Knowing thyself

Bibliometrics as knowledge management tools for research units and communities

July 4, 2023 - Guest Lecture @ University of Hagen

Philippe Mongeon
pmongeon@dal.ca
www.qsslab.ca

DALHOUSIE UNIVERSITY

CIR

Presentation outline

- Introduction to bibliometrics
- From research evaluation
- … to knowledge management
- Challenges and tips and tricks

What are bibliometrics?

- “The measurement of all aspects related to the publication and reading of books and documents.” (Otlet, 1934)
- “the application of mathematics and statistical methods to books and other media of communication.” (Pritchard, 1969)

In principle, bibliometrics could be applied to any type of documents, but in practice they are applied to scholarly outputs to **measure knowledge production, dissemination, and use.**

What’s in a name?

Bibliometrics is a widely used term to refer to the field, but it is not the only (and probably not the best) one. Other (quasi-)synonyms include:

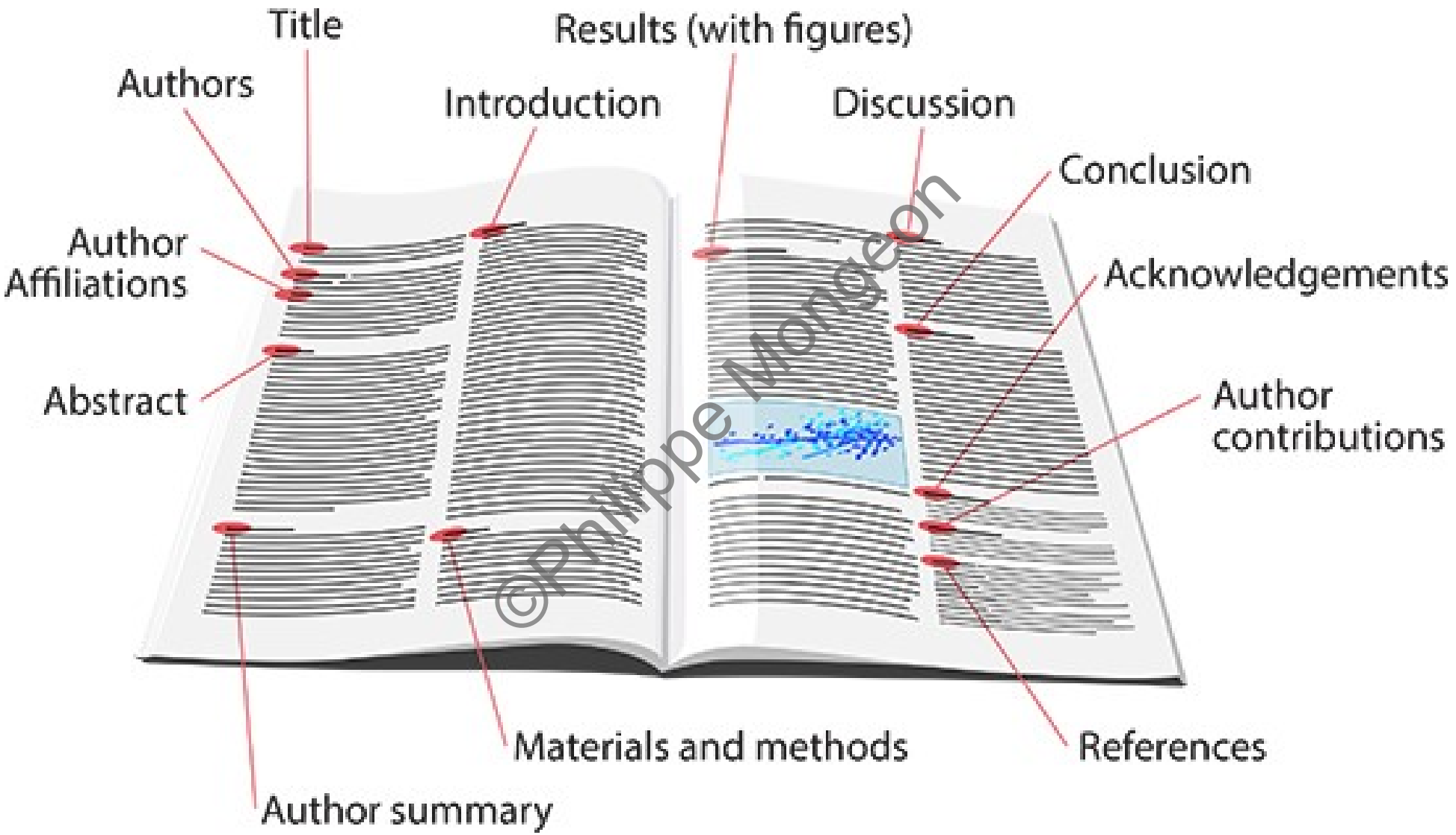
- Bibliometrics, Scientometrics, info(r)metrics, altmetrics, Quantitative Science Studies
- Science of science, research on research

Science and technology studies (STS) share the research object but not the methods (STS uses mainly qualitative methods).

