The State and Evolution of U.S. iSchools
--from Talent Acquisitions to Research Outcome

Zhiya Zuo[1], Kang Zhao[2*], and David Eichmann[3]

[1]Interdisciplinary Graduate Program in Informatics, 3087 Main Lib, The University of Iowa, Iowa City, IA 52242. zhiya-zuo@uiowa.edu.

[2]Department of Management Sciences, S224 PBB, The University of Iowa, Iowa City, IA 52242. kang-zhao@uiowa.edu. *Corresponding author. Phone: 319-335-3831. Fax: 319-335-0297.*

[3]School of Library and Information Science, 3086 Main Lib, The University of Iowa, Iowa City, IA 52242. david-eichmann@uiowa.edu

ABSTRACT

The past two decades have witnessed the emergence of information as a scientific discipline and the growth of information schools around the world. We analyzed the current state of the iSchool community in the U.S. with a special focus on the evolution of the community. We conducted our study from the perspectives of acquiring talents and producing research, including the analysis on iSchool faculty members' educational backgrounds, research topics, and the hiring network among iSchools. Applying text mining techniques and social network analysis to data from various sources, our research revealed how the iSchool community gradually built its own identity over time, including the growing number of faculty members who received their doctorates from the study of information, the deviation from computer science and library science, the rising emphasis on the intersection of information, technology and people, and the increasing educational and research homogeneity as a community. These findings suggest that iSchools in the U.S. are evolving into a mature and independent discipline with a more established identity.

Keywords

Information schools, faculty hiring, network analysis, topic modeling, evolution of disciplines.

INTRODUCTION

The term "iSchools," or "information schools," refers to a group of university academic units that are dedicated to the study of information. Many of the iSchools started as library sciences programs, but expanded their focus to information studies. While there were many milestone events in the history of the iSchool movement, which can trace back to the 1960s, the iSchool organization stated in its website (ischools.org) that the start of iSchools dates back to the formation of "Gang of Three" in 1988, which is comprised of University of Pittsburgh's School of Library and Information Science (now the School of Information Science), Syracuse University's School of Information Studies, and Drexel University's College of Information Science and Technology. The "iSchool movement" accelerated as the world moved into the information age — the iSchool community saw fast growth in the 21st century, with new members from outside the library sciences discipline (e.g., the iSchools of Penn State University and Georgia Institute of Technology). As of 2015, the iSchool community has 65 members all over the world, along with a conference — the "iConference" — held annually since 2005.

Despite the fast growth in membership, the iSchool community is still small and young compared to other well-established academic communities such as sociology and computer science (Dillon, 2012). Also, although iSchools share common research interests in "the relationship between information, technology, and people" (excerpt from http://ischools.org/about/history/motivation/), the identity of iSchools as a discipline is still unclear to many people outside, and even inside, iSchools (Cronin, 2005).

Thus our research addressed the question of whether the community is getting more mature and established as a discipline. Our focus was therefore not only on the current state of iSchools, but also on how the community evolved over time. Using iSchools in the U.S. as a sample for the whole iSchool community, this research applied text mining and network analysis techniques to analyze several key perspectives of iSchools—from acquiring talents from various disciplines, to producing research. We hope our findings can help various stakeholders better understand and guide the iSchool movement, including iSchool administrators and scholars, funding agencies, and potential employers of iSchool students.

RELATED WORK

In the literature, several papers have introduced or described the history and characteristics of iSchools as an emerging discipline. Olson and Grudin (2009) expressed their anticipation of the promising future of information with the rapid growth of computing and digital technologies. Dillon (2012) talked about the key attributes of iSchools, including intellectual coverage, interdisciplinarity and research commitment. Cronin (2005) stated his concern regarding the identity crisis of these newly formed or transformed academic units due to the fuzzy definition of information compared to mature disciplines such as law and business.

As educational institutions within universities, the education components of iSchools have been analyzed. For example, Subramaniam and Jaeger (2010) studied the syllabi of courses in the American Library Association (ALA) accredited Master of Library and Information Science (MLIS) programs at iSchools in North America. Seadle and Greifeneder (2007) proposed a unique iSchool curriculum using an anthropological approach. Wu, D., et al. (2012) focused on five facets of iSchools' graduate education: mission and vision, education program design, core course design, research interests of doctoral students and careers of graduate students. The results showed that the core of iSchools is indeed the exploration of relationships between people, information and technology.

In addition to teaching, iSchools faculty members are also active in research. Thus, several studies also analyzed the educational backgrounds and research of iSchool faculty members. Wiggins and Sawyer (2012) measured the intellectual diversity of iSchools by analyzing the PhD programs where iSchool faculty members received their doctorates. Based on the educational backgrounds of faculty members in each iSchool, they also divided iSchools into four groups: Computational, Library & Information, Sociotechnical, and Niche. Specifically, the Computational cluster contains iSchools in which 60% to 80% of the faculty graduated from Computer Science and related disciplines; the Library & Information cluster represents iSchools with over 50% of the faculty from Library, Information and Humanities; iSchools in the Sociotechnical group have faculty members from Computing, Social Sciences, and some Library and Information; Niche is a cluster of iSchools that have diverse faculty educational backgrounds and thus cannot be well represented by one or two discipline categories. Luo (2013) studied the interdisciplinarity of iSchools by analyzing iSchool faculty members' online profiles and survey data collected from 135 iSchool faculty members. Chen (2008) studied the identity of iSchools by visualizing keywords of iSchool faculty's publications. Wu, D., et al. (2012) gauged the interdisciplinarity of iSchools via the Web of Science journal classification of where iSchool faculty members publish to extract research areas of iSchools. Zhang, et al. (2013) identified the dominant areas among iSchools as Information, Computing, Management & Policy, and Library in both faculty educational backgrounds and in the manual classification of journals where faculty members published.

Holmberg, Tsou, and Sugimoto (2013) collected iSchool faculty members' research interests and conducted co-word analysis. They found that while the majority of iSchool faculty members are still concerned about traditional information topics (e.g., bibliometrics, information retrieval, and information seeking behavior, etc.), there are increasing numbers of interdisciplinary areas (e.g., data mining, artificial intelligence, social media, etc.). Most of the research has concluded that iSchools are multi- and inter-disciplinary.

Network analysis has also been used in studies of different disciplines, including iSchools. For example, Clauset, et al. (2015) found evidence of social inequality and hierarchical structures in faculty hiring networks in computer science, business, and history. From a different perspective, Yu (2013) analyzed the collaboration network of iSchools using iConference publication data and the Twitter network among official Twitter accounts of various iSchools. Similar to Zhang et al. (2013), the author discovered that there was very little collaboration between iSchools as well as between iSchools and other disciplines. However, the publication data was only from iConference between 2008 and 2013, which only partially represents scholar activities in iSchools. Wiggins, et al. (2006, 2008) compared the hiring network of iSchools' to that of 29 Computer Science (CS) departments. They found that PageRank, especially weighted PageRank, is highly correlated with the current ranking of CS programs by U.S. News and World Report (USNWR). It was revealed that, compared to CS, iSchools are more diverse in terms of hiring faculty from different areas.

