

# COSC 6323 - Statistical Methods in Research

## Project Phase - 1

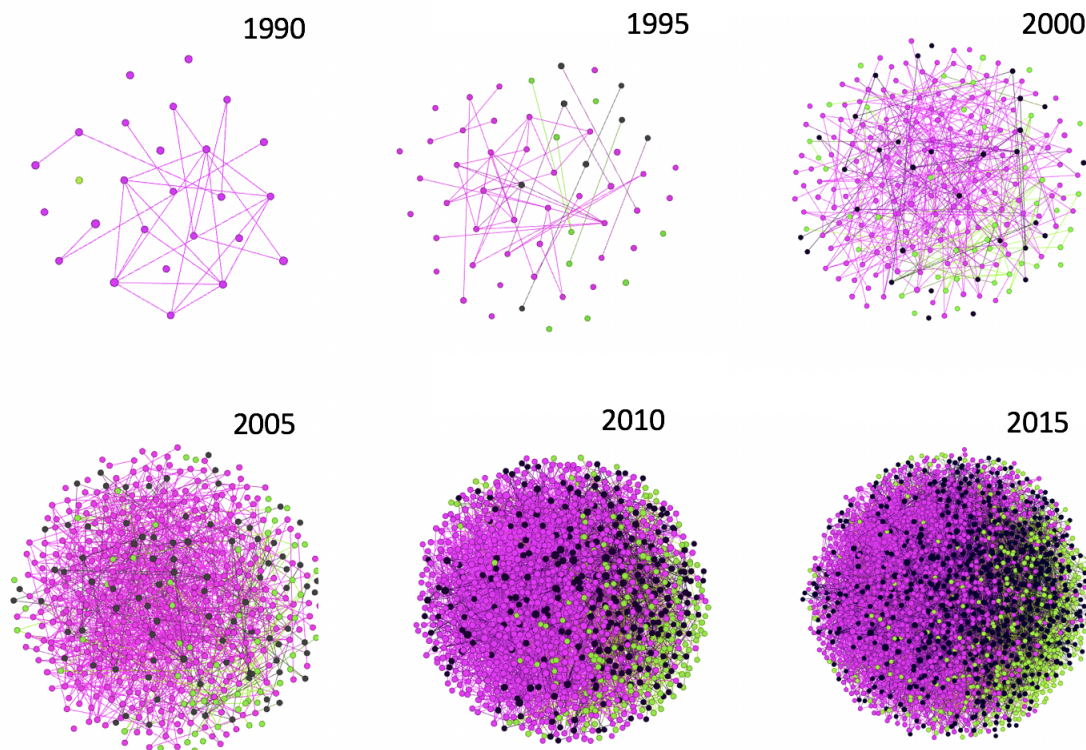
Members: Team-8

1. Md Rafiqul Islam Rabin, ID:1797648, [mrabin@central.uh.edu](mailto:mrabin@central.uh.edu)
2. Salah, ID:, @
3. Farah, ID:, @

March 08, 2019.

**Contributions:** In the first week of the project, we worked on Fig 3. At first, we drew Fig 3 with ggplot, but as we faced a problem with y-log transformation, we moved to plot\_ly. We all equally contributed to Fig 3. After that, we sit together several times to discuss the required data for Gephi and Supplementary plot. Then we individually collect data (i.e. XDIndicator, Node, Edge, Etc.) from given CSV files. We cross-checked our data for confirmation. So, we all also equally contributed to the data processing. Now, at that moment, we divided our remaining tasks among us: 2B and S2 for Rabin, 2A and S4 for Farah, S1 and S3 for Salah. Although we divided our tasks at that stage, we always shared our progress/problem and helped each other during implementation. We created a git repository and distributively contributed for the project. As we all completed our tasks and actively involved with each other all the times, we almost equally contributed to this phase of project. Here is the overall contribution percentage: Rabin(35%), Farah(33%) and Salah(32%).