Main assumptions of bibliometrics

- Peer-reviewed scholarly works are contributions to the advancement of knowledge (or units of knowledge production).
- Because researchers cite their sources, citations 1) indicate a relationship between two works and 2) can measure the use (or impact? Quality? Importance?) of a contribution to knowledge.

Bibliometric data

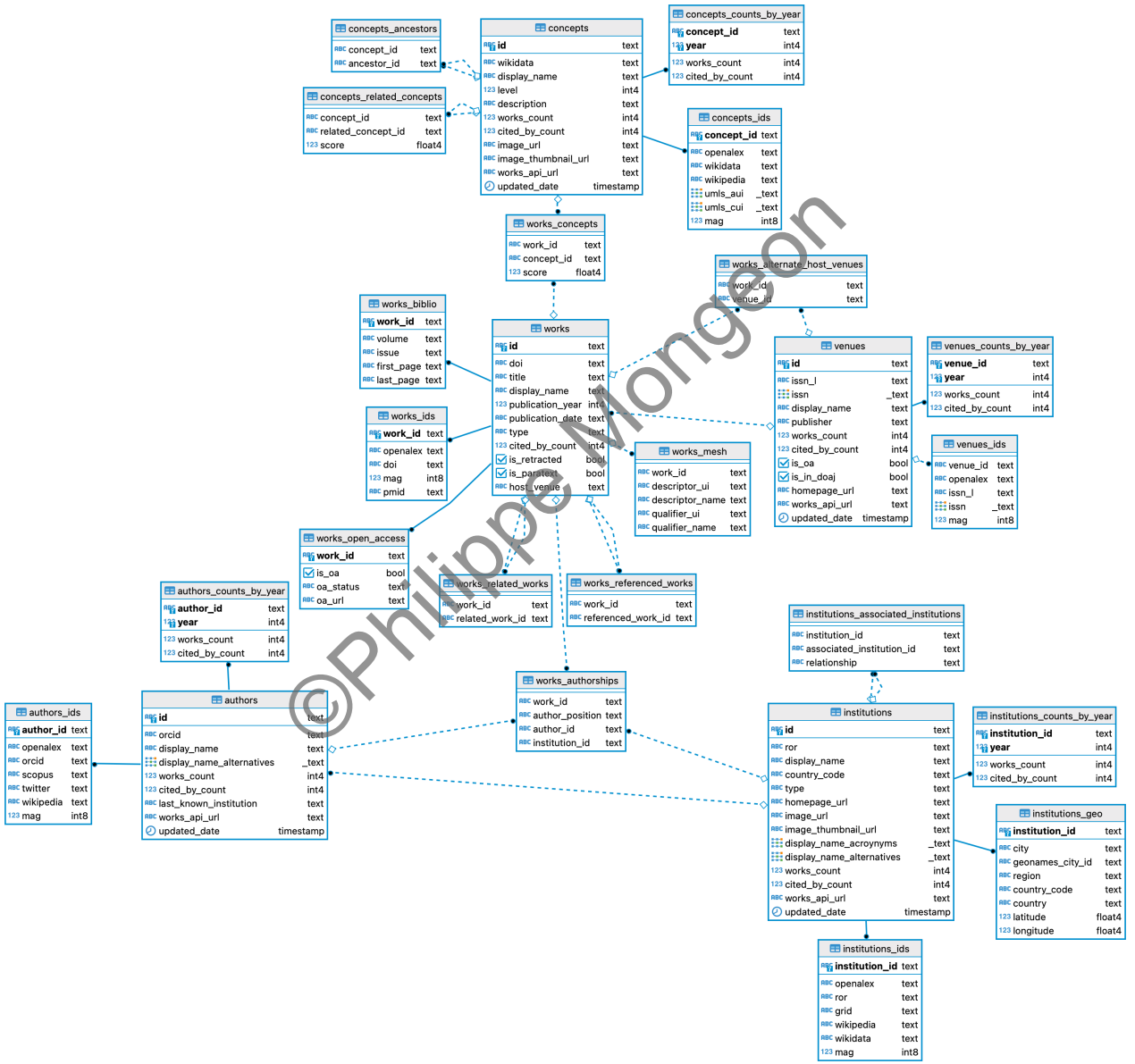


Source: https://askabiologist.asu.edu/explore/anatomy-of-an-article

Entities “knowable” through bibliometrics

- Works
- Journals
- Publishers
- Authors
- Research organizations (e.g. Universities
- Funders
- Research areas

Data structure



Bibliometric database vs other bibliograhic database

Main advantages of most of the bibliometric databases are:

- They index more metadata elements from the paper
- They enrich the metadata by adding elements (e.g., classifications, unique identifiers for authors and other entities)
- **They index citations,** which is why these databases are often called citation indexes (or scientific knowledge graph (SKG), which may be a better name since these databases generally include more than citations).

Applications of bibliometrics

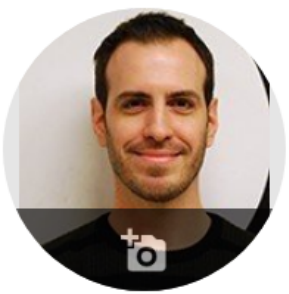
- Sociology of science
- History of science
- Science policy
- Library and Information Science
- Research evaluation
- Etc.

Bibliometric data sources

More details on my course website (in development)
https://pmongeon.github.io/bibliometrics-and-scholarly-communication/ch4.html

Google Scholar

- Probably the best coverage
- Black box
- Limited access to data



Philippe Mongeon

FOLLOWING

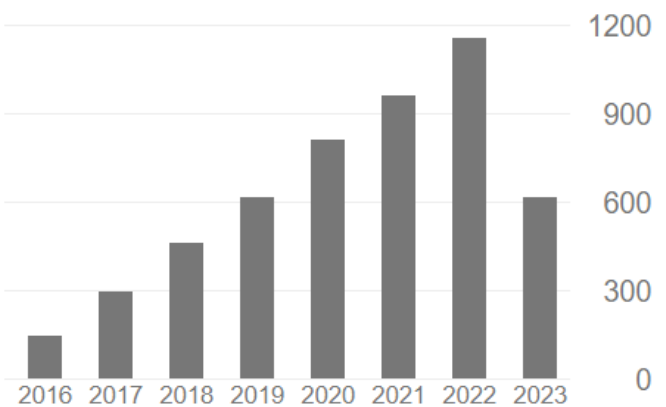
