First steps in
Identifying Academic Migration using Memento and Quasi-Canonicalization

Mat Kelly, Deanna Zarrillo, Christopher Jackson, and Erjia Yan

Drexel University
College of Computing & Informatics (CCI)
Philadelphia, PA

mkelly@drexel.edu
@machawk1

Web Archiving and Digital Libraries Workshop 2022
at the ACM/IEEE Joint Conference on Digital Libraries 2022
June 24, 2022

JCDL
COLOGNE
2022

The Problem: HBCU faculty more prone to mobility

- Historically Black Colleges and Universities (HBCUs) are the only institutions in the United States that were created for the <u>express purpose of educating Black citizens</u>. *(Gasman, 2013)*
- Brown v. Board of Education, HBCUs competing with non-HBCUs to attract and retain outstanding faculty, often losing the battle (*Seymore, 2005*)
- **Brain drain**: substantial exit of faculty who hold any expertise, considered valuable institutional assets *(Rosenblatt & Sheaffer, 2001)*
- Social mobility and civil rights movements in the past few decades have accelerated academic mobility *(Sugimoto, 2017)*

**Thus, the need to investigate academic mobility
of faculty at HBCUs is <u>ever-increasing</u>.**

DREXEL UNIVERSITY
College of
Computing & Informatics

First steps in Identifying Academic Migration using Memento and Quasi-Canonicalization
Web Archiving and Digital Libraries (WADL) Workshop 2022

@machawk1   2

The *Ultimate* Goal: Identify Brain Drain

Project Status

This project runs from 2021 to 2024.

1. **Data collection, survey preparation, website creation**
2. Data cleaning, supplementation, and classification
3. Publication identification, survey identification, data publication
4. Professor productivity identification, pilot/distribute questionnaire, disseminate outputs
5. Professor modeling, follow-up interviews, visual dashboard creation
6. Regression testing, visualization dashboard dissemination, data dissemination, conclude interviews, present results

- Examine two forms of effects of academic mobility
  - Institution Level: measured in terms of the institutional human capital change at HBCUs
    - brain drains or gains
  - Individual Level: where the impact of academic moves on professors' research activities is assessed

- Collect faculty affiliation data from **35 HBCUs** with master's- or doctoral-level programs
- Conduct surveys to reveal
  - Factors that attract and retain faculty at HBCUs
  - Factors that made faculty leave HBCUs
- Follow up with two dozen in-depth interviews with HBCU faculty.

DREXEL UNIVERSITY College of Computing & Informatics

First steps in Identifying Academic Migration using Memento and Quasi-Canonicalization
Web Archiving and Digital Libraries (WADL) Workshop 2022

@machawk1   3

Project Methodology

1. Collect faculty affiliation data from Internet Archive*
2. Identify mobile and non-mobile professors who joined and left HBCUs
3. Query professors' publications and citations in a local XML version of the Web of Science database
4. Collect institution-level profiles from CCIHE (Carnegie Classification)
5. Gather survey data
6. Transcribe interview data

Building off the integrated data, we will employ historical comparisons, bibliometrics, and regression modeling for a clear and robust examination of academic mobility, institutional stratification, and the role of organizational factors in shaping academic mobility.

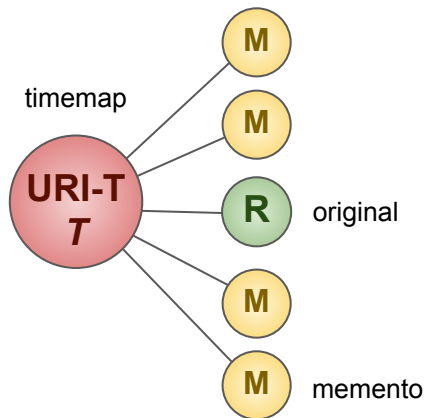**Results will help to better understanding HBCU faculty retention and mobility.**

* originally proposed, using multiple archive via aggregation instead

DREXEL UNIVERSITY
College of
Computing & Informatics

First steps in Identifying Academic Migration using Memento and Quasi-Canonicalization
Web Archiving and Digital Libraries (WADL) Workshop 2022