**Fig. 2(A) :**



**Fig. 2. Growth of cross-disciplinary social capital. (A)** Evolution of the giant component in the U.S. biology-computing network.

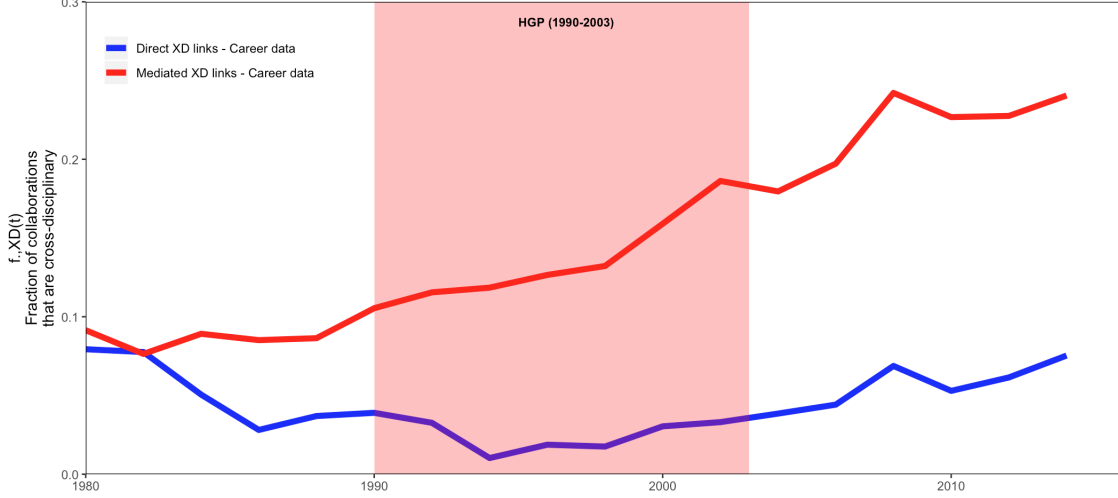
**Description of figure content:**

Here, ...

**Observations, conclusions, and hypotheses:**

Here, ...

**Fig. 2(B) :**



**Fig. 2. Growth of cross-disciplinary social capital. (B)** Evolution of the fraction of collaboration links in the  $F$  network that are cross-disciplinary.