School of Information Management, Dalhousie University
Verified email at dal.ca

Quantitative Science Studies Scholarly Communication
Bibliometrics

TITLE	CITED BY	YEAR
The journal coverage of Web of Science and Scopus: a comparative analysis P Mongeon, A Paul-Hus Scientometrics 106 (1), 213-228	2569	2016
The Oligopoly of Academic Publishers in the Digital Era V Larivière, S Haustein, P Mongeon PloS one 10 (6), e0127502	1205	2015

Cited by VIEW ALL

	All	Since 2018
Citations	5255	4648
h-index	21	21
i10-index	30	30



Web of Science and InCites

Scopus, SciVal, and the ICSR lab

- Launched in 1996 by Elsevier as the first competitor to the Web of Science.
- Better coverage than Web of Science, but still limited.
- Scopus for **information retrieval**, SciVal for **evaluation**, and the ICSR lab for **advanced bibliometric research**.

Dimensions

- Relatively new database by Digital Science (owned by Springer-Nature).
- Broader coverage than Web of Science and Scopus.
- Has a free online search interface.
- Can request access to the full database or API for research purposes.

OpenAlex

- **Fully open** bibliometric database.
- Accessible through an API or a database snapshot.
- Most comprehensive.
- Metadata quality and completeness sometimes lacking.
- Still in development.

Other useful tools/resources

- VOSviewer
- Gephi
- Scite
- And many more…

From research evaluation to knowledge management

Bibliometrics and research evaluation

- Goal is to measure and/or compare performance and effectiveness.
- University Ranking (Institutions), Journal Impact factor (Journals), H-Index (Researchers).
- Often relying on external organizations operating in a black box.