While all these studies provided valuable insights for the iSchool community, there are some missing pieces that our research would like to address. *First*, no study has explored the evolution of the community, which is important for an emerging discipline to understand how it develops over time. Most studies analyzed a snapshot of the community based on aggregated data up to the time of the study. Although Wiggins and Sawyer (2012) and Zhang et al. (2013) have studied the temporal changes of the iSchool community, they only examined the number of faculty and the number of publications. *Second*, classifications of educational background and journal are too coarse-grained as measures for research, as they fail to reflect precisely what a faculty member works on. For example, a researcher with a PhD from an iSchool may study metadata, information policies, or data mining. Similarly, a paper published in a journal such as JASIST or PLOS ONE may be about data mining, user interface design, or scientometrics. Manual classifications of journal papers could be more accurate than the Web of Science classification, but do not scale to volumes of papers. To precisely capture the research topics of iSchool faculty members, we need more fine-grained and automated analysis of text from their publications. Research interest keywords provided by faculty members (Holmberg et al. 2013) may be too coarse-grained. They did not accurately reflect the level of interests or engagement a faculty member has for each of the listed areas. Although the titles and abstracts of iSchool scholars' publications would offer better representations of their research interests, the study of Chen (2008) was limited to the lexcial level without exploring the latent structures between keywords and phrases. Topic modeling techniques have been widely used to detect "latent" topics from documents, because they provide an abstract representation of what a document is about. They are also better at handling synonyms (phrases that are semantically close, such as "text" and "document"), and polysemous words that may correspond to more than one research area (e.g., "computer" and "network"), etc. While Sugimoto et al. (2011) revealed trending topics in dissertations from Library and Information Science programs using topic modeling techniques, the topic modeling was done for dissertations in each decade. As a result, topics generated for one decade are different from those generated for another decade. This makes it difficult to keep track of each topic's popularity over time, unless domain experts examine the keywords for each topic in each decade, and manually identify the correspondence or equivalence between two topics in different decades. Third, academic units that were analyzed are the entirety of iSchools listed on the iSchool.org member directory. However, while the iSchool membership is often for schools or colleges, some member schools contain departments that do not focus on information studies. For example, the iSchool at UCLA (namely the Graduate School of Education and Information Studies) consists of the Department of Information Studies and the Department of Education, whereas the latter focuses more on education than information. Similarly, the iSchool at UC Irvine is the School of Information and Computer Science and has Informatics, Computer science, and Statistics departments. For these iSchools, the iSchool membership is in fact for one of the departments only. Thus for iSchools that have departments, it becomes necessary to focus our analysis on their information-related departments in order to avoid noise from other departments within the same administrative unit. Back to the example of UCLA and UC Irvine, we only collected data for faculty members from UCLA's Department of Information Studies and UC Irvine's Informatics department.

DATASET

In this study, we used 27 U.S. members of the iSchool community (http://ischools.org/members/directory/) as a sample. The iSchool movement originated in the U.S. and most of the early iSchool members are U.S. institutions. Although the iSchool community has been growing during the past a few years with more non-U.S. members, the U.S. still has the most members in the community. Among the current 25 invited iCaucus members, who "represent institutional leadership in the field" (http://ischools.org/members/icaucus-members/), 19 are from the U.S. Thus, we believe our analysis, even though restricted by geographical boundaries, can still provide insights to the understanding of the whole community. As mentioned earlier, we further limited the academic units of interest to departments that focus on information-related subjects and excluded non-information-related units housed in iSchools. Table 1 lists the 27 U.S. iSchools in our study.

University	Academic unit	University	Academic unit
Univ. of California, Berkeley	School of Information	Rutgers Univ.	Library and Information Science Dept.
Carnegie Mellon Univ. (CMU)	Heinz College	Simmons College	School of Library and Information Science
Drexel Univ.	College of Computing & Informatics (only the former College of Information Science & Technology)	Syracuse Univ.	School of Information Studies
Florida State Univ.	School of Information	Univ. of Tennessee, Knoxville	School of Information Sciences
Georgia Inst. of Tech. (Gatech)	School of Interactive Computing	Univ. of Texas at Austin	School of Information
Univ. of Illinois at Urbana-Champaign	Graduate School of Library and Information Science	Univ. of California, Irvine (UCI)	Department of Informatics
Indiana Univ.	School of Informatics and Computing	Univ. of California, Los Angeles (UCLA)	Dept. of Information Studies
Univ. of Kentucky	Library and information science	Univ. of Maryland, Baltimore County (UMBC)	Dept. of Information Systems
Univ. of Maryland, College Park (UMD)	College of Information Studies	Univ. of North Carolina at Chapel Hill (UNC)	School of Information and Library Science
Michigan State Univ.	Dept. of Media and Information	Univ. of North Texas (UNT)	Dept. of Library & Information Sciences
Univ. of Michigan	School of Information	Univ. of Wisconsin-Milwaukee (UWM)	School of Information Studies
Univ. of Missouri	School of Information Science & Learning Technologies	Univ. of Washington	Information School
Penn State Univ.	College of Information Sciences & Technology	Univ. of Wisconsin-Madison (WISC)	School of Library & Information Studies
Univ. of Pittsburgh (Pitt)	School of Information Sciences		

Table 1. The list of 27 U.S. iSchools.

For all full-time tenured or tenure-track faculty members from the 27 iSchools, we collected their educational backgrounds (including PhD programs and institutions) and titles from their personal or schools' websites.

Associate and full professors were considered as senior faculty while assistant professors as junior faculty. Those with titles such as emeritus, adjunct, or visiting professors were not included.

Our final dataset consists of 708 faculty members, including 201 assistant, 238 associate, and 269 full professors. Based on the taxonomy by Zhang et al. (2013), PhD programs where iSchool faculty members received their doctorates were classified into nine categories: Communication, Computing, Education, Humanities, Information, Library, Management & Policy, Science & Engineering, and Social & Behavioral. The detailed classification of PhD programs is in Table 2.

Discipline category	**PhD programs**
Communication	Media and Mass Communication, Journalism
Computing	Computer Science, Electrical Engineering, Mathematics, Computer Engineering
Education	Education, Learning Technology
Humanities	History, English, Philosophy, Literature, Music, Geography, Art, Anthropology
Information	Information Science, Information Studies, Information Transfer, Informatics
Library	Library Science, Information and Library Science
Mgmt&Policy	Business Administration, Management, Policy, Economics, City & Regional Planning, Public Administration
Sci.&Eng.	Life Sciences, Physics, Statistics, Engineering (not Electrical or Computer), Biology
Social&Behavioral	Psychology, Sociology, Law, Social Sciences, Linguistics, Political Science, Government

Table 2. The classification of PhD programs.

