Examining the academic mobility at Historically Black Colleges and Universities in the U.S.

Erjia Yan[1], Mat Kelly[2], Deanna Zarrillo[3], Jiangen He[4], Chaoqun Ni[5], Robert Palmer[6]

[1] *ey86@drexel.edu,* [2]*mkelly@drexel.edu,* [3]*dz364@drexel.edu*
College of Computing and Informatics, Drexel University, Philadelphia (USA)

[4] *jiangen@utk.edu*
School of Information Sciences, University of Tennessee, Knoxville (USA)

[5] *chaoqun.ni@wisc.edu*
Information School, University of Wisconsin, Madison (USA)

[6] *robert.palmer@howard.edu*
Department of Educational Leadership and Policy Studies, Howard University, Washington D.C. (USA)

Abstract
This RiP paper describes the current state and planned research for a project that examines academic mobility in Historically Black Colleges and Universities (HBCUs). The project has designed protocols for collecting faculty affiliation data using Internet Archive. The project will use the data to identify faculty who changed affiliations (mobile) and those who did not (non-mobile) joining or leaving HBCUs, query professors' publication, citation in a local Web of Science database, and gain in-depth understanding of HBCU faculty retention and mobility through surveys and focused interviews. The integrated data will allow for clear and robust examination of academic mobility, institutional stratification, and the role of organizational factors in shaping academic mobility.

Introduction
Before the passage of the Civil Rights Act in 1964, HBCUs were the only institutions in the U.S. that were created for the purpose of educating Black citizens (Gasman, 2013). While some research has shown that some high-profile Black scholars and coaches are choosing to work at HBCUs (Broady et al., 2021), research has generally shown that HBCUs are on the losing side of attracting the most talented Black students and faculty (Seymore, 2005) and suffering a concomitant "brain drain." For the past few decades social mobility and civil rights movements have accelerated academic mobility (Sugimoto et al., 2017; Van Noorden, 2012); however, there is scarce research on academic mobility for HBCUs. In this research-in-progress paper, we present the current state and planned research of our work on identifying academic mobility of Historically Black Colleges and Universities (HBCUs).

Mobility is driven by a variety of reasons that can be academic, job-related, family-related, or personal (Auriol, 2010). Mobile researchers who changed affiliations during their scientific career tend to have slightly higher publication and citation rates than other researchers (Aksnes, Rørstad, Piro, & Sivertsen, 2013). In HBCUs, brain drain has been a subject of constant debate and worry because many Black professors and students transition from HBCUs to non-HBCUs, but few non-Black professors and students transition to Black colleges (Allen, 1991). Brain drain in this context began as desegregation allowing Black students and faculty to enter non-HBCUs where academic quality was perceived as higher (Morris, 1972). and predominantly White Institutions were able to provide better resources to attract Black faculty due to the chronic underfunding of Black colleges (Broady et al., 2021). Academic Mobility of Black HBCU faculty has been exacerbated by the fact that elite universities in the U.S. have been keen to hire highly qualified Black faculty members to increase faculty diversity (Barrett & Smith, 2008). A 2010 U.S. Commission on Civil Rights report affirms that as resistance to

Black attendance at non-HBCUs has faded and the need for segregated schooling has declined, it is reasonable to expect a brain drain from HBCUs (Rights, 2010). However, empirical studies of this problem have rarely been conducted since the 1990s.

To fill this gap, this project intends to take advantage of modern information technologies including the Internet Archive's Wayback Machine, Carnegie Classification of Institutions of Higher Education (CCIHE), and Web of Science to collect a large, heterogenous, longitudinal data set. This project includes 10 major steps. These steps are (Figure 1): (1) select the target HBCUs for this project using CCIHE; (2) identify web links for each college or department at the selected HBCUs in Internet Archive and check the Archive's coverage and reliability; (3) build a local publication and citation database using XML Web of Science data; (4) collect professors' affiliation data using Internet Archive for each year between 2006 and 2020 and supplement the data using LinkedIn, ORCID, personal websites, and CVs; (5) use the yearly affiliation data to group professors into four categories based on their mobility and types of institutions; (6) calculate pre- and post-move publication and citation trends and identify changes in human capital at HBCUs; (7) model the effect of academic moves on professors' research activities in relation to several variables using logistic regressions; (8) gain an in-depth understanding of the effect of mobility and factors associated with move decisions through surveys and interviews; (9) design a visual dashboard to publish and communicate research data and outputs; and (10) broadly share and communicate project results.