Bibliometrics and knowledge management

- Goal is to gain knowledge about some unit.
- Networks visualizations, classification, topic models.
- Relies on comprehensive data access and understanding.

Case #1 - The breaking the silos in LIS project

CARL ABRC

mists
Maritime Institute for Science, Technology, and Society

École de bibliothéconomie et des sciences de l'information
Faculté des arts et des sciences
Université de Montréal
et du monde.

Conseil de recherches en sciences humaines du Canada

Social Sciences and Humanities Research Council of Canada

Canada

Objectives

- Build an **open and exhaustive database** of the scholarship produced by LIS academics and practitioners in Canada.
- **Promote the scholarship** produced by LIS academics and practitioners in Canada.
- **Encourage research collaboration** between academics and practitioners in Canada.

Data sources

- List of researchers from Canadian Academic Libraries and LIS department websites.
- Google Scholar and ORCID.
- OpenAlex

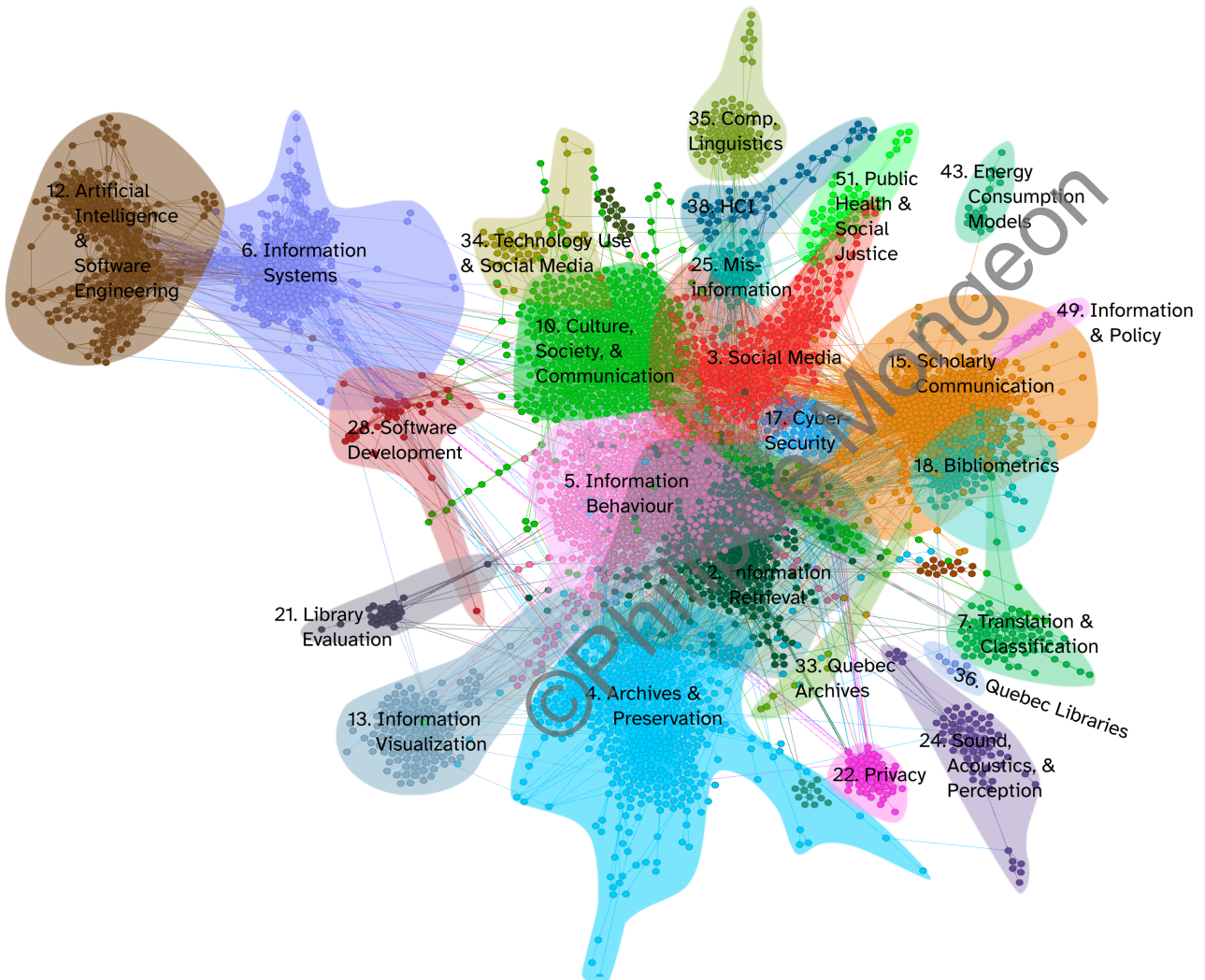
Outcome

- 2,630 individuals (2022 librarians, 608 academics) from 93 institutions.
- 6500+ publications (journal articles, books, book chapters, conference proceedings).
- OpenAlex author IDs and work IDs, Google Scholar IDs, ORCIDs, etc.
- Citation index (including links to records outside of the dataset).

Note on researcher-based field delineation

- Advantages
  - Less ambiguous than field-delineation based on topics.
  - Manageable scope.
  - Capture mutlidisciplinarity within the group/unit.
- Challenges
  - Author name disambiguation is tedious.
  - Mobility
  - Excludes LIS scholars with non-LIS affiliations and non-LIS scholars contributing to the LIS scholarship.

Map of research by LIS academics



Characterizing the clusters

- Tokenize abstract.
- Remove stop words.
- TF-IDF at the cluster level rather than the document level.
- Top TF-IDF weithed words to represent the clusters.

Label	Size	Top words
Information Behaviour	554	information; library; knowledge; study; student; lis; practice; organization; research; science; community; social; health; management; experience; public; program; read; immigrant; km; paper; professional; canadian; tag; finding; archive