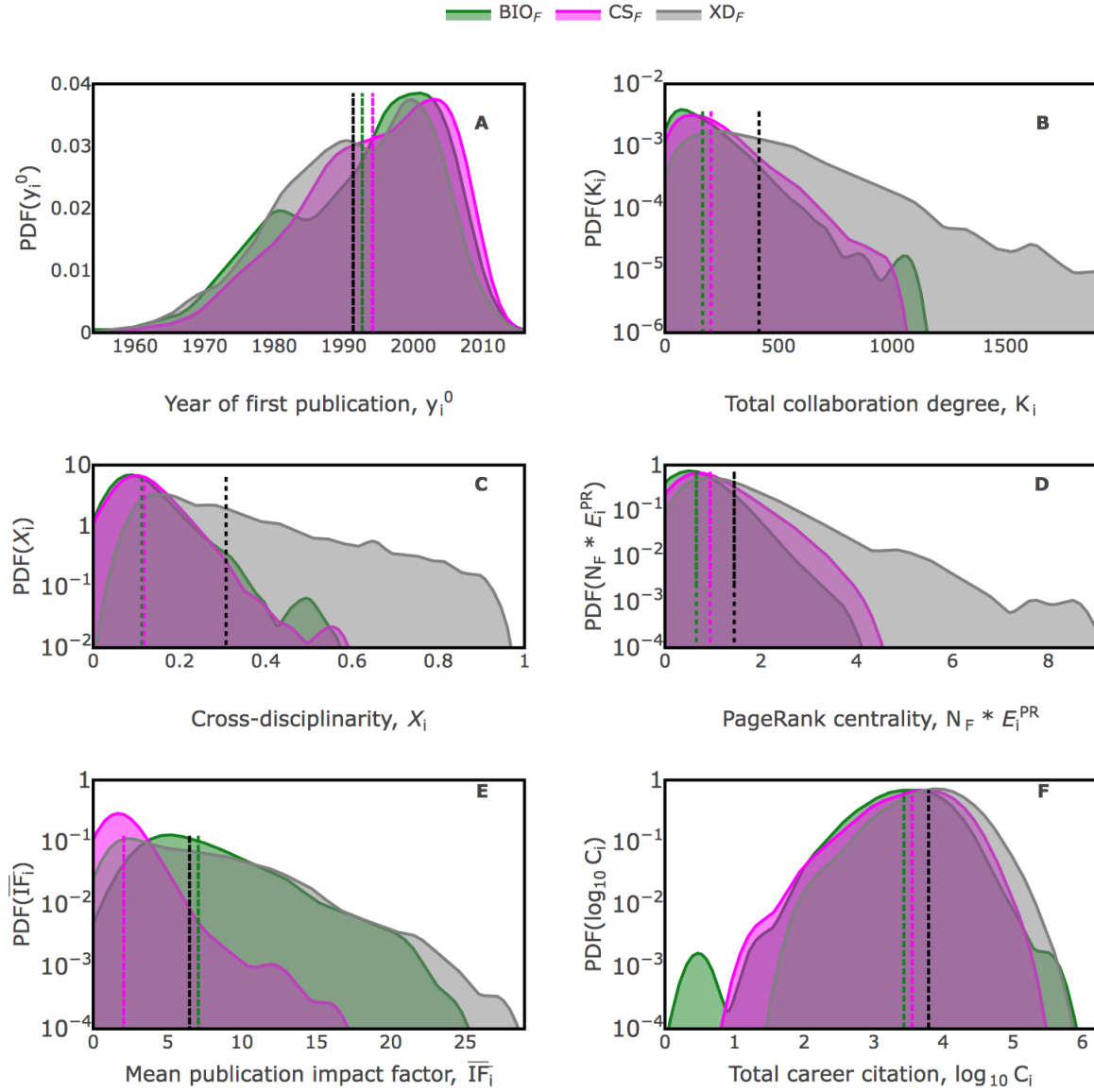
### Description of figure content:

Here, in Fig 2B, we plotted the fraction of cross-disciplinary collaboration for each nonoverlapping 2-years period through (1979-1980) to (2013-2014). The blue and red line represents the trend of Direct-XD links and Mediated-XD links, respectively. For each nonoverlapping 2-years period, we collected all the publication data from **GoogleScholar\_paper\_stats.csv** file. Then, for the Direct-XD links, we count total direct (F-F) links and total cross-discipline direct links for each period. Finally, we calculated the fraction of direct cross-disciplinary collaboration by **total F-F direct links / total cross-discipline direct links**. Similarly, for the Mediated-XD links, we count total mediated (F-P-F) links and cross-discipline mediated links for each period. Finally, we calculated the fraction of mediated cross-disciplinary collaboration by **total F-P-F mediated links / total cross-discipline mediated links**.

### Observations, conclusions, and hypotheses:

- The first thing we can see that, the growth is comparatively higher for mediated cross-disciplinary collaboration but slower for direct cross-disciplinary collaboration. This highlights that pollinators are more likely to be cross link than direct faculty persons.
- Both the direct and mediated cross-disciplinary collaboration has increased during HGP(1990-2003) period and the increasing rate also continues upward after the HGP as well. So, the impact of HGP on cross-disciplinary collaboration is clearly visible.
- The increasing rate of cross-disciplinary collaboration continues. In recent years (i.e. by 2015) the direct cross-disciplinary collaboration becomes almost 10% of total direct collaboration, and the mediated cross-disciplinary collaboration becomes almost 30% of total mediated collaboration (whether within-discipline or cross-discipline).