We used Elsevier's Scopus database as the source for publications of iSchool faculty members. According to its website (http://www.elsevier.com/solutions/scopus/content), Scopus covers broad subject areas, including science, mathematics, engineering, technology, health and medicine, social sciences, and arts and humanities. It indexes not only journal publications, but also conference papers and book chapters. Although an Elsevier commercial product, Scopus includes publication data from other major publishers, such as Springer, Nature Publishing Group, AAAS, BMJ, etc. As of February 2014, Scopus possesses over 54 million records, from over 20,000 journals, 18,000 conferences, 367 trade journals, more than 400 book series, etc. We believe Scopus' coverage of subject areas and publication types are suitable for the iSchool community, since iSchool faculty members work on many different areas and many commonly publish in conference proceedings in addition to journals.

Titles, abstracts, authors (along with their affiliations), publication dates and types of each iSchool faculty member's papers were retrieved from Scopus APIs, based on each author's name and affiliation. To improve the quality of data, we also manually inspected the dataset to exclude some obvious non-iSchool scholars whose names and affiliations are close to those of any iSchool faculty member.

Using Scopus APIs, we retrieved 26,491 papers authored by iSchool faculty members. Figure 1 depicts the distribution of publication types, with the majority being conference and journal papers. The high percentage of conference papers also highlights the importance of including conference proceedings besides journals when analyzing research in the iSchool community.

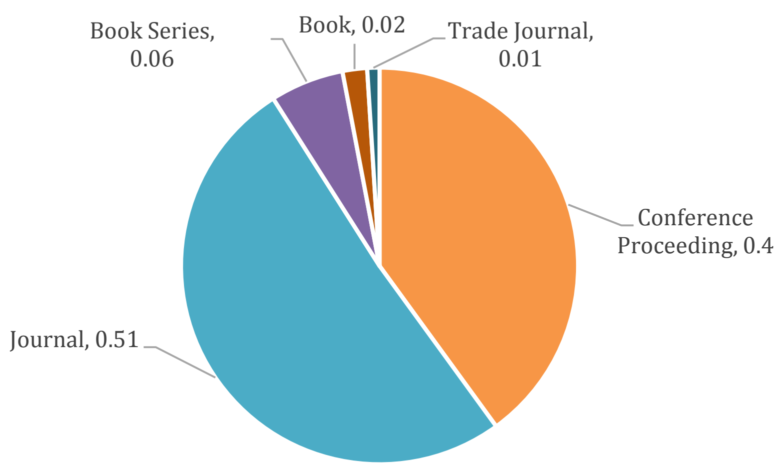

Figure 1. The distribution of publication types

ANALYSIS OF EDUCATIONAL BACKGROUNDS

Among the 708 iSchool faculty members in our dataset, the educational backgrounds of 5 are unavailable online. Based on the 703 faculty members, we analyzed iSchool faculty educational background distributions at the levels of both the community and individual iSchools.

Community-level Distribution

As shown in Table 3, computing, information and library dominate the faculty makeup at the community level. Overall, computing is the top area that produced the most iSchool faculty members, followed by information. Further, we compared the education distributions between senior and junior faculty members. A higher proportion of junior faculty members comes from computing, information and management & policy. Information saw a great increase from 15.48% to 26.87%. In other words, even though the iSchools are still hiring from other disciplines, they have started to hire more faculty from programs that focus on the study of information. Hiring from peer programs is common in most well defined disciplines. Such patterns in iSchools' hiring suggest that iSchools are gradually straying away from a mix of various disciplines and becoming a more independent discipline that can produce qualified faculty members for peer programs in the same discipline.

	Com-muni-cation	Computing	Educa-tion	Humanities	Infor-mation	Library	Mgmt &Policy	Science &Eng.	Social& Behavioral	Entropy
All	3.55%	30.07%	3.83%	6.38%	18.72%	13.62%	9.79%	5.96%	8.09%	**2.83**
Senior	3.77%	29.56%	3.97%	7.34%	15.48%	14.29%	9.52%	6.75%	9.33%	**2.88**
Junior	2.99%	31.34%	3.48%	3.98%	26.87%	11.94%	10.45%	3.98%	4.98%	**2.65**

Table 3. Faculty educational background distributions and entropies for the iSchool community.

We further utilized information entropy (Shannon 1948) to measure the diversity of faculty educational background. Information entropy for an iSchool's education distribution is calculated as $\sum_i -p_i log p_i$ where p_i, in this context, is the proportion of the ith discipline. Higher entropy is a sign of more even distribution, indicating higher diversity, and vice versa. As the right-most column in Table 3 shows, the educational background of junior faculty members is less diverse than that of senior faculty members. With more junior faculty members hired from programs that focus on the study of information, the whole community also gets less diverse. In other words, iSchools are now more likely to hire graduates from academic programs that focus on the study of information.

Distributions for Individual iSchools

We also analyzed faculty educational backgrounds for each individual iSchool and found patterns that are similar to the whole community—iSchools have been hiring more faculty members from programs that focus on the study of information. For example, by comparing faculty educational background distributions between senior and junior faculty members in each iSchool, we found that only 6 out of the 27 iSchools hired more junior faculty members from library sciences. On the other hand, 17 iSchools saw increases in faculty hiring from information.

Table 4 lists the highest 3 and lowest 3 iSchools by entropy values of faculty educational background distributions. The highest 3 iSchools are all early members of the iSchool community and have established a wellbalanced multi-disciplinary faculty body. The lowest 3 iSchools are more focused on a specific area: Missouri on education; UNT on library and information; and Georgia Tech on computing.

	Communi-cation	Computing	Education	Humanities	Information	Library	Mgmt &Policy	Science &Eng.	Social& Behavioral	Entropy
Washington	3.45%	13.79%	6.90%	3.45%	24.14%	27.59%	10.34%	6.90%	3.45%	**2.775**
Michigan	3.77%	28.30%	3.77%	7.55%	11.32%	7.55%	20.75%	0.00%	16.98%	**2.696**
Texas	5.26%	15.79%	0.00%	15.79%	10.53%	26.32%	0.00%	10.53%	15.79%	**2.676**
Missouri	0.00%	0.00%	50.00%	0.00%	21.43%	28.57%	0.00%	0.00%	0.00%	**1.493**
UNT	0.00%	0.00%	4.76%	4.76%	57.14%	33.33%	0.00%	0.00%	0.00%	**1.408**
Georgia Tech	2.33%	83.72%	0.00%	9.30%	0.00%	0.00%	0.00%	4.65%	0.00%	**0.865**

Table 4. Faculty educational background distributions for the highest 3 (in shade) and lowest 3 iSchools by entropy.

