Research Collaborations in Multidisciplinary Institutions – a Case Study of iSchools

Zhiya Zuo, Xi Wang, David Eichmann, Kang Zhao*
{zhiya-zuo, xi-wang-1, david-eichmann, kang-zhao}@uiowa.edu

The University of Iowa
Iowa City, IA 52242, USA

ABSTRACT
Although closely related, multidisciplinarity and interdisciplinarity are different. The former indicates the co-existence of multiple disciplines while the latter is more about the integration among various areas. As collaboration between researchers from different areas is one of the major approaches for interdisciplinarity, this research investigated whether higher levels of multidisciplinarity in academic institutions are related to more collaborations, especially more interdisciplinary collaborations, among its faculty members. Using U.S. iSchools as a case study, we applied social network analysis and text mining techniques to faculty members’ educational background and publication data, and proposed metrics for multidisciplinarity and collaboration interdisciplinarity. Our analysis results revealed that the multidisciplinarity of an iSchool is actually negatively correlated with the frequency and interdisciplinarity of research collaborations among its faculty members. This finding suggests that having a multidisciplinary environment alone is not sufficient to promote collaborations, nor interdisciplinary collaborations.

Keywords
collaboration; multidisciplinarity; interdisciplinarity; text mining; network analysis

1. INTRODUCTION
Traditionally, science has been characterized by the existence of different disciplines, each of which features clearly defined research domains and well established methodologies. Unlike the conventional collaboration in which researchers worked only with peers with similar educational backgrounds or expertise, however, scientists nowadays often form collaborative teams with diverse expertise to investigate novel and difficult problems that need to be addressed with an interdisciplinary approach [8].

The importance of interdisciplinary research (IDR) in major scientific advances has been widely recognized [1, 8, 24]. Interdisciplinarity does not only occur in emerging areas such as Nanotechnology [27], HIV/AIDS research [2], and Astrobiology [9]. In fact, science, as a whole, has become more interdisciplinary [18].

Another concept closely related to interdisciplinarity is multidisciplinarity, which is about the co-existence of multiple disciplines, whereas interdisciplinarity focuses more on the integration of knowledge from several disciplines into research endeavors [8, 26]. While there are many ways to promote interdisciplinarity in research institutions, such as organizational culture and promotion policies, one way this research is particularly interested in is to create a multidisciplinary institution with researchers from various domains, so that they have more opportunities to form interdisciplinary research teams driven by complex problem-oriented research [24]. While having a diverse group of researchers may increase the chance of their interdisciplinary collaborations, such collaborations can also be challenging due to the heterogeneous nature of different disciplines [11, 25].

Therefore, the goal of this study is to examine the connection between multidisciplinarity of an institution and the interdisciplinarity of collaborations within the institution. Specifically, we utilized text mining and social network analysis techniques to address two research questions: First, does a multidisciplinary environment breed more collaborations? Second, do interdisciplinary collaborations emerge in a multidisciplinary environment? Answers to these questions can help research institutions and funding agencies more effectively promote IDR.

2. RELATED WORK
Multidisciplinarity is essentially a special type of diversity based on researchers’ disciplines or educational backgrounds. Various studies have explored the importance of diversity in organizations or teams. As diversity can be based on many different individual or group attributes, such as race, gender, sexual orientation, and national origins [22], the relationship between diversity and organizational performance has been mixed [25]. Meanwhile, it has been found that diversity in educational backgrounds exerts positive influence on team success [10, 11, 25].

For scientific research, multidisciplinarity is beneficial in several ways. It was found that multidisciplinarity can increase research productivities [17, 29]. Researchers also suggested that multidisciplinarity could generate novel ideas at
the intersection of disciplinary knowledge, and enhance collaborations [21]. However, the picture is not always rosy. Analyzing data collected from questionnaires, Cummings and Kiesler [6] showed that multidisciplinary research teams achieved their goals just as well as those with less disciplines. Similarly, De Saá-Pérez et al. [7] argued that too much educational diversity in a research team has significantly negative influence on the number of published articles. Thus whether multidisciplinarity can foster interdisciplinary remains an open question.

Compared to multidisciplinarity, interdisciplinarity is a more subtle concept. Generally, there are two types of approaches to quantify interdisciplinarity [19]: top-down and bottom-up. Top-down approaches utilize predefined disciplinary categories for journals and analyze their interrelationships via citation records. This suffers from the arbitrary classification schema and fails to capture the context of citations and the extent to which a citation influenced the citing paper. Bottom-up approaches investigate interdisciplinarity by examining texts from research papers. Existing methods include keywords analysis, text clustering, topic models, network analysis, etc. [19, 13, 16, 19, 27].