**Fig. 3 :**

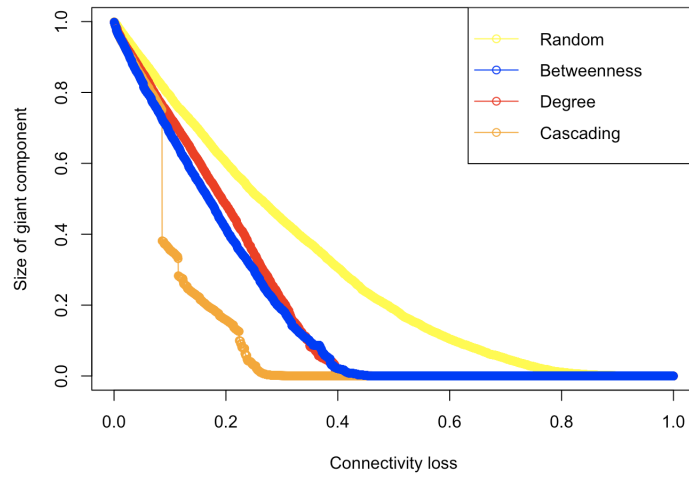


**Fig. 3. Descriptive statistics for the career data set.**

**Extended caption:**

Here, the vertical lines indicate the distribution means for the corresponding subsets. (A) ... (B) ... (C) ... (D) ... (E) ... (F) ...

**Fig. S1 :**



**Fig. S1. Robustness of the F network with respect to link removal.**

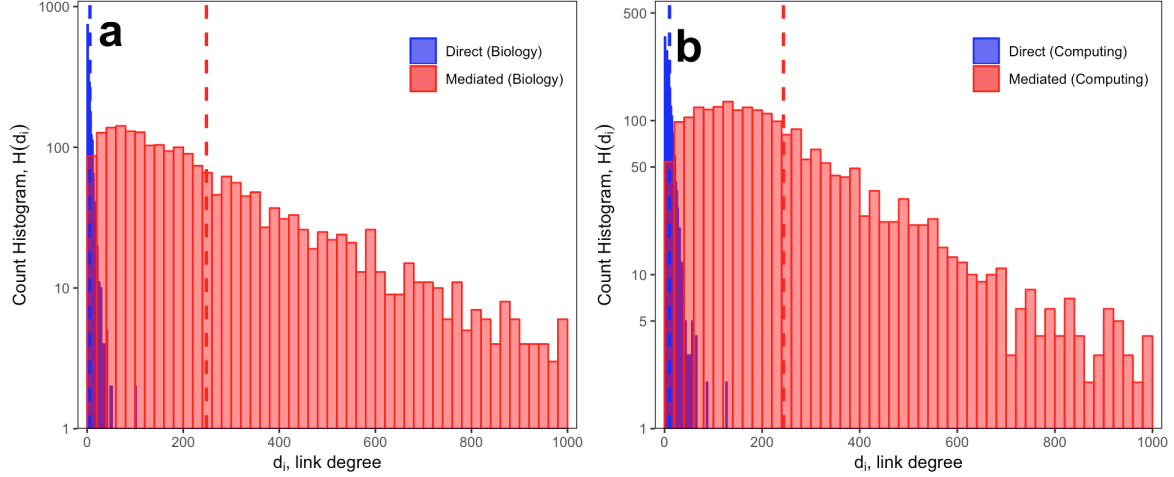
**Description of figure content:**

Here, ...

**Observations, conclusions, and hypotheses:**

Here, ...

**Fig. S2 :**



**Fig. S2.  $F$  network distributions for direct and mediated associations.**

### **Description of figure content:**

Here, in Fig S2, we showed the frequency distribution (counts) of faculty within a given link degree. Histogram (a) for biology department and (b) for computing department. At first, we read the **KDirect** and **KMediated** data from **Faculty\_GoogleScholar\_Funding\_Data\_N4190.csv** file for **BIO** and **CS** department, separately. Then we used  $\text{binwidth}=20$  for mediated counts and  $\text{binwidth}=2$  for direct counts, for both biology and computing department. The blue and red bar represent the counts for direct and mediated links, respectively. The blue and red vertical lines indicate the distribution means for direct and mediated links, respectively.

