COCS 6323: Statistical Methods in Research

Group Project

Group <number>

Department of Computer Science

University of Houston

March 8, 2019

Contents

List of Tables

List of Figures

1 Contribution

Member	**Contribution**
Brad	3A 3B 3C S2
Tung Huynh	2A 2B S3 S4
Yifan	3D 3E 3F S1

Table 1: Contribution of group member

2 Figure 2

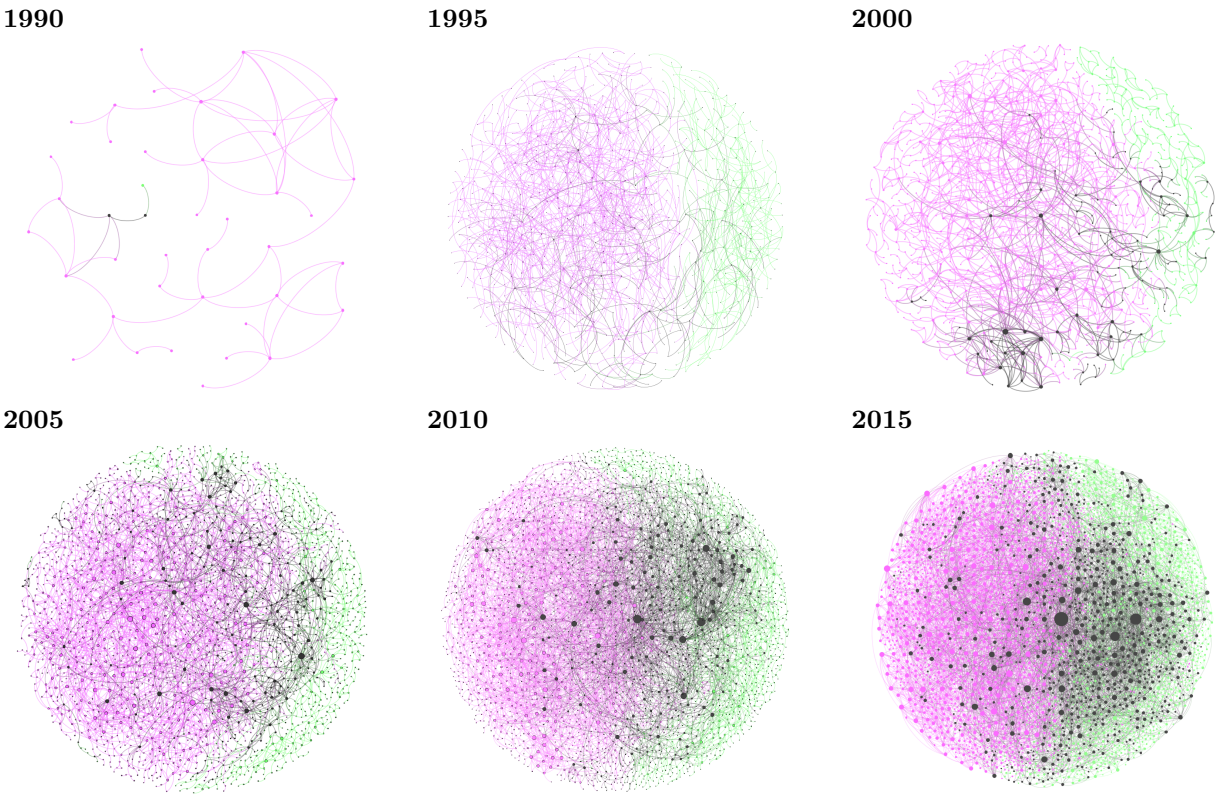

Figure 1: Growth of cross-disciplinary social capital

This figure depicts the evolution of the giant component in the U.S. biology-computing network of collaborations. It consists of six consecutive periods from 1990 to 2015. Each period illustrates the collaborations of the two departments in five previous years. While green and magenta nodes represent faculty F_i in BIO and CS department, respectively; black vertices represent faculty F_i that published at least one cross-disciplinary publication. In this figure, node size is proportional to the logarithm of the degree centrality, the total number of collaborations of the faculty.

In the first network of the year 1990, the giant component only consists of links between CS nodes. In the period from 1900 to 2005, in which the HGP happens, the networks represent the establishment of cross-disciplinary collaborations. Finally, in the last period from 2005 to 2015, the giant component does not only expands in term of the number of collaborations and robustness, but it also notably illustrates the significant role of XD group containing the largest nodes in the network.