Specifically, our contributions are three-fold: First, we are the first to evaluate the effect of multidisciplinarity on fostering collaborations, especially interdisciplinary ones. Most previous research tried to connect multidisciplinarity with productivity (often measured by the number of papers), but higher productivity does not mean higher levels of interdisciplinarity. Second, based on topic modeling, we measured multidisciplinarity and interdisciplinarity from the perspective of individuals’ research interests. Although topic modeling was adopted, the study by Nichols [16] still relied on predetermined directorate structure of National Science Foundation (NSF) and is highly limited due to accessibility of the topic model results. A previous study [4] also tried to measure interdisciplinarity based on research interests, but they only used researchers’ affiliations to approximate such interests. Although educational backgrounds, departmental/institutional affiliations, and subject categories of journals can, to some extent, reflect a researcher’s interests, they are too coarse-grained and can be inaccurate since many disciplines, such as information science, and journals, such as PLOS ONE, already feature very diverse research directions. Third, we proposed to measure the interdisciplinarity of collaborations in a social network by analyzing the difference in collaborators’ research interests before the tie was formed, so that we can still capture the degree of interdisciplinarity of a collaboration tie even though a collaborator’s research interest may have changed over time.

3. COLLABORATIONS WITHIN iSCHOOLS

Before examining the relationship between multidisciplinarity and collaborations, we first inspected the frequency of internal collaborations within each iSchool. We calculated the percentage of collaborative papers that represent internal collaborations for each iSchool. We considered a paper to be an internally collaborative one if its co-authors include more than one faculty member in the same iSchool. Figure 1 shows each iSchool’s percentage of internally collaborative papers among all papers. The highest percentage among the 27 iSchools is only 13%, with an average of 5.58%.

We also calculated the percentage of internally collaborative papers for each faculty member in an iSchool and averaged the percentage for all faculty members in the iSchool. The average percentage and its corresponding 95% confidence interval (CI) for each iSchool are shown in Figure 2, along with their sizes (measured by the number of faculty members). Wide CIs for many iSchools suggest high variations in faculty members’ collaborative patterns: there are highly collaborative ones who publish many papers with colleagues, while some prefer working independently or with external collaborators. However, these two percentages are influenced by not only the pervasiveness of internal collaborations, but also on the outcome of collaborations, i.e., the number of papers published. As we mentioned before, such outcomes could be confounded by many other factors after a collaboration was established.

In order to better capture the extent of internal collaborations, we took a social network approach and built an internal collaboration network for each iSchool. In such a network for an iSchool, each node represents a faculty member and there will be an edge between two nodes if they have one or more co-authored publications. Therefore, these internal collaboration networks are unweighted and undirected. While the number of collaborative papers can be used as edge weights, having such weights has little influence on our subsequent analyses in this paper. Such a network will help us understand the collaboration relationship among faculty members in an iSchool and how knowledge and expertise can be exchanged among faculty members. The more connected the network is, the more collaborative an iSchool is.
Figure 2: Avg. Percentage of Internally Collaborative Papers and 95% Confidence Intervals for Individual Faculty Members

For each network, three metrics for network connectedness were calculated. We first used network density to capture the pervasiveness of internal collaborations. It is defined as the ratio between the number of actual edges and the number of possible edges. We also calculated sizes of the largest connected components (LCCs) of these networks, in order to evaluate whether edges in a network can connect many nodes. A connected component of a network is a sub-network where nodes are linked to each other by paths. A LCC is the largest such sub-network with the most nodes. A larger LCC indicates better network connectedness. While 13 iSchools’ LCCs cover at least half of their faculty members, the rest have LCCs below 50%, which means these networks may be isolated into disconnected sub-networks. To measure how close nodes are to each other, we first calculated the popular metric of average path length in the LCC. Specifically, we used the inverse of average path length for simplicity and consistency with other network metrics we used – larger inverse average path length indicates better network connectivity and hence more collaborative.

4. MULTIDISCIPLINARITY AND COLLABORATION

As we have shown in the previous section, internal collaboration networks for the 27 iSchools have various structures. While such structural differences may be due to many reasons, we wanted to see whether they are related to each iSchools’ multidisciplinarity. To examine whether the multidisciplinary environment in an iSchool can foster collaborations, we correlated two multidisciplinarity metrics with metrics of iSchool’s internal collaboration networks using Spearman Rank Correlation.

