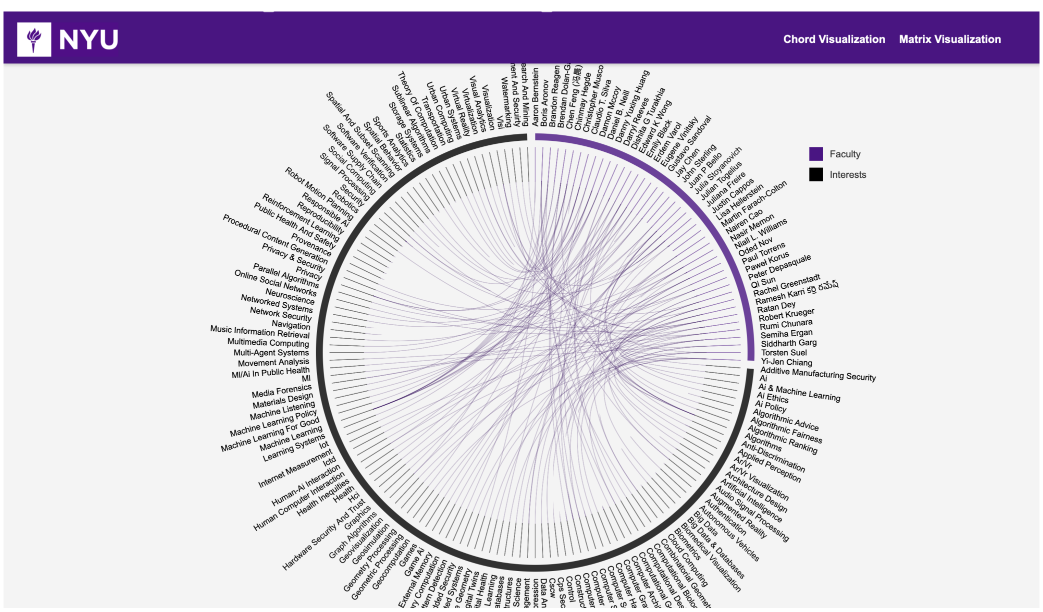

Figure 2: NYU Tandon School Of Engineering Chord Diagram Visualization (NYU CSE)

4.3.1 Visual Elements and Techniques

- **Arcs and Ribbons**:
  - **Arcs**: Represent nodes in the dataset, categorized into two groups—faculty members and research interests. Faculty arcs are colored in a rich purple (`#500F87`), while research interest arcs use a deep indigo (`#1F0535`) to distinguish the two.
  - **Ribbons**: Indicate connections between faculty and research interests. The width of a ribbon is proportional to the weight of the connection, effectively encoding the strength of collaboration or shared interest.
- **Legend**: A legend is positioned to the right of the chord diagram to clarify the color coding for faculty and interests. It includes labeled colored rectangles for each category to ensure easy interpretation.
- **Tooltip**: A dynamic tooltip provides additional information about the hovered arc, such as the faculty member's name, type (faculty or interest), and related entities. The tooltip is styled with a dark background and rounded corners for better readability.

4.3.2 Interactivity Tools

- **Hover Effects**: Hovering over an arc highlights its related connections, while dimming unrelated elements. This selective emphasis helps users focus on specific relationships within the dataset.
- **Clickable Links**: Clicking on a faculty arc redirects to their Google Scholar profile, allowing users to explore their academic contributions in detail. Clicking on a research interest arc initiates a Google search for that topic, providing broader context.
- **Dynamic Tooltip**: The tooltip dynamically updates its content and position based on the hovered arc or ribbon. It transitions smoothly to enhance user experience.

4.3.3 Algorithmic Details

- **Data Processing**: The input CSV was preprocessed into two JSON files:
  - **Nodes JSON**: Contains nodes representing faculty and research interests.
  - **Links JSON**: Defines the connections between nodes and their weights.
- **Chord Layout**: The D3.js `chord` layout algorithm calculates the angles for arcs and ribbons based on the input matrix, ensuring that the connections are visually proportional to their weights.
- **Matrix Generation**: A connection matrix was generated from the links JSON, with each cell representing the weight of the relationship between a pair of nodes. This matrix serves as the input for the chord layout algorithm.