### **Observations, conclusions, and hypotheses:**

- In higher link degree (i.e after 200 link degree), the mediated link counts are still significant but there is almost no direct link counts, for both biology and computing department.
- More than 90% co-authors are pollinators, for both biology and computing department.
- The mean line, for both biology and computing department, clearly shows that the mediated link degree are higher than the direct link degree.
- Finally, the significant impact of pollinators in network distributions is clearly visible.

**Fig. S3 :**



**Fig. S3. Three perspectives on the centrality of  $F_i$  in the direct collaboration network.**

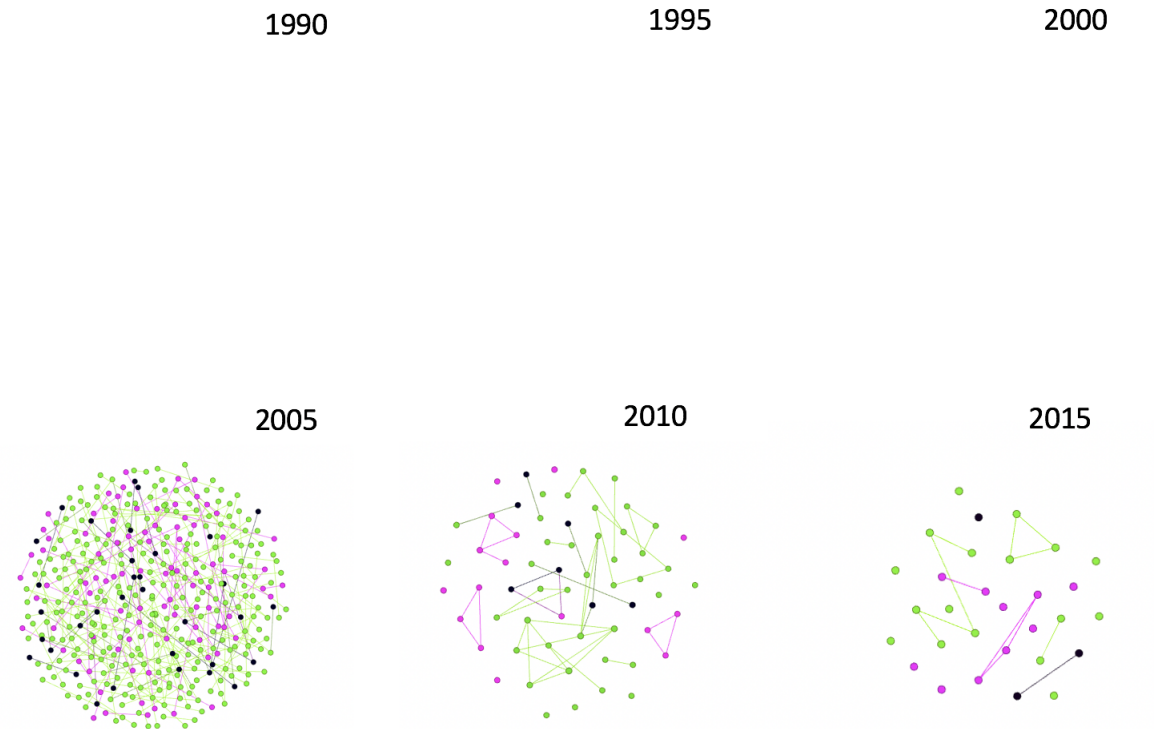
**Description of figure content:**

Here, ...

**Observations, conclusions, and hypotheses:**

Here, ...

**Fig. S4 :**



**Fig. S4. Evolution of the nongiant components in the F network.**

**Description of figure content:**

Here, ...

**Observations, conclusions, and hypotheses:**

Here, ...