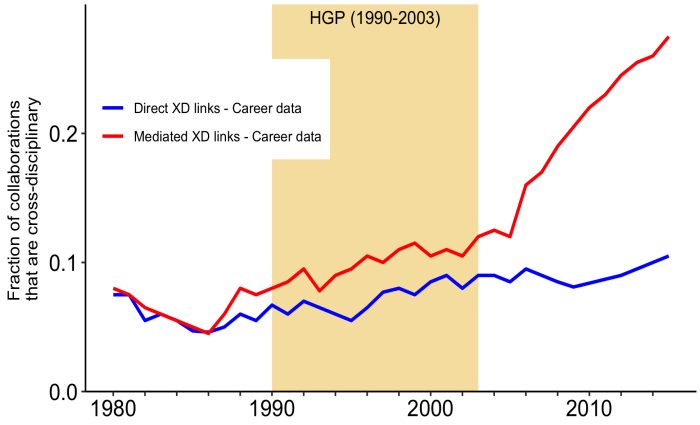

Figure 2: Evolution of the fraction of cross-disciplinary collaboration links

This figure depicts the evolution of the fraction of collaboration links in the network that are cross-disciplinary. While blue line illustrate the direct XD links, the red line represent the mediated XD links by pollinators. The orange area in the middle annotates the HGP project period from 1990 to 2013.

3 Figure 3

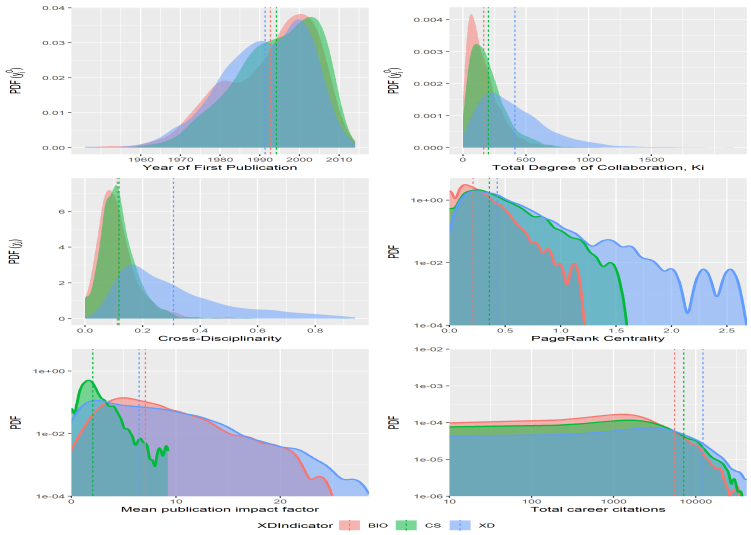

Figure 3: Descriptive statistics of the career dataset

Figures 3 a-F, are probability distributions of the following data: year of first publication, total number of collaborators, the fraction of collaborators who are XD, PageRank centrality of professors within the network, and mean impact factor of the publication record of professors within the network. The data visualization above supports a conclusion that XD professors are more collaborative than BIO or CS specific professors. Figures 3B, C, and D show that XD professors are more collaborative than their single-discipline counterparts. XD professors are also shown to have greater total degrees of collaboration, a greater fraction of collaborators who themselves are XD, and higher overall PageRank centrality. XD professors also outperform BIO and CS in terms of mean publication impact factor. Figure 3A suggests that XD professors published their first paper in genomics on average before BIO and CS.

4 Supplementary S1

Add your content here

Figure S1a depicts the ratio of the largest component other than the initial giant component within the F network. By randomly removing links, the robustness of the network is tested through the observation of the relative size of the remaining giant component to the next largest subnetwork. Error bars are fitted to points on the curve depicting mean and standard deviation. This ratio slowly declines until roughly (placeholder), from this observation we can assert (placeholder). Figure S1b shows when the network breaks up into significantly smaller subnetworks and the giant component itself degrades. This measured peak breaking point coincides with the steep decline observed in S1a, confirming that at this fraction of links removed, the ratio of giant component to next largest subnetwork declines as well.