@machawk1

4

Our Approach, Step 0: Get TimeMaps

- Start with 35 HBCU URI-Rs →
- Fetching TimeMaps
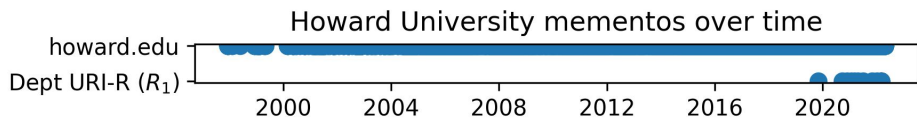

- Default set of archives queried
- Run from localhost

Alam & Nelson, "MemGator - A Portable Concurrent Memento Aggregator: Cross-Platform CLI and Server Binaries in Go", Presented at **WADL 2016**.

HBCU homepage URI-Rs

Howard University	https://howard.edu/
Florida Agricultural and Mechanical University	https://www.famu.edu/
North Carolina A & T State University	https://www.ncat.edu/
Jackson State University	https://www.jsums.edu/
Tennessee State University	https://www.tnstate.edu/
Hampton University	https://www.hamptonu.edu/
Morgan State University	https://www.morgan.edu/
Tuskegee University	https://www.tuskegee.edu/
Clark Atlanta University	https://www.cau.edu/
North Carolina Central University	https://www.nccu.edu/
Alabama A & M University	https://www.aamu.edu/
Prairie View A & M University	https://www.pvamu.edu/
Xavier University of Louisiana	https://www.xula.edu/
Southern University and A & M College	https://www.subr.edu/
Delaware State University	https://www.desu.edu/
Norfolk State University	https://www.nsu.edu/
Texas Southern University	http://www.tsu.edu/
University of Maryland Eastern Shore	https://wwwcp.umes.edu/
Virginia State University	https://www.vsu.edu/
Winston-Salem State University	https://www.wssu.edu/
University of the District of Columbia	https://www.udc.edu/
Fayetteville State University	https://www.uncfsu.edu/
Alabama State University	https://www.alasu.edu/
Langston University	https://www.langston.edu/
Fort Valley State University	https://www.fvsu.edu/
Savannah State University	https://www.savannahstate.edu/
South Carolina State University	https://scsu.edu/
Bowie State University	https://www.bowiestate.edu/
Alcorn State University	https://www.alcorn.edu/
Albany State University	https://www.asurams.edu/
Grambling State University	https://www.gram.edu/
Lincoln University	https://www.lincoln.edu/
Coppin State University	https://www.coppin.edu/
Southern University at New Orleans	https://www.suno.edu/
Mississippi Valley State University	https://www.mvsu.edu/

First steps in Identifying Academic Migration using Memento and Quasi-Canonicaliz
Web Archiving and Digital Libraries (WADL) Workshop 2022

DREXEL UNIVERSITY
College of
Computing & Informatics
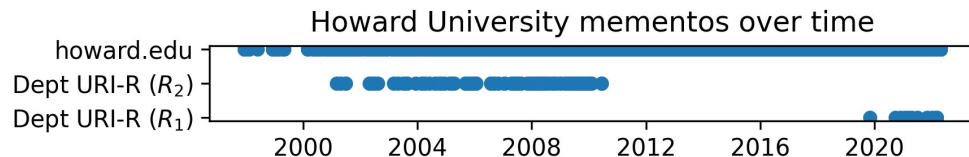
Our Approach, Step 1: Identify Dept URI-Rs

- ~~Start with 35 HBCU URI-Rs →~~
- ~~Fetching TimeMaps~~
- Manually collect URI-Rs of depts on live web
- Get TimeMaps for departments

Howard University mementos over time

howard.edu
Dept URI-R (R_1)