Figure 1. Flow chart showing key design steps

Current state of the project

HBCU selection

There are currently 101 HBCUs according to the National Center for Education Statistics. Based on CCIHE, of these 101 HBCUs, 11 are doctoral-level institutions, 24 are master's-level institutions, and 66 are associate's-level or four-year institutions. Because the goal of this project is to measure the impact of academic moves on research activities, an institution must have at least moderate research intensity to be included. For this reason, we selected the

35 master's or doctoral degree granting HBCUs with potentially higher levels of research intensity.

Using Memento to collect university home URIs[1]

We began the data collection process by manually gathering the current URI of each HBCU homepage and its respective department-level pages into a spreadsheet. This data serves as input into a Python script that utilizes the MemGator (Alam & Nelson, 2016) aggregator to create and retain a Memento (Van de Sompel et al., 2013) TimeMap for each URI that passes through the program. The semantic and temporal context of these URI associations is represented in the TimeMap as relative relation attributes. For each HBCU homepage, we were able to collect a complete time range of historical URIs, which supported the later web scraping phase of data collection. Across 35 homepages, we collected a total of 182,980 mementos with a mean of 5,228 mementos per original URI and a median of 5,322. However, while collecting and identifying the temporal extent of coverage, we found some department-level TimeMaps did not contain a complete set of captures from our time range. Accordingly, to ensure data quality, we used the TimeMaps generated for each HBCU homepage to manually collect these archival department-level URIs (URI-Ms).

Collecting faculty affiliation data

Once university URIs were obtained and stored, we used the following procedures to collect professor data. As of December 2022, the team has collected about 50% of the professor affiliation data (from 2021 to 2012 backwards) for the 11 doctoral-level HBCUs.

1. Go to the university URI data sheet, choose a capture made in the fall term (between September and December) of 2021.
2. Locate the web page that contains the list of schools and colleges, typically called "Academics" or "Schools and Colleges" links.
3. From the list of colleges or schools, select each college or school iteratively and document the faculty information until professors' information from all colleges and schools of that university have been collected.
4. From the school/college page, locate the faculty page or directory from the college level links. If the faculty page is embedded within each department of that college, then go to each department and collect the faculty data.
5. Input the faculty data in the format shown in Figure 2.
6. Once the numbers are tallied and input in the table, move to the university's homepage's capture in the previous year and repeat steps 1-6. Finish collecting this university until the year 2005's data is collected and documented.

[1] A detailed technical description of the process can be found at: Zarrillo, D., Kelly, M., Jackson, C., Yan, E. (2022). Collecting Diachronic Affiliation Data for Faculty at HBCUs Using Memento. 85th Annual Meeting of the Association for Information Science & Technology, Oct. 29-Nov. 1, 2022, Pittsburgh, PA.

Home capture URI	faculty page URI	School or college name	Department name	Faculty name	Faculty rank	Faculty email	Note
https://web.archive.org/web/20211103213315/https://howard.edu/	https://web.archive.org/web/20211118012047/https://history.howard.edu/faculty	College of Arts and Sciences	Department of History	Ana Lucia Araujo	Professor	aaraujo@howard.edu	
https://web.archive.org/web/20211103213315/https://howard.edu/	https://web.archive.org/web/20211118012047/https://history.howard.edu/faculty	College of Arts and Sciences	Department of History	Elizabeth Clark-Lewis	Professor	eclark-lewis@howard.edu	
https://web.archive.org/web/20211103213315/https://howard.edu/	https://web.archive.org/web/20211118012047/https://history.howard.edu/faculty	College of Arts and Sciences	Department of History	Jeffrey Kerr-Ritchie	Professor	jkerr-ritchie@howard.edu	

Figure 2. Sample data collection worksheet

While Internet Archive's coverage is generally consistent for the selected universities based on our sampled URLs in the pilot analysis, there may be cases where URL captures are unavailable. We have designed the following strategies to address this issue. First, considering that academic moves are a low-volatility event, if the gap between two captures is more than one year but less than two years, we assume no change of affiliation was made between the two captures. For instance, if one capture is in October 2009 and the next is in September 2011, we assume the departmental affiliation in Fall 2010 is the same as the affiliation in October 2009. Second, if the gap between two captures is more than two years, we will check the affiliation of each professor on the webpage through secondary sources including LinkedIn, ORCID, personal websites, and CVs and use the collected information to fill the gaps. The combination of these three methods should fill most of the gaps left by Internet Archive. However, after applying these methods, we will flag the record of any professor whose employment history has remaining gaps of more than two years and exclude these records from the analysis.