4.1 Educational Multidisciplinarity

Educational background has been widely used to approximate faculty members’ research directions and expertise. Classification schemas of educational background make it easy and straightforward to examine multidisciplinarity, although such schemas are usually arbitrary. In this case, we first measure the level of multidisciplinarity using this traditional approach by classifying iSchool faculty members’ PhD degrees using a schema of disciplines introduced by [29]. To measure the educational multidisciplinarity for each iSchool, we calculated Shannon Entropy [5], which is a popular metric to evaluate the evenness of distributions. The higher the entropy value is for an iSchool, the more multidisciplinary the iSchool is in terms of faculty educational background distributions.

The middle 2 columns in Table 1 show the correlations between educational multidisciplinarity and metrics of internal collaborations for the 27 iSchools. All the coefficients are negative, albeit non-significant, indicating that higher levels of educational multidisciplinarity are not associated with more internal collaborations.

4.2 Research Multidisciplinarity

Although straightforward and convenient, educational multidisciplinarity was based on a top-down approach that utilizes predefined categories of disciplines. Besides being arbitrary, classifications of educational backgrounds are not good proxies for faculty members’ actual research interests, at least for many in iSchools [29].

Besides educational backgrounds, we measured the level of multidisciplinarity of an iSchool based on how diverse faculty members’ research interests were prior to joining the current iSchool. We focused on research interests prior to joining the current iSchool mainly because hiring is one of the keys to create a multidisciplinary environment. When an iSchool makes a decision on whom to hire, each candidate is represented by her previous research interests. After joining an iSchool, collaborations with peers could change a faculty member’s research interests and confound the measure of research multidisciplinarity.

While the year in which a faculty joined her current iSchool can be inferred from affiliation changes in her papers, we still need to capture each faculty’s research interests over time, in order to get her research interests prior to joining the current iSchool. We decided to adopt topic modeling techniques, which can extract latent topics from texts of faculty members’ publications. We used Latent Dirichlet Allocation (LDA) [3] for this study. Among variations of LDA, dynamic author-topic models [14] can give authors’ topic distributions over time, but topics in these models also change over time. By contrast, we need a set of static topics, because we would compare faculty members’ topic distributions at dif-
Table 1: Spearman Correlation between Internal Collaborations and Multidisciplinarity

<table>
<thead>
<tr>
<th>Metrics of internal collaborations</th>
<th>Educational Multidisciplinarity</th>
<th>Research Multidisciplinarity</th>
</tr>
</thead>
<tbody>
<tr>
<td></td>
<td>Correlation Coeff.</td>
<td>p-Value</td>
</tr>
<tr>
<td>% of internally collaborative papers</td>
<td>-0.160</td>
<td>0.425</td>
</tr>
<tr>
<td>Avg. % of internally collaborative papers for each faculty</td>
<td>-0.032</td>
<td>0.875</td>
</tr>
<tr>
<td>Density of collaboration networks</td>
<td>-0.261</td>
<td>0.188</td>
</tr>
<tr>
<td>Size of LCC (%) in collaboration networks</td>
<td>-0.316</td>
<td>0.109</td>
</tr>
<tr>
<td>Inverse average path length in in the LCC</td>
<td>-0.075</td>
<td>0.712</td>
</tr>
</tbody>
</table>

Table 2: Spearman Correlation between Collaboration Interdisciplinarity and Other Measures

<table>
<thead>
<tr>
<th>Measures</th>
<th>Correlation Coeff.</th>
<th>p-Value</th>
</tr>
</thead>
<tbody>
<tr>
<td>% of internally collaborative papers</td>
<td>-0.005</td>
<td>0.981</td>
</tr>
<tr>
<td>Avg. % of internally collaborative papers for each faculty</td>
<td>-0.003</td>
<td>0.988</td>
</tr>
<tr>
<td>Density of collaboration networks</td>
<td>-0.156</td>
<td>0.436</td>
</tr>
<tr>
<td>Size of LCC (%) in collaboration networks</td>
<td>-0.178</td>
<td>0.375</td>
</tr>
<tr>
<td>Inverse average path length in in the LCC</td>
<td>0.155</td>
<td>0.432</td>
</tr>
<tr>
<td>Educational multidisciplinarity</td>
<td>0.179</td>
<td>0.370</td>
</tr>
<tr>
<td>Research multidisciplinarity</td>
<td>-0.406</td>
<td>0.036</td>
</tr>
</tbody>
</table>

Different time points with each other. If the underlying topics vary over time, we would be unable to compare since different time points were involved. Therefore, we modified LDA by collecting a faculty member’s publications up to a certain year, and used the average topic distribution of these papers to represent the faculty member’s research interests till that year. Titles and abstracts of papers retrieved from Scopus were preprocessed (stop words removal and stemming) before being fed to LDA as inputs. The number of topics was simply set to 20, because we were more interested in the differences in topic distributions among faculty members.