Clustering Analysis

Clustering based on faculty educational background has been used to discover different types of iSchools by putting iSchools with similar faculty educational background distributions into the same group (Wiggins and Sawyer, 2012). However, the goal of our clustering is different. We focused on the temporal dynamics of clusters, namely how the number of clusters and the overall fragmentation of the iSchool community changed over time. Here, we represented each iSchool with its faculty educational background distribution vector and applied an affinity propagation algorithm (Frey and Dueck, 2007), which can automatically choose an optimal number of clusters for the data. The Silhouette Coefficient (Rousseeuw, 1987) was used to measure the quality of grouping (a.k.a., clustering) instances based on similarity. For an instance i, which corresponds to an individual iSchool in our study, its silhouette is defined as $s(i) = \frac{b(i)-a(i)}{max\{b(i),a(i)\}}$ where $a(i)$ measures how dissimilar i is to other iSchools in the same cluster by calculating the average distance from i to other instances in the same cluster, and $b(i)$ indicates how dissimilar i is to iSchools in the closest neighboring cluster. The closest neighboring cluster of i is the cluster, whose instances have the lowest average dissimilarity with i. The Silhouette Coefficient is the average silhouette of all iSchools and bounded within [-1, 1]. Values closer to -1 indicate incorrect assignments, as $b(i) \ll a(i)$ means that iSchools are more similar to those in other clusters. Values near 1 indicate proper clustering (average intra-cluster distance a(i) is close to zero). In other words, better separations among clusters that are more compact will lead to a higher Silhouette Coefficient.

Table 5 shows the clustering results when we considered all faculty and only senior faculty members' educational background distributions. With the addition of junior faculty members to the iSchool community, the number of clusters drops from 5 to 4 and the Silhouette Coefficient decreases by 15%. The lower number of clusters and the decrease of Silhouette Coefficient suggest that iSchools become more similar to each other and it gets more difficult to group them into well-separated yet compact clusters. In other words, the iSchool community is getting less fragmented and more close-knit in terms of faculty training.

	Number of Clusters	Silhouette Coefficient
Senior only	5	0.451
All faculty	4	0.383

Table 5. iSchool clustering results for both all and senior faculty members based on educational backgrounds.

ANALYSIS OF RESEARCH TOPICS

As we mentioned earlier in this paper, using faculty educational backgrounds cannot accurately capture one's research interests. This was also echoed by Zhang et al. (2013), who manually coded the themes of research in iSchool faculty members' journal papers. Instead of manual coding, we adopted an automated topic modeling technique – Latent Dirichlet Allocation (LDA), a generative model used extensively for topic discovery (Blei, Ng, and Jordan, 2003). The input for LDA is the text corpus of titles and abstracts of iSchool faculty members' papers. The output is a group of topics, each represented by a probabilistic distribution over words. Those words with high probabilities on a topic are considered representative keywords for this topic. Similarly, each document is assigned a probabilistic distribution over all the topics.

In our analysis, we set the number of topics to be 20. Our interpretations of topics with top 5 keywords for each topic are shown in Table 6. The topics discovered by LDA cover the majority of the diverse research by iSchool faculty members. Most of the topics are related to the study of information, and many of them are indeed areas involving the relationship between information, technology, and people. The topic in the last row of the table (labeled as “Others”) makes little sense for iSchools and may have been caused by author name ambiguity in the Scopus database.

Topic Interpretation	Representative Keywords
IT for collaboration & communication	information, technology, communication, practice, collaborate
Software and system engineering	design, system, develop, software, process
Information privacy and policy	privacy, policy, government, market, internet
Social networks and media	social, community, online, media, network
Machine learning and data mining	measure, perform, test, predict, data
Information retrieval and recommendation	information, user, search, web, query
Computing infrastructure	application, system, service, compute, distributive
Cyber-security and computer networks	network, security, scheme, attack, node
Digital library and library science	library, digital, public, author, collection
User interface and experience	user, design, interface, interact, mobile
Text mining	document, retrieve, text, term, topic
Algorithms	algorithm, optimal, time, space, efficiency
Data storage and visualization	data, visual, analysis, information, collect
Education and learning technology	learn, student, education, compute, school
Robotics and cognitive Systems	robot, human, agent, game, behavior
Health informatics	health, patient, care, medical, information
Programming languages	program, language, type, function, structure
Spatial and multimedia data analytics	image, location, spatial, video, object
Bioinformatics	sequence, protein, gene, genome, structure
Others	simulation, energy, measure, process, structure

Table 6. Topics and corresponding keywords discovered from iSchool publications. Topics are in the order of descending prevalence.

Topic Evolution

Next, we examined how topics evolved over time. Our longitudinal analysis on topic evolution was conducted on annual basis from 1988, when the “Gang of Three” was formed, to 2014. We calculated the proportion of publications for each topic in each year over the 27 years. Specifically, we calculated each year’s topic distribution by averaging topic distributions of all papers that were published in that year. Each topic’s proportions across the 27 years were then used to show its evolution during this time span. For each trajectory of topic, we fit a linear regression and calculated the slope of the trajectory. As the original slope for each topic trend is small, we applied a linear transformation to standardize the values into a standard normal distribution. For the 20 slopes for 20 topic trajectories, we calculated the average slope $\bar{y}$ and sample standard deviation s. Each slope was then standardized using $y'_i = \frac{y_i - \bar{y}}{s}$ where y'_i is the standardized value of the i-th topic trajectory's slope y_i. The 7 rising topics all have standardized slopes above 0.5 and the 3 declining topics have slopes below -1. By contrast, slopes for other topics lie in the range of [-0.4, 0.2]. With much lower magnitudes in slopes, they were recognized as stable topics. We found that research topics in iSchools are indeed evolving over time. To simplify our visualization, we show curves only for selected topics that feature obvious rising (Figure 2) and declining trends (Figure 3).