4.3.4 Design Justifications

- **Color Choices**: The colors were chosen to align with NYU branding while maintaining sufficient contrast between faculty and research interests for clear differentiation.
- **Interactivity**: The combination of hover effects, tooltips, and clickable links makes the visualization intuitive and engaging, enabling users to explore relationships without overwhelming visual clutter.
- **Legend Placement**: The legend is positioned to the right of the diagram to avoid overlapping with the main visualization while remaining easily accessible.

4.3.5 Implementation Details

- **Technologies Used**: The visualization was implemented using **D3.js** for rendering and interactivity, along with **HTML** and **CSS** for layout and styling.
- **SVG Elements**: The arcs, ribbons, and legend components are rendered using scalable vector graphics (SVG) for crisp and scalable visual output.

- **Custom Event Handlers**: Interactive features such as hover, click, and tooltip updates were implemented with D3.js event handlers to ensure a seamless user experience.

The interactive chord diagram effectively visualizes the relationships between faculty and research interests at NYU. Its thoughtful design and interactivity enable users to uncover patterns, explore collaborations, and gain insights into interdisciplinary research. The techniques and tools used in this project make it a valuable resource for academic analysis.

4.4 Matrix Visualization

The Matric visualization serves as one of the core elements of this project, crafted to illustrate the connections between NYU faculty members and their research interests. This section delves into the visual components, interactive features, algorithmic implementation, and design choices that enhance the functionality and impact of the visualization.

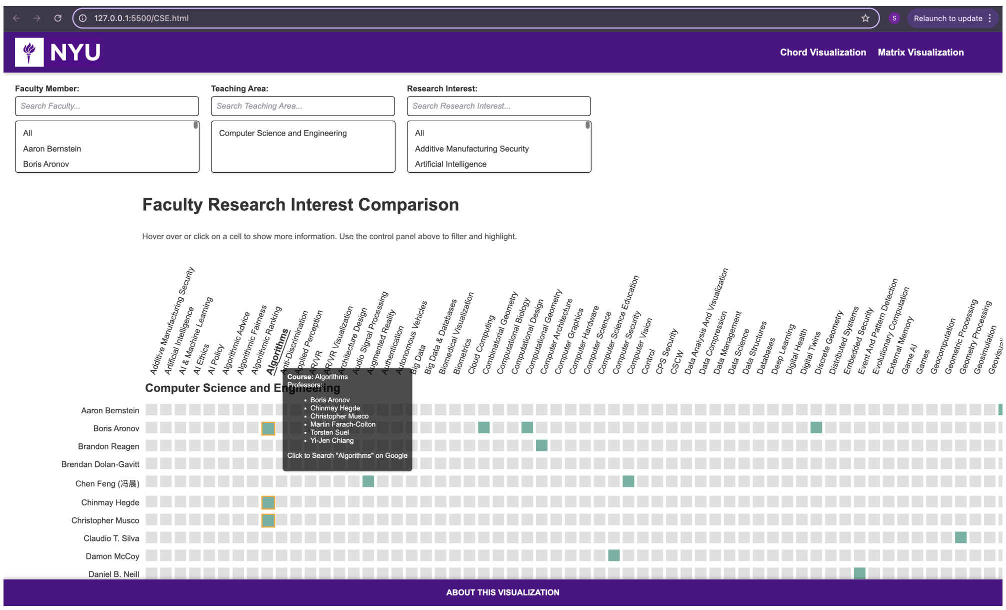

Figure 3: Faculty-Research Matrix Visualization (NYU CSE)

4.4.1 Data Preprocessing:

One-Hot Encoding (Python)

1. **Parse Research Interests:** Split comma-separated values for each faculty.
2. **Generate Features:** Extract all unique research topics.
3. **Encoding:** Assign values (random generated number if a topic exists, otherwise `0`). Assign non-binary weights (> `0.0`) for relevance.
4. **Output Matrix:** Create a sparse matrix with faculty as rows and research topics as columns, then export as CSV for D3.js.

Matrix Creation (D3.js)

1. **Load Data:** Read the CSV and extract unique faculty and topics.
2. **Grid Layout:** Create a CSS grid with faculty as rows and topics as columns.
3. **Populate Matrix:** Add `div` cells, style based on values (e.g., color for > `0.0`).
4. **Dynamic Updates:** Rebuild the grid efficiently on filter changes, ensuring smooth interactivity.