2000 2004 2008 2012 2016 2020

Alabama State University	
College of Liberal Arts & Social Sciences	https://www.alasu.edu/clas/college-liberal-arts-social-sciences
Psychology	https://www.alasu.edu/clas/psychology/department-psychology
Criminal Justice & Social Sciences	https://www.alasu.edu/clas/sociology-criminal-justice/department-criminal-justice-and-social-sciences
Communications	https://www.alasu.edu/clas/communications/department-communications
History & Political Science	https://www.alasu.edu/clas/history-political-science/department-history-and-political-science
Languages & Literatures	https://www.alasu.edu/clas/languages-literatures/department-languages-literatures
Social Work	https://www.alasu.edu/clas/social-work/department-social-work
College of Business Administration	https://www.alasu.edu/college-business-administration
Accounting & Finance	https://www.alasu.edu/cba/accounting-finance/department-accounting-finance
Business Administration	https://www.alasu.edu/cba/business-administration/business-administration
Computer Information Systems	https://www.alasu.edu/cba/computer-information-systems/department-computer-information-systems
College of Education	https://www.alasu.edu/coe/college-education
COE Accreditation	https://www.alasu.edu/coe/coe-accreditation/college-education-accreditation
Instructional Support	https://www.alasu.edu/coe/instructional-support/instructional-support-programs
Regional 8 Inservice Center	https://www.alasu.edu/coe/regional-education-inservice-center
Health, Physical Education & Recreation	https://www.alasu.edu/coe/health-physical-education-recreation/department-health-physical-education-recreation
Teacher Certification	redirect to 2021
COE Programs	https://www.alasu.edu/coe/coe-programs/college-education-programs
Biology and Secondary Education	https://www.alasu.edu/coe/curriculum-instruction-program-available-online/med-biology-secondary-education
COE Grants	https://www.alasu.edu/coe/coe-grants/college-education-grants
Elementary Education	https://www.alasu.edu/coe/curriculum-instruction-program-available-online/elementary-education
Early Childhood Center	https://www.alasu.edu/coe/zsecc
Curriculum & Instruction (Program available onl	https://www.alasu.edu/coe/curriculum-instruction-program-available-online/department-curriculum-instruction
History and Secondary Education	https://www.alasu.edu/coe/curriculum-instruction-program-available-online/med-history-education
College of Health Sciences	https://www.alasu.edu/chs/college-health-sciences
Physical Therapy	https://www.alasu.edu/chs/physical-therapy/physical-therapy
Prosthetics and Orthotics	redirect to 2021
Maternal and Child Health Pipeline Training Pro	https://www.alasu.edu/chs/maternal-and-child-health-pipeline-training-program/maternal-and-child-health-pipeline
Biomechanics and Motor Control Lab	https://www.alasu.edu/chs/biomechanics-and-motor-control-lab/biomechanics-and-motor-control-laboratory
Health Information Management	https://www.alasu.edu/chs/health-information-management/health-information-management
Rehabilitation Studies	https://www.alasu.edu/chs/rehabilitation-studies/department-rehabilitation-studies
Occupational Therapy	https://www.alasu.edu/chs/occupational-therapy/occupational-therapy
College of Science, Technology Engineering & M	https://www.alasu.edu/college-science-technology-engineering-mathematics
Biological Sciences	https://www.alasu.edu/cstem/biological-sciences/department-biological-sciences
Physical Sciences	https://www.alasu.edu/cstem/physical-sciences/physical-sciences-department
Math & Computer Science	https://www.alasu.edu/cstem/math-computer-science/department-mathematics-computer-science
Biomedical Engineering	https://www.alasu.edu/cstem/biomedical-engineering/welcome-biomedical-engineering
College of Visual & Performing Arts	https://web.archive.org/web/20200505082005/https://www.alasu.edu/cvpa/college-visual-performing-arts
Music	https://www.alasu.edu/cvpa/music/department-music
Theatre Arts	https://www.alasu.edu/cvpa/theatre-arts/department-theatre-arts