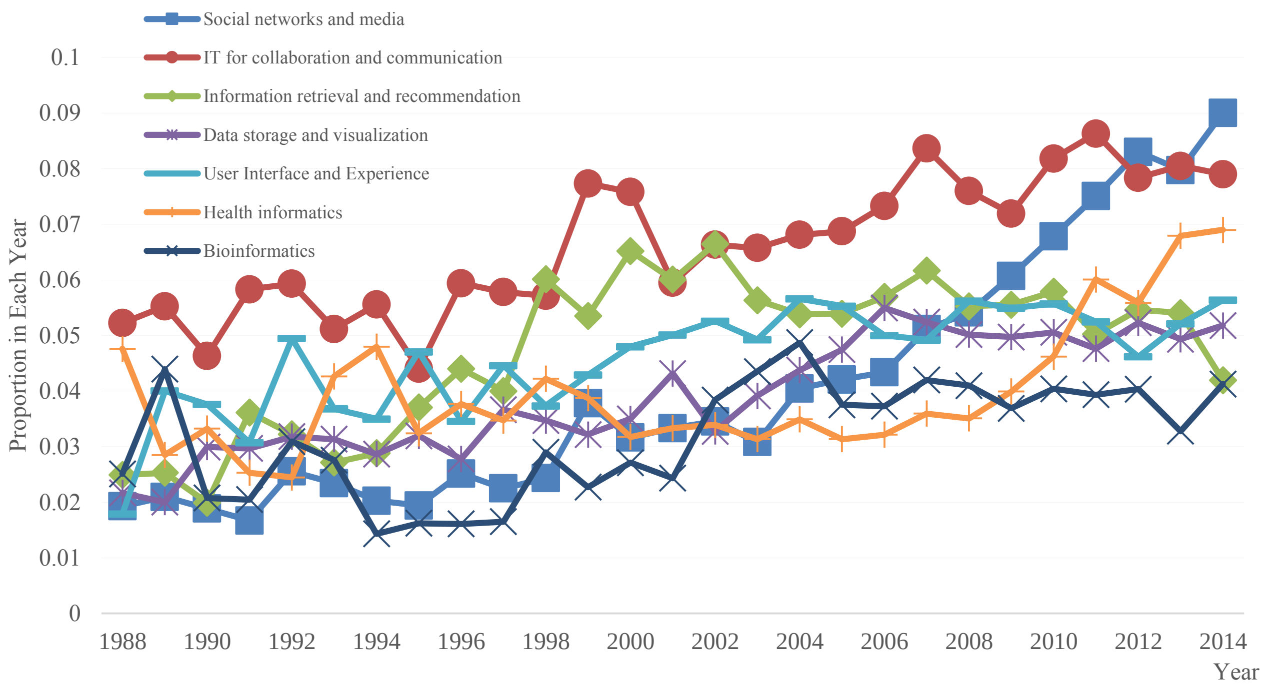

Figure 2. Rising iSchool research topics over 27 years.

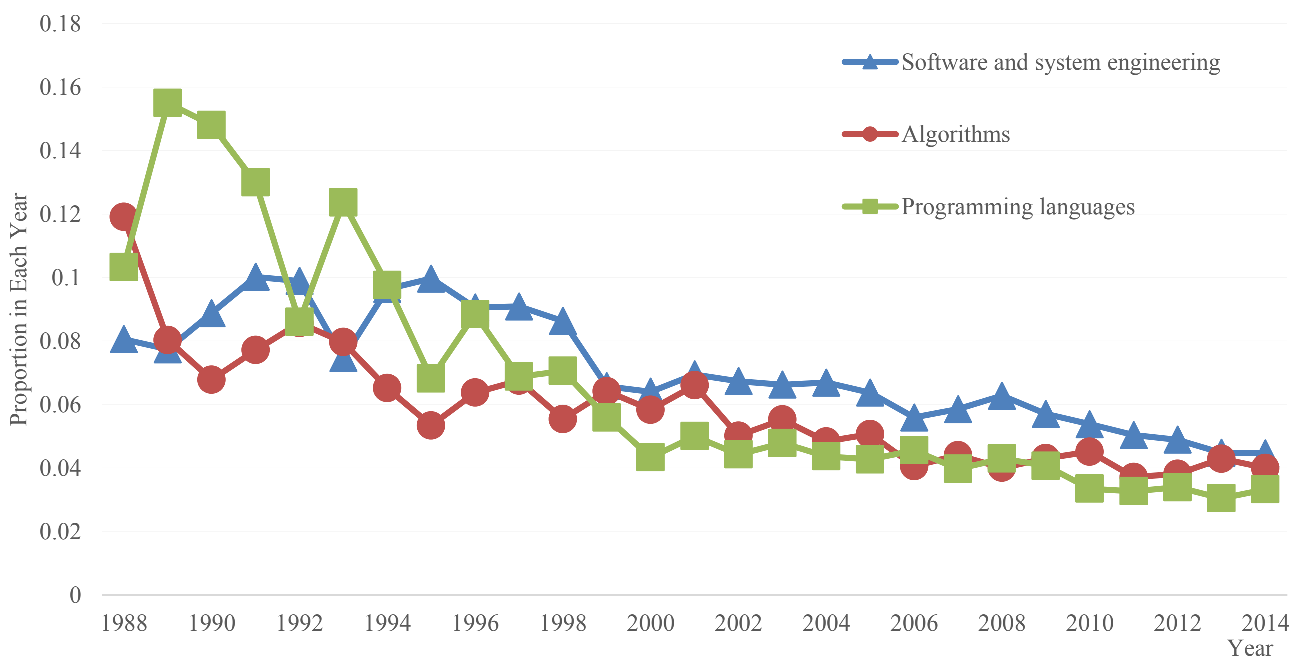

Figure 3. Declining iSchool research topics over 27 years.

Again, we see that research areas that explore the relationships between information, technology and people are on the rise. For example, information technology for communication and collaboration, social network analysis and media, user interface and experience, are all about how people use and interact with information and technologies. Meanwhile, among the three declining topics, two are typical computer science areas (algorithms and programming languages) and the third (software and system engineering) is also closely related to computer science. Although some of the topics also draw more attention from computer scientists in recent years, such as social network and media, and data storage and visualization, their focus is more on algorithms, which are on the decline among iSchool researchers. Therefore, the iSchool community has started to part ways with traditional computer science research. This is additional evidence that iSchools are establishing their own identity as a discipline that studies the Information-Technology-People triangle. We also found that such trend was consistent when we chose different numbers of topics for topic modeling (e.g., 30, 50, and 100).

The Similarity among iSchool Research over Time

In our above analysis, we showed that iSchools are getting less diverse in terms of faculty members' educational backgrounds. Does the trend of becoming more homogeneous exist in research topics as well? To address this question, we measured inter-iSchool similarity in terms of research topics. For each iSchool, we pooled all papers by its faculty members. Then for each year, we extracted the topic distributions for an iSchool based on all papers published by its faculty in that year. In the end, each iSchool has a topic distribution vector for each year from 1988 to 2014. On an annual basis, we calculated pair-wise cosine similarities between topic distribution vectors of all possible pairs of iSchools. The cosine similarity between iSchool i and j is defined as $sim(i,j) = \frac{\vec{t_i} \cdot \vec{t_j}}{\|\vec{t_i}\| \|\vec{t_j}\|}$ where $\vec{t_i}$ is the topic distribution vector of iSchool i. For 27 iSchools, that means 27*26/2=351 pairs of iSchools.

Figure 4 shows the temporal trend of average pair-wise similarities between iSchools' research topic distributions across years. The error bars show the upper and lower bounds of the 95% confidence intervals. The trend is very clear: iSchools are becoming more similar to each other in terms of what types of research they do. This is yet further evidence that the iSchool community is emerging as a more mature discipline with more common research interests.