Extracting publication and citation data

The XML-format of Web of Science data has been processed and stored in the form of a MySQL database. We queried each HBCU in the database: the average number of publications for doctoral-level institutions since 1980 is 5,150; and 1,250 is the corresponding figure for master's-level institutions. Considering the relatively small faculty size of HBCUs, these numbers should provide enough data points for analysis.

Planned research

Grouping professors by mobility and institution type

After collecting professor affiliation data for the 35 HBCUs, we will compare the yearly affiliation change at the institution level and identify incoming and outgoing professors. For incoming professors, we need to know previous affiliations; for outgoing professors, we need to know affiliations post-move. We will use Web of Science to collect such data. To identify incoming professors' previous affiliation and outgoing professors' next affiliation, we will search professors' names in "Author Search" in Web of Science. The search results are disambiguated researcher profiles with their affiliations and publication venues. The results also show past and current organization affiliations, including duration of affiliation, based on publication data. Affiliation information will be recorded to identify types of academic moves.

Despite Web of Science's efforts in disambiguating authors names, mismatches do occur, and we have designed rules to minimize errors: namely, using (1) research areas, (2) publication venues, and (3) publication history to disambiguate authors so that the target author's publications should fall within the same research areas, the same set of venues, and the same time span. For outgoing professors, if no new affiliation can be identified post-move, we will record the move as a dropout.

After the affiliation and duration data is collected and disambiguated, we will cross-reference institutions with CCIHE to obtain institution-level data, including HBCU status. The integrated data will be used to identify four categories of professors: (1) non-mobile professors (staying within one HBCU since 2006); (2) mobile professors within HBCUs; (3) mobile professors from non-HBCU to HBCU; and (4) mobile professors from HBCU to non-HBCU. Each data point is at the per-move level, which means a professor with an employment history of "HBCU to non-HBCU to HBCU" will be counted as two data points: HBCU to non-HBCU and non-HBCU to HBCU.

Historical comparisons

Historical comparison is an effective approach to examining the longitudinal differences of productivity and impact of different mobility groups. To understand true differences in the productivity and impact of professors in the four categories of mobility, we need to recreate the entire longitudinal trajectory. This project makes this possible as it collects complete publication and mobility data for professors in the selected HBCUs. To obtain professors' yearly publication and citation data, we will need to run batch queries in our local XML Web of Science database. We will use a combination of author names, institution names, and affiliation duration from Section 3.2 to query (e.g., "last_name, first_name_initial"+Univ X+Year:200X–201X). If we identify noticeable noise in the query results, we will add to the query combination a discipline qualifier to flag any publications that are not in the author's home domain, and we will manually screen those flagged publications for enhanced accuracy.

We will employ an approach of a previous work that used historical comparison and collect six types of characteristics for each professor: total productivity, career length, annual productivity, total impact, academic age, and dropout (Huang, Gates, Sinatra, & Barabási, 2020). To properly compare the longitudinal differences of productivity and impact of professors in different mobility groups, we will use coarsened exact matching (CEM) (Iacus, King, & Porro, 2012) to balance the covariations of the abovementioned six factors for the four mobility groups.

Because of financial constraints, most HBCU faculty have tended to have a higher teaching load than those at non-HBCUs (Gasman, 2013). Heavy teaching loads may consume much research time for faculty and hinder research productivity. Our pilot study showed that between R2 (Doctoral Universities – High research activity) and master-level institutions, R2 institutions on average published more papers indexed in Web of Science (Currently, no HBCU holds R1 status). If professor-level publication data is insufficient for meaningful analysis, we will select only professors from the 11 R2 institutions or professors with sufficient publications or citations for individual-level analysis.

Contextualizing findings through surveys and interviews

To understand factors that help attract and retain faculty at HBCUs and factors that made faculty leave HBCUs, we need to resort to additional research instruments. The project will design and conduct one set of surveys and interviews. The two tasks, integral to this project, will be closely

connected with the previous steps. When we collect mobility data from the Internet Archive, we will collect professors' email addresses that can be used to distribute questionnaires. The survey will ask if respondents have changed their primary affiliations in the past 20 years. Based on the answer to this question, two branches of questions will be prepared: one for mobile professors and the other for non-mobile professors. For mobile professors, the survey will ask reasons to move and levels of institutional support pre- and post-move, including teaching load and compensation. For non-mobile professors, the survey will ask reasons to stay at current HBCUs. We will also collect demographic information including race and ethnicity. The collected information will be used as factors in modeling the effect of academic moves on professors' productivity and impact. Possible reasons and types of support will be listed, and a 5-point Likert scale will accompany the items on the list. At the end of the questionnaire, we will query respondents' interest in participating in a follow-up interview. From those interested, we will sample 24 interviewees, 12 at master's-level institutions (6 pre-tenure and 6 tenured) and 12 at doctoral-level institutions (6 pre-tenure and 6 tenured). These in-depth interviews (45–60 mins) will be semi-structured. Each interview will inquire about (1) the role of academic moves in the interviewee's research activities, (2) differences in institutional support of research and scholarly activities, and (3) observations of institutional human capital changes at HBCUs and non-HBCUs.