Culture, Society, & Communication	483	medium; worker; digital; policy; technology; labor; social; public; communication; cultural; labour; ai; design; platform; privacy; music; system; political; television; economy; datum; information; form; right
Archives & Preservation	443	record; archival; preservation; digital; archive; blockchain; authenticity; information; electronic; library; curation; genre; document; datum;

Canadian LIS department specialization (publications)

Cluster	Dal	McGill	UofA	UBC	UdeM	uOttawa	UofT	Western
Information Behaviour	1.63	1.16	1.57	0.84	0.88	0.08	0.91	1.27
Culture, Society, & Communication	0.06	0.07	0.29	0.28	0.03	0.42	1.42	2.32
Archives & Preservation	0.13	0.82	0.36	2.43	0.93	0.94	1.26	0.18
Information Retrieval	1.61	2.09	2.89	1.58	0.88	0.33	0.39	0.79
Scholarly Communications	2.64	0.15	0.44	0.18	5.76	3.82	0.23	0.44
Information Systems	0.00	0.08	0.00	0.00	0.10	0.00	2.45	0.00
Artificial Intelligence and Software Engineering	0.31	0.03	0.00	0.00	0.00	0.25	2.42	0.03
Social Media	1.07	0.34	1.74	0.91	0.13	0.25	0.84	2.58
Information Visualization	0.00	0.06	0.00	1.26	0.15	0.15	0.22	4.92
Computational Linguistics	0.00	0.22	0.00	9.86	0.00	0.00	0.20	0.06
Bibliometrics	1.36	0.23	9.11	0.00	1.23	2.40	0.30	0.58

Canadian LIS department specialization (authors)

Cluster	Dal	McGill	UofA	UBC	UdeM	uOttawa	UofT	Western
Information Behaviour	1.30	0.76	1.46	0.88	1.35	0.36	0.71	1.48
Culture, Society, & Communication	0.03	0.25	0.36	0.27	0.02	0.79	1.60	1.72
Archives & Preservation	0.24	1.13	0.42	1.62	1.49	0.93	1.15	0.35
Information Retrieval	2.03	1.83	1.75	1.17	1.40	0.24	0.57	0.69
Scholarly Communications	1.47	0.99	0.80	0.70	2.45	3.36	0.72	0.47
Information Systems	0.00	1.97	0.00	0.00	0.58	0.00	2.02	0.00
Artificial Intelligence and Software Engineering	2.40	0.08	0.00	0.00	0.00	1.60	2.04	0.40
Social Media	1.09	0.30	2.17	1.94	0.29	0.37	1.08	0.73
Information Visualization	0.00	0.02	0.00	2.32	0.44	0.66	1.08	1.51
Computational Linguistics	0.00	0.04	0.00	5.23	0.00	0.00	1.08	0.01
Bibliometrics	0.72	0.11	5.54	0.00	0.17	8.26	0.39	1.26