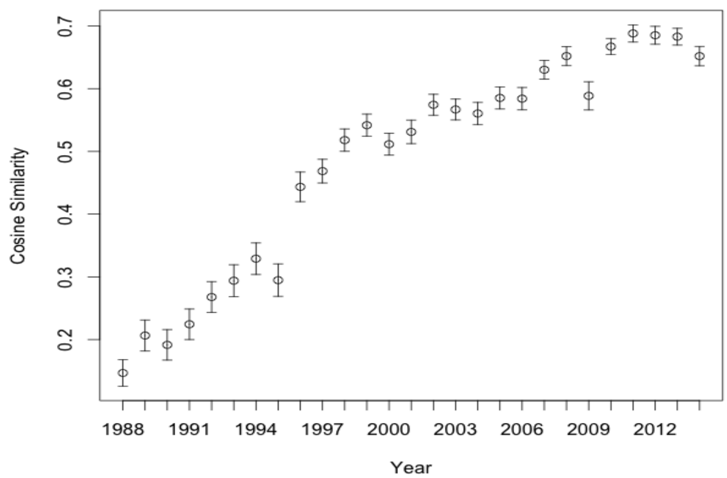

Figure 4. Avg. pair-wise similarity in iSchool research topics (vertical bars indicate 95% confidence intervals).

THE ISCHOOL HIRING NETWORK

The analysis on faculty educational backgrounds and research topics treated each iSchool as individuals, while they are in fact connected. Because iSchools are hiring more faculty members from peer programs, we built a hiring network with only iSchool-to-iSchool connections to better understand the relationships and interactions between iSchools. We chose hiring because iSchool scholars do not often collaborate with peers from other iSchools (Zhang et al., 2013; Yu, 2013). Consequently, a collaboration network among iSchools will be very sparse. We investigated the network by examining its topology together with each individual iSchool's faculty educational backgrounds and research topics.

Our hiring network has 27 nodes, each representing an iSchool in our study. An edge between a source and a target node means that the source iSchool hired doctorate student(s) from the target iSchool (as shown in Figure 5; sizes of nodes are proportional to their degrees and node colors indicate the network cluster they belong to). Note that this hiring network only considers faculty members who graduated from U.S. iSchools. An iSchool faculty member who received his/her PhD from non-iSchool programs will not be reflected in this iSchool-to-iSchool hiring network, even though the university from which he/she graduated does have an iSchool (e.g., an iSchool faculty member who graduated from Carnegie Mellon University's Computer Science program). This is different from university-to-university hiring networks in Clauset et al. (2015) and Wiggins et al. (2006, 2008), and offers a cleaner picture of relationships among iSchools. The weight of an edge is the number of PhDs produced by the target and hired by the source. We also built a similar hiring network that only reflects the hiring of senior faculty members for the purpose of comparison (Figure 6).

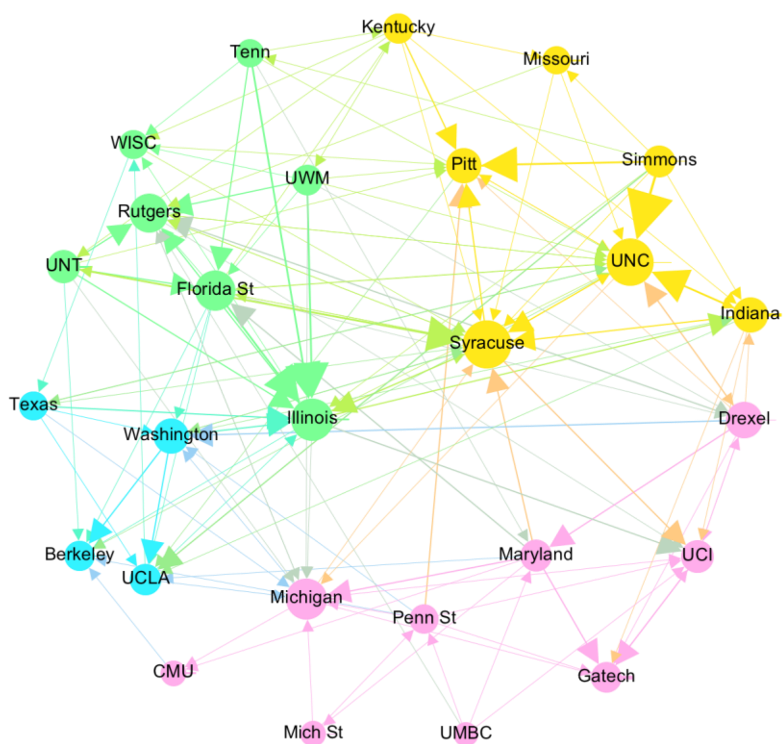

Figure 5. The iSchool hiring network.

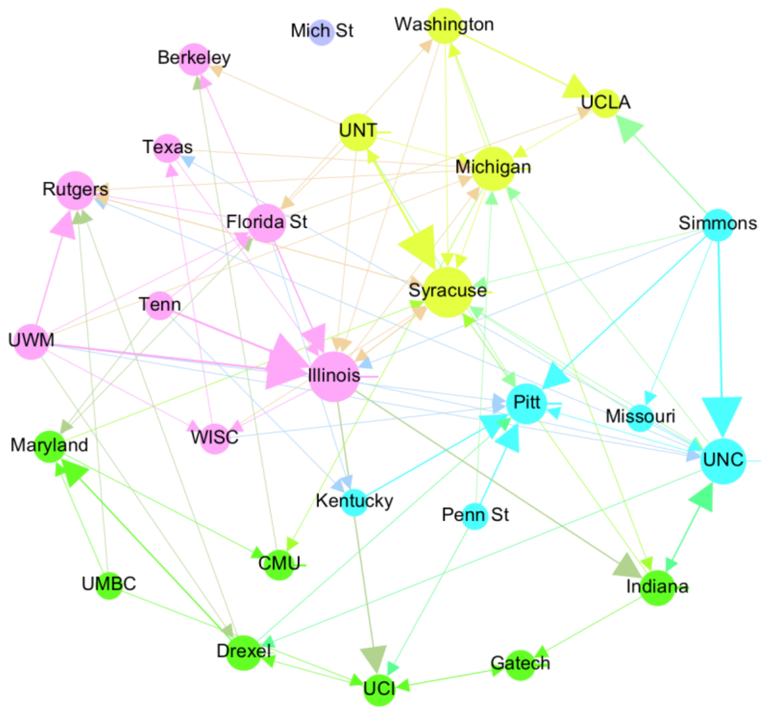

Figure 6. The iSchool hiring network with only senior faculty members.