Then we used Kullback-Leibler (KL) Divergence [12] to measure the level of research multidisciplinarity. KL Divergence is used extensively to detect the differences between two probability distributions. For each iSchool, we calculated KL Divergence between all pairs of faculty members’ topic distributions prior to their current iSchool employment, and used the average KL Divergence to measure the iSchool’s research multidisciplinarity. The higher the KL Divergence value is, the more multidisciplinary an iSchool is on research.

The last 2 columns in Table 1 show the correlations between research multidisciplinarity and internal collaboration metrics. Again, there is no statistically significant correlation between research multidisciplinarity and internal collaborations. The difference between the middle 2 and last 2 columns in Table 1 also shows that the two multidisciplinarity measures indeed reflect multidisciplinarity from different perspectives.

5. INTERDISCIPLINARITY OF INTERNAL COLLABORATIONS

The previous section failed to find significant correlation between the level of multidisciplinary and collaborations. The next question we considered relates to the nature of collaborations – Are collaborations interdisciplin ary in a multidisciplinary environment? Edges in our collaboration networks clearly showed collaboration relationships between individual faculty members. However, some collaborations may be between those who work on very similar areas. We believe that interdisciplinary collaborations are those that occurred between faculty members who have different research interests. The more diverse two connected faculty members’ research topics are, the more interdisciplinary their collaboration is.

Such diversity at the dyadic level can be measured by assortative mixing patterns of collaboration networks. Assortativity is the tendency of nodes to connect to similar others in a network [15, 30]. In our study, assortativity of a collaboration network is the likelihood of faculty within the same iSchool with similar topic distributions to co-author papers. To make it more intuitive, we adopt the opposite of assortativity – disassortativity, which quantifies the extent to which dissimilar nodes are connected to each other. The more disassortative an iSchool is, the more interdisciplinary its internal collaborations are.

To calculate disassortativity for a collaboration network, we specified each faculty member’s topic distribution as node attributes. Previously, we used each faculty member’s topic distribution prior to joining her current iSchool to measure the iSchool’s research multidisciplinarity. For disassortativity, however, we wanted to capture research interests of two faculty members before their first collaborative paper – after they co-authored a paper, their research interests inevitably get closer to each other than before.

Topic distributions take the form of vectors, while the traditional method of assortativity computation considers node attributes as scalar. This makes the traditional way infeasible for this study. Instead, we used a method proposed in [28]: For each edge, the cosine distance between nodes’ attributes, i.e., authors’ topic distributions prior to formation of their edge was calculated. The disassortativity of a network will be the average cosine distance over all edges in the network. We then correlated disassortativity, which represents interdisciplinarity of internal collaborations, with internal collaboration metrics in Section 4 and multidisciplinarity measures in Section 5.

Table 2 shows the results. First, having more internal collaborations does not mean these collaborations are interdisciplinary in multidisciplinary environments like iSchools. While no coefficient for collaboration network metrics is sta-
**Table 3:** Top 5 iSchools by Educational Multidisciplinarity, Research Multidisciplinarity, and Collaboration Interdisciplinarity respectively.

<table>
<thead>
<tr>
<th>Educational Multidisciplinarity</th>
<th>Research Multidisciplinarity</th>
<th>Interdisciplinarity of Collaborations</th>
</tr>
</thead>
<tbody>
<tr>
<td>U WASH.</td>
<td>CMU</td>
<td>KENTUCKY</td>
</tr>
<tr>
<td>MICHIGAN</td>
<td>ILLINOIS</td>
<td>UNT</td>
</tr>
<tr>
<td>FSU</td>
<td>PITT</td>
<td>PSU</td>
</tr>
<tr>
<td>TEXAS</td>
<td>UCLA</td>
<td>PITT</td>
</tr>
<tr>
<td>PITT</td>
<td>BERKELEY</td>
<td>TEXAS</td>
</tr>
</tbody>
</table>