Case study #2 - Race-related research in LIS

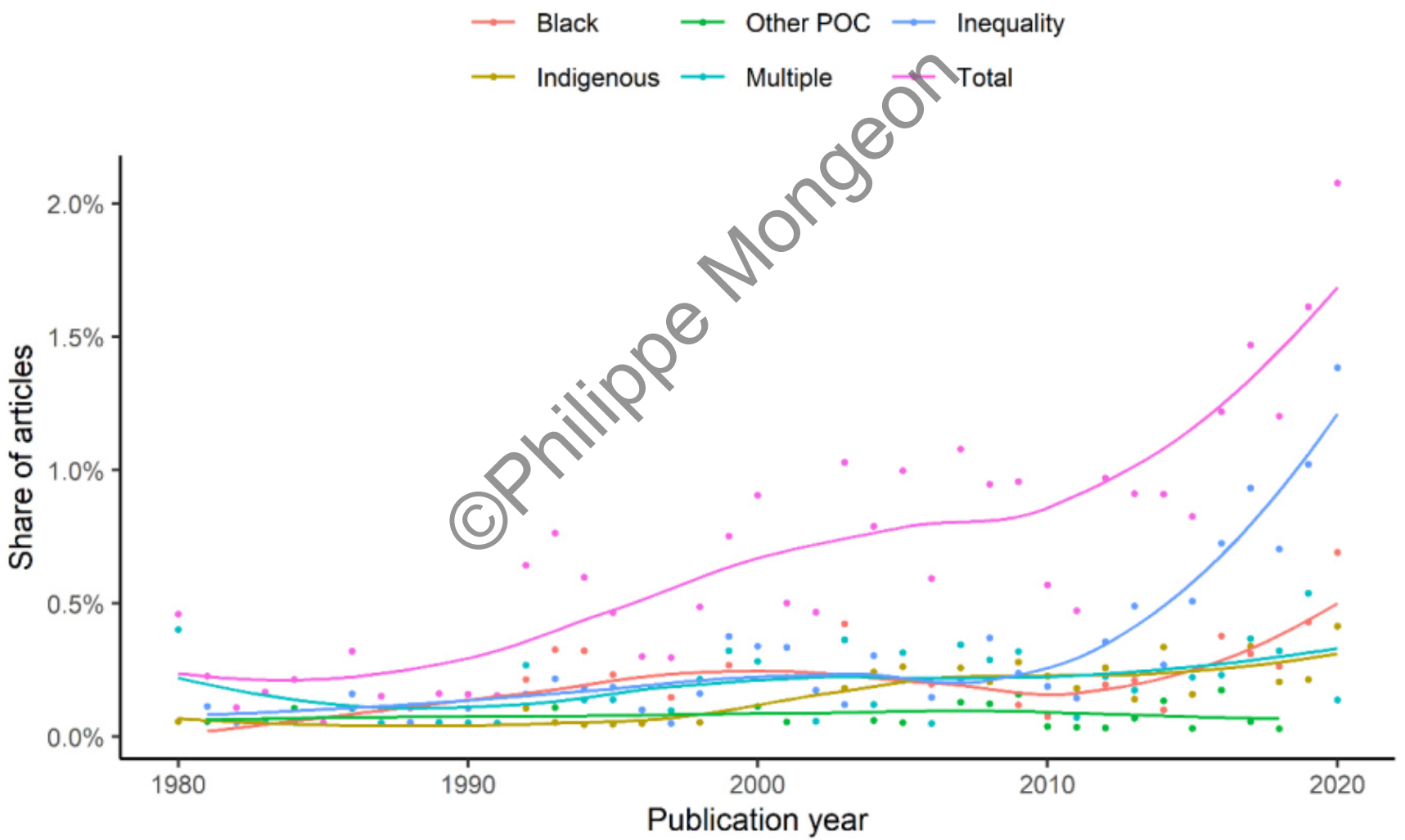
Objectives

1. Provide empirical evidence of the extent of LIS scholarship that includes race and/or racial inequity as an area of focus.
2. Analyzing the distribution of this research across subareas of the field.