5 Supplementary S2

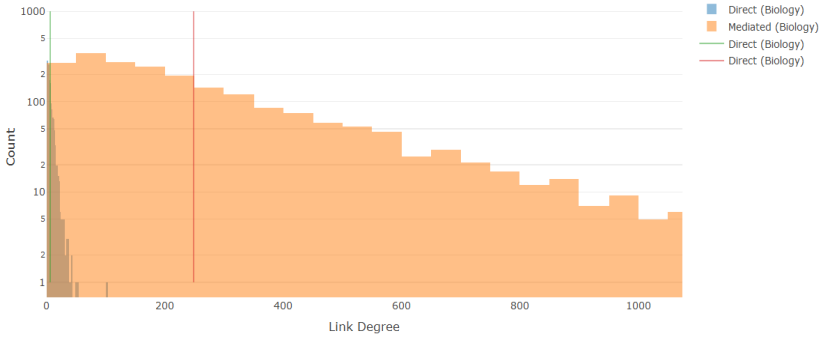

Figure 4: Network Distributions for Direct and Mediated Associations of Biologists

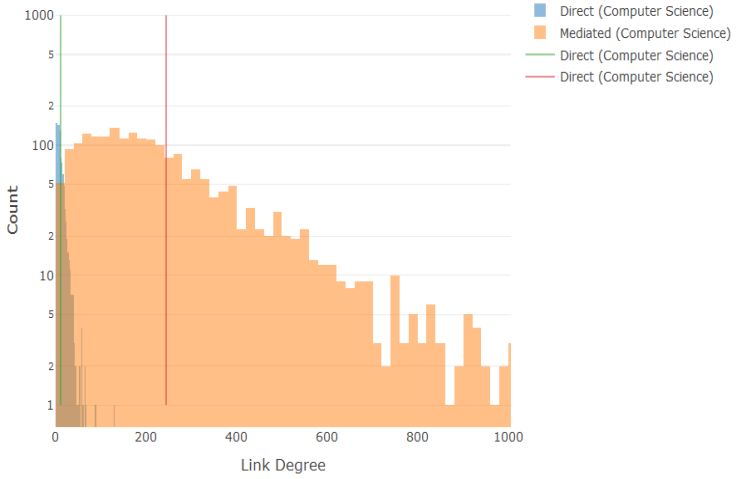

Figure 5: Network Distributions for Direct and Mediated Associations of Computer Scientists

Figure S2 depicts the number of associations between professors within BIO departments (a), and CS departments (b) with other researchers studying genomics. The dark area shows counts of the direct connections between professors, while the lighter orange area shows counts of mediated connections. Vertical lines represent means of each distribution. Direct links are made when professors within the Fi dataset collaborate to publish a paper. Mediated links are established when professors have a collaborator who exists outside of Fi in common. The histograms above demonstrate the importance of the mediated connections, with regards to the robustness of the network. Thus, professors who are outside of the dataset at hand make an important impact on the social network studied here.

6 Supplementary S3

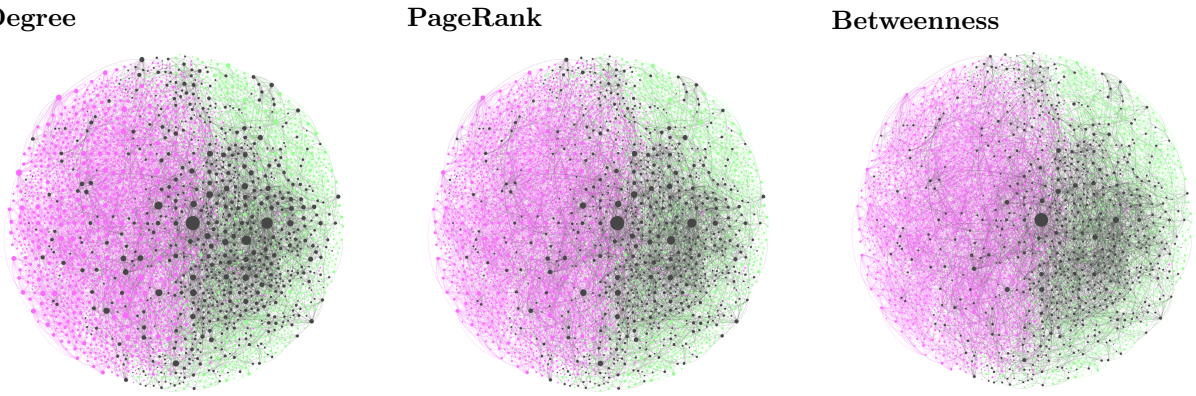

Figure 6: Three perspectives on the centrality of F_i in the direct collaboration network

This figure shows the giant connected component of the faculty network using all data up to 2015 from three different perspectives. The position of nodes and links are unchanged in the three networks. And the node sizes change respectively to the three centrality measure including Degree, PageRank, and Betweenness.

While Degree centrality treats the importance of each node equally, PageRank takes into account the prominence of each faculty. Therefore, with the same number of links, a node size of faculty F_i in the network using PageRank centrality measure changes to smaller or larger than its counterpart in the network using Degree centrality measure according to the prominence of his collaborators.