**Table 4: OLS Regression Results**

<table>
<thead>
<tr>
<th>Variable</th>
<th>Coef.</th>
<th>Standard Dev.</th>
<th>p-Value</th>
<th>95% Confidence Interval</th>
</tr>
</thead>
<tbody>
<tr>
<td>Carnegie Classification</td>
<td>0.037</td>
<td>0.025</td>
<td>0.158</td>
<td>(-0.016, 0.090)</td>
</tr>
<tr>
<td>Num. of faculty</td>
<td>-0.040</td>
<td>0.022</td>
<td>0.082</td>
<td>(-0.086, 0.006)</td>
</tr>
<tr>
<td>Avg. num. of papers</td>
<td>0.060</td>
<td>0.024</td>
<td>0.021</td>
<td>(0.010, 0.111)</td>
</tr>
<tr>
<td>Network density</td>
<td>-0.083</td>
<td>0.024</td>
<td>0.002</td>
<td>(-0.134, -0.035)</td>
</tr>
<tr>
<td>Educational multidisciplinarity</td>
<td>-0.060</td>
<td>0.017</td>
<td>0.749</td>
<td>(-0.040, 0.030)</td>
</tr>
<tr>
<td>Research multidisciplinarity</td>
<td>-0.049</td>
<td>0.016</td>
<td>0.006</td>
<td>(-0.083, -0.016)</td>
</tr>
<tr>
<td>Intercept</td>
<td>0.283</td>
<td>0.073</td>
<td>0.001</td>
<td>(0.131, 0.435)</td>
</tr>
</tbody>
</table>

Statistically significant, suggesting most of the collaborations were not between scholars with different interests. Second, a more multidisciplinary environment cannot guarantee that collaborations are more interdisciplinary. Research multidisciplinarity is negatively and significantly (p-value < 0.05) correlated with interdisciplinarity of internal collaborations. This suggests that in an iSchool with faculty members who work on very different topics, their internal collaborations tend to be less interdisciplinary on average. Although the relationship between diversity in educational backgrounds and interdisciplinarity of collaborations is positive, it is not significant. As an example to show the discrepancies between multidisciplinarity of iSchools and interdisciplinarity of their internal collaborations, we listed top 5 iSchools by educational multidisciplinarity, research multidisciplinarity, and collaboration interdisciplinarity in Table 3. Only the iSchool at the University of Pittsburgh is ranked as top 5 by all the three measures.

To better understand the role of multidisciplinarity in promoting interdisciplinary collaboration, we also ran a regression analysis using ordinary least square (OLS) regression. Each iSchool’s interdisciplinarity of internal collaborations is the dependent variable while their educational multidisciplinarity and research multidisciplinarity are independent variables. We also controlled iSchool size (measured by the number of faculty members), average number of papers per faculty, the density of internal collaboration networks, and university classification according to the Carnegie Classification of Institutions of Higher Education [20]. Results in Table 4 further confirmed our earlier finding – educational multidisciplinarity is a non-significant predictor of interdisciplinarity of collaborations, while research multidisciplinarity has significant negative effect.

### 6. CONCLUSIONS

Using social network analysis and text mining techniques, this research consists of a three-step analysis on the relationship between multidisciplinarity and interdisciplinary collaborations in U.S. Schools. We first provided an overview of internal collaborations within each iSchool. Characteristics of each iSchool’s internal collaboration network were then correlated with its level of multidisciplinarity, which is based on educational and research diversities, to examine the correlations between multidisciplinarity and collaborations. Finally, we applied topic-based assortativity analysis to examine whether collaborations in multidisciplinary environments are indeed interdisciplinary.

As the results suggest, different iSchools feature different levels of internal faculty collaborations. However, neither educational nor research multidisciplinarity was significantly correlated with collaborativeness. In addition, we found a negative correlation between the research multidisciplinarity in an iSchool and the interdisciplinarity of its faculty collaborations.

Admittedly there are other factors that could affect the establishment of collaborations and the interdisciplinarity of collaborations within a research institution. For example, a possibility is that iSchools, as a newly founded discipline, need more time to create chemistry among their faculty members. Nevertheless, our analysis at least suggested that a multidisciplinary institution alone does not necessarily lead to a more collaborative environment, or collaborations that are more interdisciplinary. More coordination, management, and incentives may be needed to fully exploit the benefits of institutional multidisciplinarity in stimulating collaboration and interdisciplinarity. The lack of significance of most results may also indicate the relationship between multidisciplinarity and collaboration interdisciplinarity is not simply monotonic but more sophisticated. Further analysis is needed to explore the nature of this relationship.

This study has its limitations too. First, our analysis was limited to 27 iSchools in the U.S. and the results may not be applicable to all research areas. Second, the study analyzed empirical data using correlation and regression analyses, and thus cannot claim or infer a causal relationship between multidisciplinarity and interdisciplinary collaborations. Third, by employing co-authorship as a proxy for collaboration, we assumed that all the co-authors have the same level of interests to the paper’s topic, which might not be true in the reality. Finally, our analyses did not take the impact of publications into consideration.
7. REFERENCES