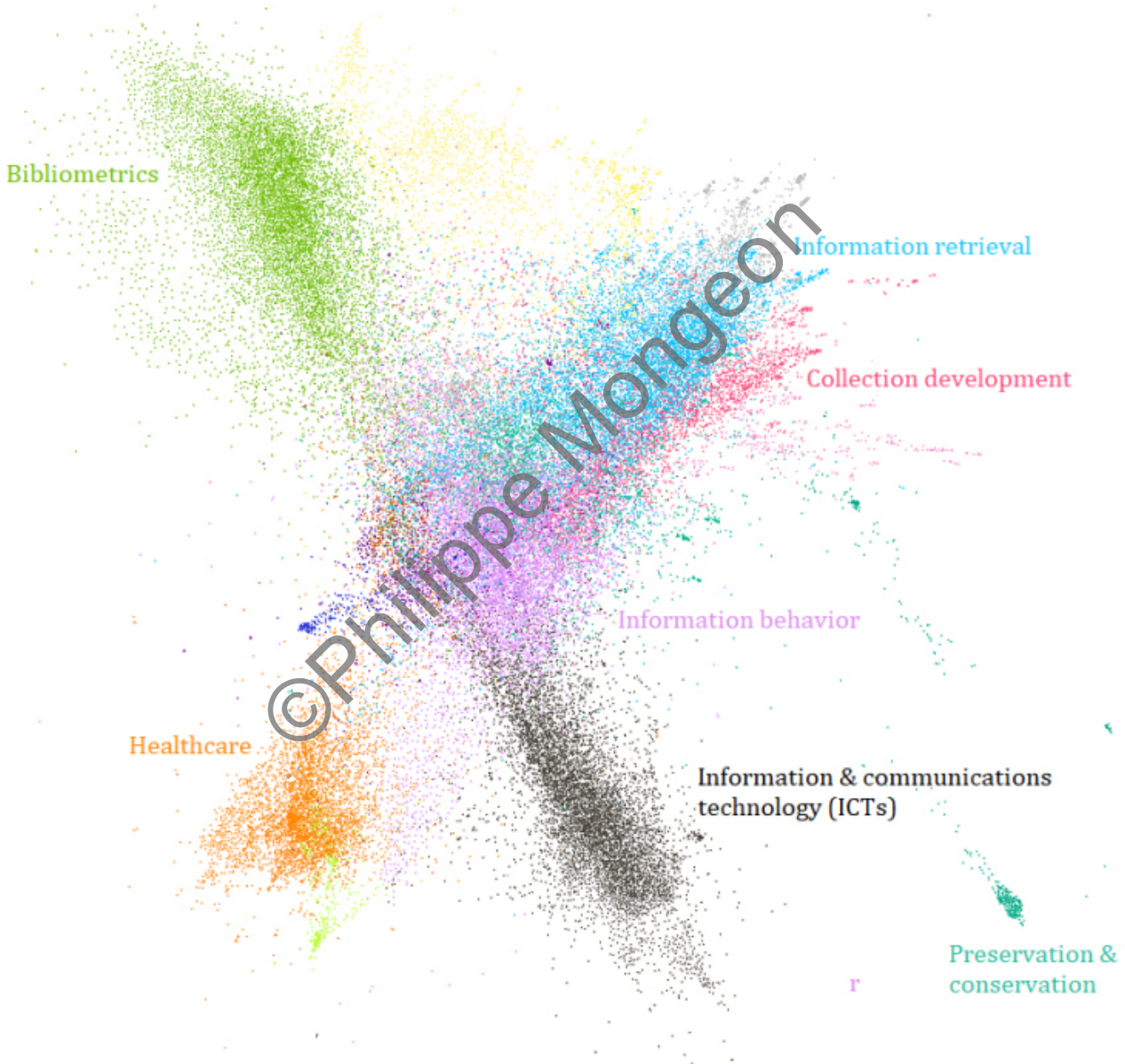
Data

- All publications from the Web of Science in the Library and Information Science specialty (NSF classification).
- A list of terms related to race and inequality.