DREXEL UNIVERSITY
College of
Computing & Informatics

First steps in Identifying Academic Migration using Memento and Quasi-Canonicalization
Web Archiving and Digital Libraries (WADL) Workshop 2022
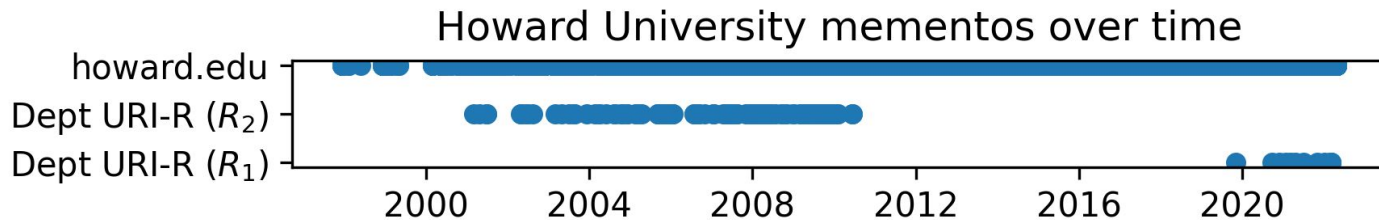
@machawk1

6

Our Approach, Step 2: Discover Old Depts' URI-Rs

- Start with 35 HBCU URI-Rs →
- Fetching TimeMaps
- Manually collect URI-Rs of depts on live web
- Get TimeMaps for departments
- Consult older URI-Ms of homepages
  - **DISCOVER** historical locations (URI-Rs) of depts
  - Ultimately use **this** as the basis for finding faculty's name and associations

Howard University mementos over time

First steps in Identifying Academic Migration using Memento and Quasi-Canonicalization
Web Archiving and Digital Libraries (WADL) Workshop 2022

@machawk1

7

DREXEL UNIVERSITY
College of
Computing & Informatics

Quasi-Canonicalization

- Can we associate the prior URI-Rs of a dept together to get an accurate picture of the dept over time?
- What about departmental splitting, merging, creation, and disbandment?
- Typically, <u>URI variations</u> (example.com:80, www.example.com) are basis of association
  - Here: **different URI-Rs** represent the same dept. over time

Howard University mementos over time

DREXEL UNIVERSITY
College of
Computing & Informatics

First steps in Identifying Academic Migration using Memento and Quasi-Canonicalization
Web Archiving and Digital Libraries (WADL) Workshop 2022

@machawk1

Canonicalization Caveats

- HTTP/HTTPS transition
- Archived HTTP 3XXs
- Other patterns previously identified*
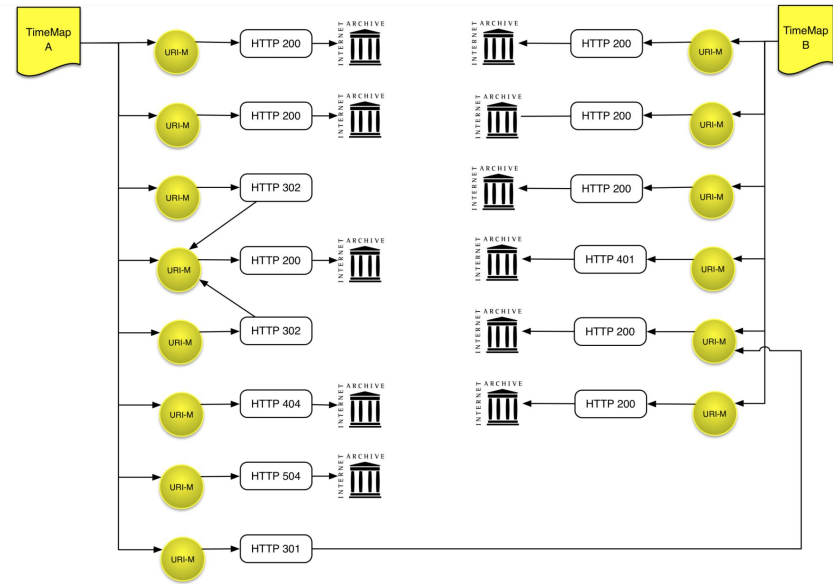
- **This investigation**: suitability of multiple URI-Rs association ("quasi-canonicalization") upon discovery of departmental movement.