Conclusion

This project will examine the effect of academic mobility on the productivity, impact, and career paths of professors employed at HBCUs. The emphasis on mobility of HBCU professors is distinct from previous research that largely focused on mobility at majority universities. Results from this large-scale, longitudinal analysis will provide important evidence regarding the career paths of professors at HBCUs, including those moving to HBCUs and those moving away from HBCUs. This project will conduct surveys and interviews to identify and understand factors associated with mobility and retention decisions for HBCU faculty and deliver meaningful interpretations. This project will provide data-backed evidence to support a diverse, inclusive, and equitable scientific workforce.

To ensure the sustainability of this project, we will make our data materials, statistical codes, tools, and research outputs available on our project website, as well as in open access platforms (Figshare and GitHub). The results of this project will provide insights for HBCU administrators. Outputs of this project should help facilitate a transparent dialogue around the challenges faced by minority-serving institutions and underrepresented faculty.

Acknowledgments

This paper is supported by National Science Foundation awards no. 2122525, 2121861, and 2122691.

References

Aksnes, D. W., Rørstad, K., Piro, F. N., & Sivertsen, G. (2013). Are mobile researchers more productive and cited than non-mobile researchers? A large-scale study of Norwegian scientists. *Research Evaluation, 22*, 215-223. doi:10.1093/reseval/rvt012

Alam, S., & Nelson, M. L. (2016, June). MemGator—A portable concurrent memento aggregator: Cross-platform CLI and server binaries in Go. In 2016 IEEE/ACM Joint Conference on Digital Libraries (JCDL) (pp. 243-244). IEEE.

Allen, H. L. (1991). The Mobility of Black Collegiate Faculty Revisited: Whatever Happened to the" Brain Drain"? *The Journal of Negro Education, 60*(1), 97-109.

Auriol, L. (2010). Careers of Doctorate Holders. Employment and Mobility Patterns. *OECD Science, Technology and Industry Working Papers*, 29. doi:10.1787/5kmh8phxvvf5-en

Barrett, T. G., & Smith, T. (2008). Southern coup: Recruiting African American faculty members at an elite private Southern research university. *American Educational Research Journal, 45*(4), 946-973.

Broady, K., Perry, A. M., & Romer, C. (2021). Underfunding HBCUs leads to underrepresentation of Black faculty. Brookings.

Gasman, M. (2013). The changing face of historically Black colleges and universities.

Huang, J., Gates, A. J., Sinatra, R., & Barabási, A.-L. (2020). Historical comparison of gender inequality in scientific careers across countries and disciplines. *Proceedings of the National Academy of Sciences, 117*(9), 4609-4616.

Iacus, S. M., King, G., & Porro, G. (2012). Causal inference without balance checking: Coarsened exact matching. *Political analysis*, 1-24.

Morris, E. W. (1972). The contemporary Negro college and the brain drain. *The Journal of Negro Education, 41*(4), 309-319.

Palmer, R. T., & Griffin, K. (2009). Desegregation policy and disparities in faculty salary and workload: Maryland's historically Black and predominantly White institutions. Negro Educational Review, 60(1-4), 7-21.

Rights, U. C. o. C. (2010). The educational effectiveness of historically Black colleges and universities. In: Author Washington, DC.

Seymore, S. B. (2005). I'm confused: How can the federal government promote diversity in higher education yet continue to strengthen historically Black colleges. *Wash. & Lee J. Civil Rts. & Soc. Just., 12*, 287.

Sugimoto, C. R., Robinson-Garcia, N., Murray, D. S., Yegros-Yegros, A., Costas, R., & Larivière, V. (2017). Scientists have most impact when they're free to move. *Nature, 550*, 29-31. doi:10.1038/550029a

Van de Sompel, H., Nelson, M., Sanderson, R.: HTTP Framework for Time-Based Access to Resource States – Memento. IETF RFC 7089 (December 2013)

Van Noorden, R. (2012). Global mobility: Science on the move. *Nature News, 490*, 326.