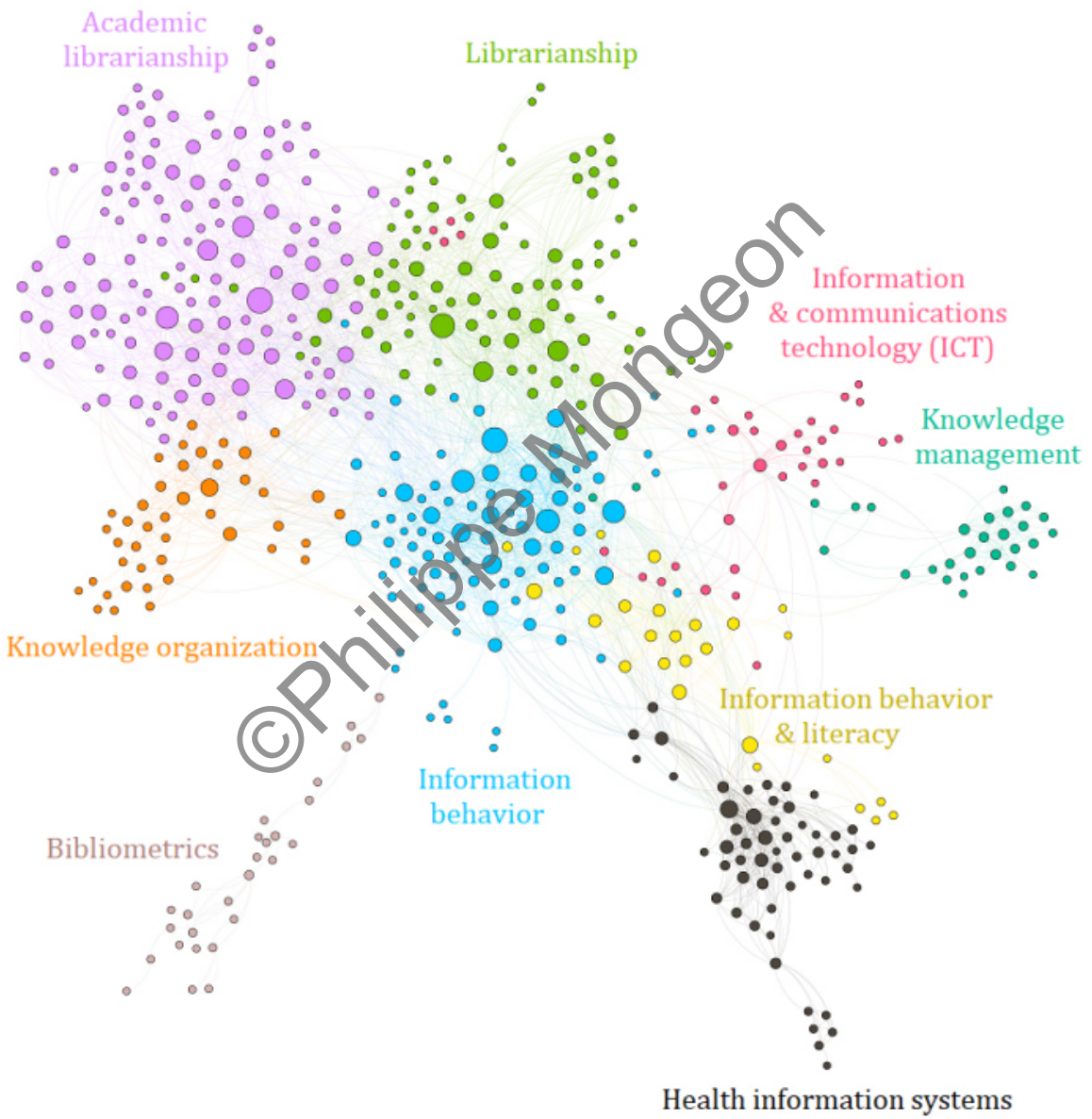
Trends in publications talking about race or inequality



Mapping the field



Map of race-related research only



Distribution and relative impact in the broader context

Group	All publications	BIPOC publications		
		N	%	ARC
Information behavior	7,265	159	2.19	0.96
Healthcare	4,480	84	1.88	1.61
Librarianship	1,629	75	4.6	0.64
Knowledge organization	2,529	68	2.69	0.69
ICTs	7,071	35	0.49	0.5
Bibliometrics	7,504	32	0.43	0.63
Health information	867	12	1.38	0.17
Digital libraries	2,015	10	0.5	0.74
Collection development	223	9	4.04	0.16
Preservation	348	7	2.01	0.5
Information retrieval	2,413	7	0.29	3.03

Starting a bibliometric project on the right foot

- Start with a clear goal and stick to it as much as possible.
- Choose data sources that meet your need (metadata available, coverage).
- Remember that technologies/tools can be overrated (focus on the goal and the data).
- Manage your urge to automate processes that can/should be done manually.
- Consult or collaborate with a bibliometrician.

Thank you!

Philippe Mongeon, Dalhousie University
pmongeon@dal.ca
www.qsslab.ca