4.4.2 Choice of Visualization:

The matrix visualization was chosen to represent faculty research interests across multiple dimensions, such as teaching areas and specific research topics. This visualization is well-suited for data sets where the relationships between entities (faculty and research topics) need to be highlighted, providing an intuitive way to explore overlaps and gaps. The choice ensures clarity and scalability, accommodating a large dataset without overwhelming the viewer.

4.4.3 Visual Encoding:

- **Rows and Columns:** Rows represent faculty members, while columns correspond to research topics. The grid structure ensures easy navigation and cross-referencing.
- **Cell Encoding:** Filled cells indicate a match between a faculty member and a research topic. These cells are visually differentiated using color, while empty cells remain neutral to reduce visual clutter.
- **Highlighting:** Hovering over a row, column, or specific cell dynamically highlights associated entities, improving discoverability and understanding of relationships.

4.4.4 Additional Considerations:

- **Data Accessibility:** Alphabetical ordering of both faculty names and research topics enhances navigability. Users can quickly locate desired information without excessive scrolling or search.
- **User Experience:** Dropdown filters allow users to refine the visualization based on specific teaching areas, faculty, or research interests. This targeted exploration minimizes cognitive overload while supporting diverse use cases.
- **Scalability:** The matrix can scale to include additional faculty members and research topics without significant loss of usability, as the encoding remains consistent.

4.4.5 Colour Encoding:

- **Cell Colors: Active cells** are highlighted using a distinct, high-contrast color (e.g., **green**), drawing immediate attention to relationships while ensuring accessibility.
- **Hover and Interaction Colors: Hover states** use subtle shading to indicate focus without overshadowing the data itself. This transient feedback ensures users stay oriented during exploration.
- **Neutral Background:** The **background** remains neutral (**gray/white**) to avoid interference with data encoding and maintain a clean aesthetic.

4.4.6 Interaction Techniques:

- **Dropdown Filters:** Users can dynamically filter data and multiple selection by faculty, teaching area, or research interest, tailoring the visualization to their specific needs.
- **Hover Effects:** Hovering over a cell highlights corresponding rows and columns, revealing immediate relationships and context. Additional details, such as faculty names and associated topics, are displayed in a tooltip.
- **Clickable Cells:** Clicking on a research topic redirects users to a Google search, while clicking on a faculty member's name links to their Google Scholar profile (if available). This feature connects the visualization with external resources for deeper exploration.
- **Collapsible Info Bar:** A collapsible bar at the bottom provides context for the visualization, including usage instructions and its broader purpose.

4.4.7 Insights:

- **Unique vs. Shared Research Areas:**

- **Unique Expertise:** Some topics have only one faculty member associated, indicating niche expertise. For example, if **Chen Feng** is the only person linked to **Autonomous Vehicles**, it reflects a specialized focus in this area.
- **Shared Research:** Topics like **Algorithmic Fairness** may show associations with multiple faculty members, suggesting collaborative opportunities within the department.

- **Faculty Research Profiles:**

  Faculty members like **Aaron Bernstein** might have fewer highlighted cells, suggesting a more focused research scope. In contrast, someone with many filled cells (e.g., **Christopher Musco**) may have a more diverse or interdisciplinary research portfolio.

5 Discussion

Our visualization project of NYU Tandon faculty connections revealed several significant insights and challenges worth discussing:

5.1 Key Findings

- **Interdisciplinary Patterns:** We discovered substantial overlap in research interests across departments, particularly between Computer Science and Electrical Engineering. This suggests strong potential for cross-departmental collaboration.
- **Research Clusters:** The visualization revealed distinct clusters of research expertise, with emerging technologies like artificial intelligence and cybersecurity showing strong representation across multiple departments.
- **Department Connectivity:** Some departments showed stronger interdisciplinary connections than others, potentially highlighting areas where cross-department collaboration could be enhanced.

5.2 Technical Challenges

During the development process, we encountered several technical challenges:

- **Data Collection:**
  - Managing rate limits during web scraping
  - Handling inconsistent profile formats
  - Standardizing research interest terminology
- **Visualization Implementation:**
  - Optimizing the chord diagram for large datasets
  - Balancing visual complexity with usability
  - Ensuring responsive performance across different devices

5.3 Limitations

Our current implementation has several limitations:

- **Data Completeness:** Not all faculty members maintain updated Google Scholar profiles, potentially leading to incomplete representation.
- **Temporal Aspects:** The visualization represents a snapshot in time and doesn't capture the evolution of research interests.
- **Scalability:** The chord diagram becomes visually complex with large numbers of connections, potentially affecting readability.