Properties of the iSchool Hiring Network

Table 7 lists basic properties (calculated as undirected networks) of the two hiring networks in Figure 5 and 6. Overall, this is a very close-knit community. The network with all faculty members is able to connect all the 27 U.S. iSchools and no one is isolated from others. The average shortest path (the average length of the shortest paths between all possible pairs of nodes) is below 2 and the diameter (the maximal shortest path length between all possible pairs of nodes) is 4. These mean that each iSchool is on average fewer than 2 hops away from each other, and the maximum number of hops between any two iSchool is only 4. In addition to characteristics of the hiring network with both senior and junior faculty members, we are more interested in how the senior-only hiring network changed after the addition of junior faculty members. With the hiring of junior faculty members, the network gets more edges. As a result, the all-faculty network has a shorter average shortest path and a smaller diameter. This is natural as the two metrics are monotonically non-increasing when the number of edges increases and the number of nodes is fixed.

Network	Num. of Edges	Diameter	Avg. shortest path length	Modularity
Senior and junior faculty	155	3	1.70	0.48
Senior faculty only	99	4	1.93	0.64

Table 7. Basic properties of iSchool hiring networks.

To accurately capture whether the iSchool community is indeed getting closer, we also calculated the maximal modularity values for both networks. As a measure for the strength of community structures in networks, modularity is defined as the fraction of within-community edges minus such fraction if nodes were randomly connected (Newman and Girvan, 2004). It has been widely used to measure the quality of network clustering or community discovery. Community structures for the two hiring networks were also generated by maximizing the modularity of each network. It also serves as an indicator of whether a network features obvious communities that have many more intra-community edges than inter-community edges. The higher the maximal modularity is, the more fragmented a network is into sub-communities with more internal connections than external connections. Different from the average shortest path length and diameter, modularity is not directly correlated with the number of edges in a network and hence can better capture the overall connectedness of a network. It turns out with the addition of junior faculty members, the modularity of the hiring network decreases by 25%, from 0.64 to 0.48. This echoes our deduction that iSchools are getting closer to and more connected with each other as a community.

Assortativity of the iSchool Hiring Network

Whether it is the decreasing number of clusters based on faculty's educational backgrounds, or the increasing research topic similarity, our analysis above has shown that at the community level, iSchools are getting more homogeneous. We believe the hiring network can also help to explain the tendency toward higher homogeneity. Specifically, in addition to topological analysis in previous subsections, we also incorporated iSchools' educational background and research topic distributions into the hiring network as node attributes to examine its mixing pattern. We examined whether iSchools tend to hire faculty members from peer schools that are similar to or different from themselves. This tendency can be measured by assortativity. A network is assortative if nodes in this network tend to connect to others with similar characteristics, and dis-assortative otherwise (Newman, 2002; Zhao et al, 2010). It is sometimes referred to as “homophily” or “birds of a feather” in social science literatures (Mcpherson, Smith-Lovin, and Cook, 2001). In our case, the similarity between iSchools is based on their faculty’s educational backgrounds and research topics.

As node attributes, both educational backgrounds and research topics are in the form of vectors. Therefore, classic ways to calculate the assortativity coefficient of a network cannot be directly applied. Adopting the method proposed by Zhang and Pelechrinis (2014), we calculated the average cosine similarity between connected nodes based on the two vectors respectively, and compared it with the expected similarity when edges are placed at random. To obtain the latter, we conducted Monte Carlo simulations to sample 10,000 random graphs with the same number of nodes and edges as in the iSchool hiring network. A 95% confidence interval (CI) is calculated for the average similarity in sampled random graphs. The similarity will have values between zero and one, with zero meaning dis-assortativity and one indicating perfect assortativity. If the real-world hiring network has higher similarity score than the average similarity of random graphs, we conclude that the real-world network exhibits assortative mixing patterns.

We calculated assortativities of the senior-only and the junior-only hiring networks, and compared them with their counterparts in random graphs. Note that the calculation of assortativity for both hiring networks was based on senior faculty members’ distributions of educational backgrounds and research topics. It is also worth noting that the addition of a new faculty member in an iSchool could potentially have some effects on the research interests of current senior faculty members in the same school, because of potential collaboration after such addition. The collaboration between the new junior and senior faculty members usually lead to the publication of co-authored research papers. Including these co-authored papers in topic models inevitably makes the research topics of senior members and the junior faculty member closer to each other, which will lead to a junior-only network that is more assortative than it should be when junior faculty members were first hired. Thus, we extracted the topic distribution for an iSchool from its faculty members’ publications up to the year 2009 so that we can better approximate what kind of research senior faculty members were doing before junior faculty members joined their current iSchool. We chose the year 2009 with the assumption that it takes 6 years for a junior faculty member to get tenured and become a senior faculty member.

As shown in Table 8, assortativities of real-world hiring networks are significantly higher than those of the corresponding random graphs – the assortativity scores from actual networks are higher than the upper bounds of 95% confidence intervals from simulated networks. In other words, iSchools indeed prefer PhD graduates from peer iSchools whose faculty members share similar educational backgrounds and research topics. Meanwhile, lower assortativities in the junior-only hiring network compared to the senior-only network suggest that iSchools have been trying to be less assortative in hiring faculty members in recent years. Such endeavors in bringing in talents to complement or expand, instead of just reinforcing, their existing faculty's expertise may have contributed to the increasing educational and research homogeneity of the community as a whole.

Network	Education	Education (random)	Research Topic	Research Topic (random)
Senior only	0.6650	[0.4718, 0.4726]	0.7947	[0.7071, 0.7076]
Junior only (based on senior's profiles)	0.6027	[0.4720, 0.4730]	0.7655	[0.7074, 0.7080]

Table 8. Assortativity of the real-world hiring network compared to the 95% CIs of assortativity in random networks.

CONCLUSIONS AND FUTURE WORK

In this paper, we analyzed the current state, as well as the evolution of U.S. iSchools by examining the talents acquired and the research produced by iSchools. Our analysis covered three perspectives: faculty educational backgrounds, research topics, and hiring networks, by leveraging data from faculty profiles and publications, including conference papers. The findings suggest that iSchools are gradually finding their identity as a cohesive discipline and progressively straying away from other closely related disciplines such as computer science and library sciences. For instance, more and more new faculty members at iSchools are trained by peer iSchools. Research at iSchools is also gaining independence from other related disciplines and getting more homogenous as a group, with a focus on the intersection of information, technology and people. The community, as a network, is getting less fragmented and less assortative.

The major contributions of this research are in the following three areas: (1) To the best of our knowledge, this study represents the first to conduct longitudinal analysis on the evolution of iSchools. Different from analysis based on a snapshot of the community, a longitudinal approach can better capture directions to which iSchools are heading. Similar approaches could also be used to study other emerging and fast-changing disciplines. (2) From the standpoint of collecting and analyzing bibliographical data, we retrieved publication data from various types of venues (e.g., journals, conferences, and book chapters), and applied automated topic modeling techniques to reveal research topics for the whole community and each individual iSchool by examining text from these publications. Compared with relying only on journal papers and arbitrary classifications of journals into disciplines (Wu, He, Jiang, Dong, & Vo, 2012; Zhang, Yan & Hassman, 2013), our approach is more suitable for a multi- and inter-disciplinary community like the iSchools, whose researchers have diverse preferences regarding publication venues. Our computational analysis also enables us to analyze a large number of publications in an automated way. (3) Although there have been studies on academic disciplines using network analysis techniques (Clauset, Arbesman, & Larremore, 2015; Wiggins, Adamic & McQuaid, 2006; Wiggins, McQuaid & Adamic, 2008; Yu, 2013), we are the first to incorporate educational backgrounds and research topic distributions into the analysis of a discipline's hiring network. The inclusion of these nodal attributes into topological analysis of a hiring network helps to better capture characteristics and the evolution of a discipline.