* Mat Kelly, Lulwah M. Alkwai, Sawood Alam, Michael L. Nelson, Michele C. Weigle, and Herbert Van de Sompel, "Impact of URI Canonicalization on Memento Count," Presented at the Web Archiving and Digital Libraries Workshop **(WADL 2017)**. Toronto, Canada, June 2017. arXiv:1703.03302

DREXEL UNIVERSITY
College of
Computing & Informatics

First steps in Identifying Academic Migration using Memento and Quasi-Canonicalization
Web Archiving and Digital Libraries (WADL) Workshop 2022
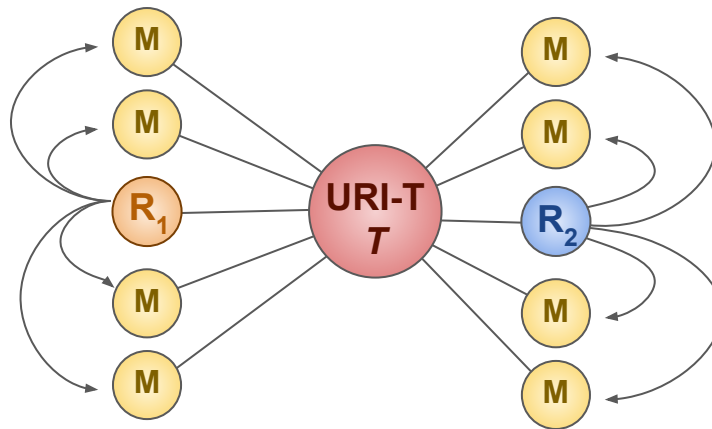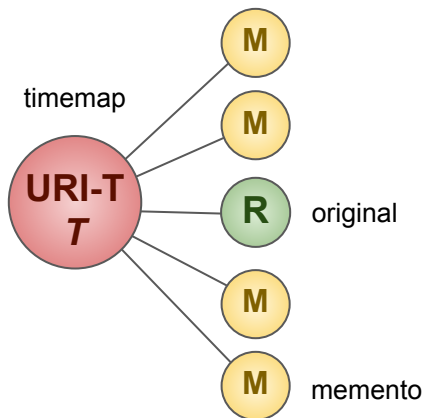
@machawk1   9

Premise & Hypotheses

- Identifying past departmental URI-R can act as basis for identifying faculty
  - (TBD)
- Anticipated faculty movement
  - Inter-university
  - Inter-department (at same university)
  - Leave academia
  - Other patterns to-be-explored

First steps in Identifying Academic Migration using Memento and Quasi-Canonicalization
Web Archiving and Digital Libraries (WADL) Workshop 2022

@machawk1   10

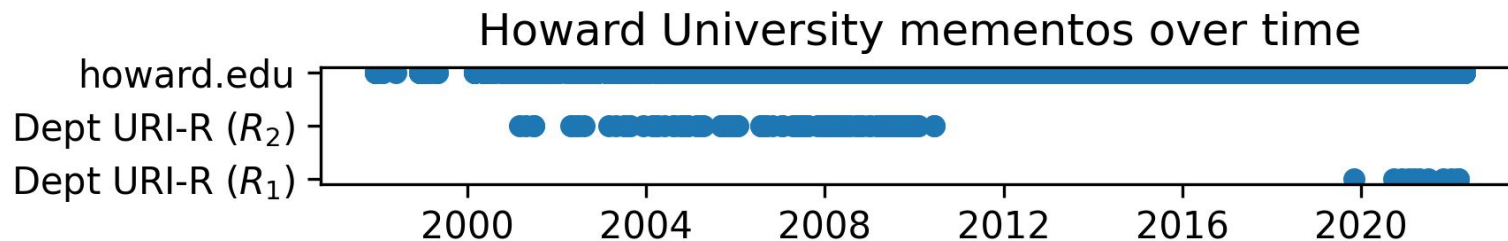DREXEL UNIVERSITY
College of
Computing & Informatics

Memento-Compatible TimeMaps?

- Per-department TimeMap rather than (in addition to) per URI-R
- Maybe moot, as end-goal is to not serve these mementos

First steps in Identifying Academic Migration using Memento and Quasi-Canonicalization
Web Archiving and Digital Libraries (WADL) Workshop 2022

@machawk1   11

Memento-Compatible Navigation?