5.4 Future Improvements

Based on our findings, we propose several potential enhancements:

- **Dynamic Updates:** Implement automated data collection to keep the visualization current.
- **Enhanced Filtering:** Add more sophisticated filtering options to focus on specific research areas or time periods.
- **Alternative Views:** Develop additional visualization modes for different analytical perspectives.
- **Collaboration Features:** Add functionality to facilitate direct connection between researchers with shared interests.

5.5 Impact and Applications

The visualization tool has several potential applications:

- **Student Guidance:** Helping students identify potential advisors and research opportunities.
- **Research Collaboration:** Facilitating interdisciplinary research partnerships.
- **Administrative Planning:** Supporting strategic decisions about department development and faculty hiring.
- **Resource Allocation:** Identifying areas of strength and potential gaps in research coverage.

This project demonstrates the value of interactive visualization in understanding academic relationships and research patterns within a university setting. While there are areas for improvement, the current implementation provides valuable insights for various stakeholders in the academic community.

6 Conclusion

Our visualization project successfully transformed the traditional faculty directory into an interactive, data-driven exploration tool. Through the implementation of chord diagrams and matrix visualizations, we achieved our primary objectives of:

- Creating an intuitive interface for exploring faculty connections
- Highlighting interdisciplinary research opportunities
- Facilitating student-faculty matching based on research interests

The project demonstrates the power of combining web scraping techniques with modern visualization approaches to create meaningful academic tools. If given more time, we would enhance the project by:

- Implementing real-time data updates from Google Scholar profiles
- Adding publication network visualization features
- Incorporating citation metrics into the visual encoding
- Developing filters for temporal analysis of research trends

This visualization tool serves as a foundation for future developments in academic network analysis and could be extended to other schools within NYU or adapted for different academic institutions.

7 Workload Distribution

The project work was distributed among team members based on their expertise and interests:

Dhairyasheel Patil (dp3979):

- Developed comprehensive web scraping scripts using Python, BeautifulSoup4, and Selenium
- Implemented data collection pipeline for faculty profiles and research interests
- Created data preprocessing and cleaning procedures
- Contributed to the project documentation and final report

Kaartikeya Panjwani (kp3291):

- Designed and implemented the interactive chord diagram visualization
- Developed D3.js-based visualization components
- Created hover effects and interactive features
- Contributed to the technical documentation and final presentation

Sanyukta Tuti (st5442):

- Implemented the matrix visualization component
- Created the color encoding scheme and interaction techniques
- Developed filtering and search functionality
- Coordinated the project documentation and report writing

All team members participated in regular meetings, contributed to design decisions, and collaborated on integrating the various components into a cohesive visualization system.

8 Supplementary Details

To explore the interactive visualization and access the project's codebase, the following resources are provided:

- **Interactive Visualizations:** Both the chord diagram and matrix visualization and their interactive features can be accessed online at the following link: https://kaartikeya15.github.io/info-viz-final/.
- **GitHub Repository:** The full codebase, including data preprocessing scripts, visualization code, and supporting files, is available on GitHub: https://github.com/kaartikeya15/info-viz-final.

9 Works Cited

References

[1] X. Chen et al. "Visualizing Academic Networks: Insights and Challenges". In: *Journal of Visualization Research* (2020).

[2] Harvard School of Engineering and Applied Sciences. *Faculty Connections*. Accessed: 2024-11-26. 2024. URL: `https://facultyexplorer.seas.harvard.edu/faculty-connections/`.

[3] NYU Tandon School of Engineering. *Staff Directory*. Accessed: 2024-11-26. 2024. URL: `https://engineering.nyu.edu/staff-directory`.

[4] A. Smith and B. Jones. "Scalable Visualization Techniques for Large Academic Datasets". In: *Proceedings of the ACM Symposium on Visual Analytics*. 2019.

[5] Kai Wang et al. "CHORDination: Evaluating Visual Design Choices in Chord Diagrams for Network Data". In: *17th International Symposium on Visual Information Communication and Interaction* (2024).