**Methods**

**Data**

We use the datasets provided by Scholars@Duke and The Graduate School for our analysis and visualization. To analyze the publications and collaborations, only the following datasets and columns are used:

* 'DUID', 'APPOINTMENT\_TYPE' (with rows that represent each faculty’s principle appointment only), 'SCHOOL\_NAME', and 'ORG\_DISPLAY\_NAME' columns from faculty dataset
* 'DUID', 'PUBLICATION\_URI', and 'PUBLICATION\_DATE' columns from publications dataset

The 'DUID' column is our reference to find the schools and organizations associated with each publication.

To analyze the graduate committees, the following datasets and columns are employed:

* 'DUID', 'APPOINTMENT\_TYPE' (with rows that represent each faculty’s principle appointment only), 'SCHOOL\_NAME', and 'ORG\_DISPLAY\_NAME' from faculty dataset
* 'Student random ID', 'Degree Nbr', 'Degree', 'Advisor Role', 'Advisor', 'Acad Org', and 'Advisor Duke UID' from dissertation committees dataset

Again, we use 'Advisor Duke UID' and 'DUID' from these two datasets to determine the affiliations of each student’s committee members.

**Procedure:**

First we look at the number of publications per year for each school at Duke. Using the datasets above, we can group the publications based on the year they were published and the schools their authors are affiliated with. Then we can count the total number of unique publications for each school and plot them versus the year they were published in. Since the datasets do not provide information for all months in 2012 and 2017, we disregard the data from these years and use the data from 2013-2016 to predict the number of publications for 2017 and 2018. Linear regression works for all schools except Divinity School, and for the data from this school we use exponential fit. Our predictions for 2017, agree with the results of those few schools that have provided information for more than 10 months in 2017.

Using the same columns, we create a matrix where each row represents the publications of each school per year. By calculating the frequency of repeated publications in each row we can find the papers that have more than one author affiliated with the same school. Then we can use their DUID and 'ORG\_DISPLAY\_NAME to recognize if these collaborations have happened between different departments. We keep track of the number of inter-departmental publications in each school. Then by comparing the unique list of publications at each school to the list of publications at other schools and counting the number of intersecting elements, we can find the number of publications that cross school boundaries. This number is added to the count of interdepartmental collaborations within the school, divided by the total number of publications, and plotted versus year to show what percentage of publications in each school are interdisciplinary.

There are two graphs representing the publications at Duke, and they are both generated by Gephi, a software that is widely used for network analysis and visualizations. To produce the input for these two graphs, we generate a weighted matrix where the rows and columns represent the schools/departments and each matrix element [i,j] is the logarithm of number of unique co-authorships between schools/departments i and j. Each node in the graph represents a school or departments and its size is proportional to the total number of publications in that unit. The edges represent the collaborations and their thickness is proportional to the number of unique co-authorships between the two schools or departments over the course of five years.

To analyze the overlaps between different departments and schools, we also generate a matrix based on dissertation committee members, where each row represents the student department and each column represents a school. The [i,j] element in this matrix, is the number of students in department i who have selected their dissertation committee members from school j. By inputting this matrix into Gephi, we can create a directed graph where each node represents either a department or a school and directional edges show where the students from each department go to select a committee member. We have used a dual circle layout for this plot, where the schools are located in the outer circle and the departments are arranged on the inner circle. The size of each node is proportional to the number of ingoing edges to the node and the width of each edge is proportional to the number of students in the source node who have selected their advisor or committee members from the target node (school).