The implications of our study are not limited to helping the iSchool community, funding agencies and employers better understand and guide the development of the new discipline. The approaches above to retrieve data, and to analyze the state and evolution of a discipline can also be easily adopted in the study of other disciplines, especially those facing rapid changes.

Admittedly, this study is not without limitations. First, we only examined U.S. iSchools. The inclusion of iSchools outside the U.S. could potentially give us better ideas of the evolution of the internationally emerging community. Second, in our longitudinal analysis of iSchools, we did not consider the mid-career movement of faculty members who moved to another iSchool or joined an iSchool from another discipline, or the exact year of one's transition from junior to senior members. The reason is that the exact time in which one was hired by an iSchool or got tenured can be difficult to obtain for many faculty members, whose full CVs are not available online. Instead, we assumed a typical career path of “assistant professor - associate professor - full professor” within the same institution. The promotion from assistant professors to associate professors is presumed to take six years. Although such assumptions are true for many faculty members, accurate data of each faculty member’s employment history would certainly help us calibrate the longitudinal analysis.

There are also interesting future research directions. To see if such topic evolution is specific to iSchools, it is intriguing to apply the methods above to similar departments such as computer science and library sciences. While information is usually considered a multi- and inter-disciplinary area, the evidence for interdisciplinarity is rare. To examine the existence of interdisciplinary nature, we can further explore topics from iSchool faculty publications to examine research that spans traditional disciplines. It is also interesting to explore whether having a diverse faculty body facilitates interdisciplinarity and whether interdisciplinary research leads to high scholar impact. The iSchool community, in this case, can serve as an ideal case study to address these two questions.

REFERENCES

Blei, D. M., Ng, A. Y., & Jordan, M. I. (2003). Latent Dirichlet allocation. Journal of Machine Learning Research, 3(4-5), 993–1022.

Chen, C. (2008). Thematic maps of 19 iSchools. Proceedings of the ASIST Annual Meeting, 45.

Clauset, A., Arbesman, S., & Larremore, D. B. (2015). Systematic inequality and hierarchy in faculty hiring networks. Science Advances, 1(1), e1400005–e1400005.

Cronin, B. (2005). An Identity crisis? The information schools movement. International Journal of Information Management, 25(4), 363–365.

Dillon, A. (2012). What it Means to be an iSchool. Journal of Education for Library & Information Science, 53(4), 267–273.

Frey, B. J., & Dueck, D. (2007). Clustering by Passing Messages Between Data Points. Science, 315, 972.

Holmberg, K., Tsou, A., & Sugimoto, C. (2013). The conceptual landscape of iSchools: Examining current research interests of faculty members. Information Research, 18(3), paper C32.

Luo, L. (2013). Being interdisciplinary: A look into the background and experiences of iSchool faculty members. Libres, 23(2), 1-20.

McPherson, M., Smith-Lovin, L., & Cook, J. M. (2001). Birds of a feather: homophily. in Social Networks." Annual Review of Sociology 27:415–44.

Newman, M. E. J. (2002). Assortative mixing in networks. Physical Review Letters, 89(20), 208701.

Newman, M. E. J., & Girvan, M. (2004). Finding and evaluating community structure in networks. Physical Review E - Statistical, Nonlinear, and Soft Matter Physics, 69(2 2), 026113–1–026113–15.

Olson, G. M., & Grudin, J. (2009). The information school phenomenon. Interactions, 16(2), 15–19.

Rousseeuw, P. J. (1987). Silhouettes: A graphical aid to the interpretation and validation of cluster analysis. Journal of Computational and Applied Mathematics, 20, 53–65.

Seadle, M., Greifeneder, E. (2007). Envisioning an iSchool Curriculum. Information Research, 12(4), paper colise02.

Shannon, C. E. (1948). A mathematical theory of communication. Bell System Technical Journal, The, 27(3), 379–423. http://doi.org/10.1002/j.1538-7305.1948.tb01338.x

Subramaniam, M. M., & Jaeger, P. T. (2010). Weaving diversity into LIS: An examination of diversity course offerings in iSchool programs. Education for Information, 28(1), 1–19.

Sugimoto, C. R., Li, D., Russell, T. G., Finlay, S. C., & Ding, Y. (2011). The shifting sands of disciplinary development: Analyzing North American Library and Information Science dissertations using latent Dirichlet allocation. Journal of the American Society for Information Science and Technology, 62(1), 185–204.

Wiggins, A., Adamic, L., & McQuaid, M. (2006). The Small Worlds of Academic Hiring Networks. In Applications of social network analysis: Proceedings of the 3rd conference on applications of social network analysis.

Wiggins, A., McQuaid, M. J., & Adamic, L. A. (2008). Community Identity: Peer Prestige and Academic Hiring in the iSchools.

Wiggins, A., & Sawyer, S. (2012). Intellectual diversity and the faculty composition of iSchools. Journal of the American Society for Information Science and Technology, 63(1), 8–21.

Wu, D., He, D., Jiang, J., Dong, W., & Vo, K. T. (2012). The state of iSchools: An analysis of academic research and graduate education. Journal of Information Science, 38(1), 15–36.

Yu, S.-Y. (2013). Detecting collaboration patterns among iSchools by linking scholarly communication to social networking at the macro and micro levels. Libres, 23(2), 1-13.

Zhang, K., & Pelechrinis, K. (2014). Understanding Spatial Homophily: The Case of Peer Influence and Social Selection. In Proceedings of the 23rd International Conference on World Wide Web (pp. 271–282). Geneva, Switzerland. http://doi.org/10.1145/2566486.2567990.

Zhang, P., Yan, J. L. S., & Hassman, K. D. (2013). The intellectual characteristics of the information field: Heritage and substance. Journal of the American Society for Information Science and Technology, 64(12), 2468–2491.

Zhao, K., Ngamassi, L.-M., Yen, J., Maitland, C., & Tapia, A. (2010). Assortativity Patterns in Multi-dimensional Inter-organizational Networks: A Case Study of the Humanitarian Relief Sector. Proceedings of the Advances in Social Computing--Proceedings of the 2010 International Conference on Social Computing, Behavioral Modeling, and Prediction (SBP'10), Springer. 265-272.