- Resolving URI-R to a past datetime beyond what is available
  - Perhaps URI-R_1 resided at URI-R_2
- The key might be in the "key"
  - URI-R as primary parameter could produce TimeMap containing URI-Ms that do not conventionally canonialize, e.g., www.examples.com:80 → org,example)/

Howard University mementos over time

First steps in Identifying Academic Migration using Memento and Quasi-Canonicalization
Web Archiving and Digital Libraries (WADL) Workshop 2022

@machawk1   12

Ethics of Investigating this Public Data

- Balancing transparency (as required by NSF) and responsible data dissemination
    - What if faculty don't want to be identified? Public then ⇏ Free to use now
    - What sort of anonymization will be sufficient for findings dissemination?
    - Many other outstanding questions!
- Another study! But look to recent work by:
    - Tiidenberg 2018, Maemura 2018, Cilliers & Viljoen 2021, among others

First steps in Identifying Academic Migration using Memento and Quasi-Canonicalization
Web Archiving and Digital Libraries (WADL) Workshop 2022

@machawk1   13

DREXEL UNIVERSITY
College of
Computing & Informatics

Learn more at the project web site:

https://HBCUmobility.github.io

Extended version of this initial work to be presented at

Project Status

This project runs from 2021 to 2024.

1. **Data collection, survey preparation, website creation**
2. Data cleaning, supplementation, and classification
3. Publication identification, survey identification, data publication
4. Professor productivity identification, pilot/distribute questionnaire, disseminate outputs
5. Professor modeling, follow-up interviews, visual dashboard creation
6. Regression testing, visualization dashboard dissemination, data dissemination, conclude interviews, present results

asis&t
OCTOBER 29 - NOVEMBER 1, 2022
WYNDHAM GRAND, PITTSBURGH, PA
85TH ANNUAL MEETING

These slides available at matkelly.com/wadl2022

Research support by
NSF SoS:DCI #2122525

Mat Kelly, **Deanna Zarrillo, Christopher Jackson, and Erjia Yan**

DREXEL UNIVERSITY
College of
Computing & Informatics

First steps in Identifying Academic Migration using Memento and Quasi-Canonicalization
Web Archiving and Digital Libraries (WADL) Workshop 2022

@machawk1   14

References

- Sawood Alam and Michael Nelson, "MemGator - A Portable Concurrent Memento Aggregator: Cross-Platform CLI and Server Binaries in Go", Presented at the Web Archiving and Digital Libraries Workshop (**WADL 2016**). Newark, NJ, June 2016.
- Liezel Cilliers and Kim Viljoen. "A framework of ethical issues to consider when conducting internet-based research." South African Journal of Information Management 23.1 (2021): 1-9.
- Marybeth Gasman, The changing face of historically Black colleges and universities. 2013.
- Mat Kelly, Lulwah M. Alkwai, Sawood Alam, Michael L. Nelson, Michele C. Weigle, and Herbert Van de Sompel, "Impact of URI Canonicalization on Memento Count," Presented at the Web Archiving and Digital Libraries Workshop (**WADL 2017**). Toronto, Canada, June 2017. arXiv:1703.03302
- Emily Maemura, "What's cached is prologue: Reviewing recent web archives research towards supporting scholarly use." Proceedings of the Association for Information Science and Technology 55.1 (2018): 327-336.
- Zehava Rosenblatt and Zachary Sheaffer, Brain drain in declining organizations: Toward a research agenda. Journal of Organizational Behavior: The International Journal of Industrial, Occupational and Organizational Psychology and Behavior, 2001. 22(4): p. 409-424.
- Sean B. Seymore, 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., 2005. 12: p. 287.
- Cassidy R. Sugimoto, et al., Scientists have most impact when they're free to move. Nature, 2017. 550: p. 29-31.
- Katrin Tiidenberg. "Ethics in digital research." The SAGE handbook of qualitative data collection (2018): 466-479.

DREXEL UNIVERSITY College of Computing & Informatics

First steps in Identifying Academic Migration using Memento and Quasi-Canonicalization
Web Archiving and Digital Libraries (WADL) Workshop 2022

@machawk1   15