Besides, because the Betweenness centrality measures the number of shortest paths passing through a node, the network using this measure only emphasize on those nodes helping to increase the robustness of the whole network.

Remarkably, regardless of the change of centrality measure using in each network, the group of founders of HGP including Eric Lander always demonstrate their significant role of connecting the cross-disciplinary links with the highest central.

7 Supplementary S4

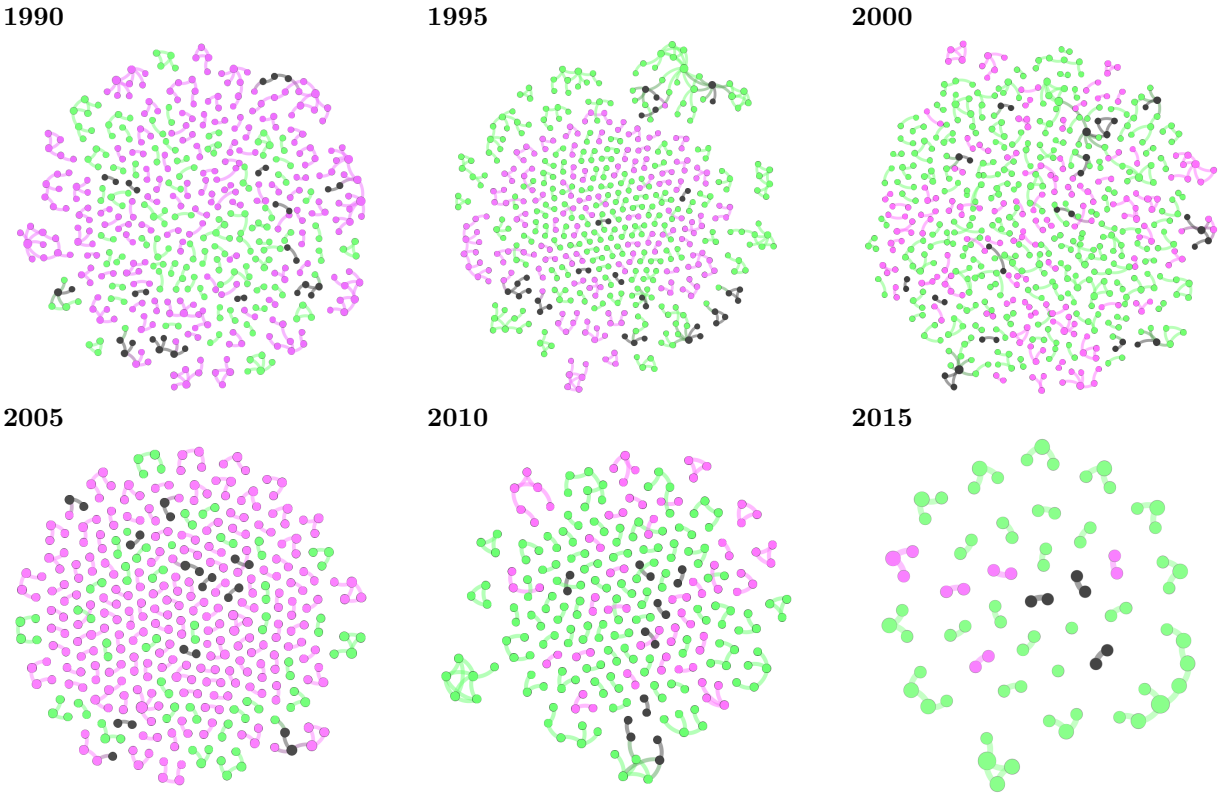

Figure 7: Evolution of the nongiant components in the network

This figure depicts the development of the nongiant components in the direct collaboration network. Similarly to Figure 2A and Figure S3, green and magenta nodes represent faculty F_i in BIO and CS department, respectively, while black nodes represent faculty F_i in the XD group.

Notably, the number of nodes and links decreases significantly from 1990 to 2015. If in the year 1990, due to the lack of cross-disciplinary collaborations, there are some local strong groups of internal collaboration. In the 1991-1995 period, the network also presents the establishment of cross-disciplinary collaboration. However, those groups of collaboration are still local and disjointed. From 2000 to 2010, when the size of local disjointed groups increases and there were links connecting these groups so that they joined to the giant component and disappear in the network of nongiant in those years. Finally, in the year 2015, there are only a few faculties who are left out of the giant robust component.