Mapping SDGs by Faculty to find new interdisciplinary collaborations, a type of Linked Literature Analysis

Outline

Part 1: Mapping SDGs by Faculty

- 1. SDG publications by faculty :: *interdisciplinarity*
- 2. Identification of potential collaborations
- 3. Expert input in selecting collaborators

Part 2: Parallels with Linked Literature Analysis

- 1. Can LLA be automated?
- 2. Application to Bibliometrics & Scholarly Communication
- 3. LLA in LLMs

Take-home message:

- 3 techniques you can use to analyze/identify collaborations.
- **Predictive analytics** as new conceptual approach to Bibliometrics.

Mapping SDGs by Faculty - 3 steps

Demaine, J., **Bhatia**, Y., & **Whalen**, K. (2024). Mapping publications by sustainable development goal at the faculty level to highlight inter-faculty collaborations. *International Journal of Sustainability in Higher Education*. https://doi.org/10.1108/IJSHE-01-2024-0058

1. SDG publications by faculty :: *interdisciplinarity*

 Match records from Dimensions to records in McMaster's Research Information Management System (RIMS) based on DOI

2. Identification of potential collaborations

- Of all possible combinations of inter-Faculty co-authorship...
 - Join by:
 - SDG
 - ANZSRC 2020 "Fields of Research" categories Australian and New Zealand Standard Research Classification

3. Expert input in selecting collaborators

• Visualize bibliographic-coupled network of authors from 2 Faculties

16 SDGs categories for research





University output as categorized by SDG used in



Times Higher Education Impact Rankings

McMaster University (2022):

- 9th in Canada
- 37th in World

...but 80th in



Some institutions do better in Impact than World rankings.

Why interdisciplinarity?

New Frontiers in Research Fund

• **\$142 million** for "Canadian-led, **interdisciplinary** research projects" (March 13, 2025)

Tri-Agency Interdisciplinary Peer Review Committee

- **SSHRC** Insight Grants
- **CIHR** Project Grant
- **NSERC** Discovery Horizons

"Applications... must represent research across disciplines and subject areas... that clearly articulate **interdisciplinary approaches**."

"Interdisciplinarity must be a key characteristic of proposed projects, where the project goals could not be achieved without an interdisciplinary approach."

Methods

Dimensions: 32,605 publications

- 1+ author @ McMaster
- 6 years: January 2018 \rightarrow December 2023
- Research Article, Review Article, Conference Paper
- Matched 25,406 publications by DOI in RIMS
- 8,594 had SDG category by Dimensions

Data & Methods available on Figshare:

Demaine, Jeffrey; Bhatia, Yash; Whalen, Kate (2024). Supplementary materials: "Mapping publications by Sustainable Development Goal at the faculty level to highlight inter-faculty collaborations". figshare. Dataset. <u>https://doi.org/10.6084/m9.figshare.25075727.v1</u>



1- SDG publications by Faculty

- **8,594** publications map to **9,345** categorizations-by-affiliations
 - Because...

Distribution is highly skewed:

- Health Sciences: 70.5% of McMaster's total.
 - Of these, 88.7% address SDG #3 Good Health and Well Being.
- **Engineering**: 1,179 publications (12.6% of total)
 - of which 725 (61.5%) in SDG #7- Affordable and Clean Energy.
- **Science**: 10.6% of output
 - 47% are also SDG #3 Good Health and Well Being



1 - SDG publications by Faculty

	Health					
SDG	Sciences	Engineering	Science	Sciences	Business	Humanities
1 No Poverty	5	1	2	5	1	1
2 Zero Hunger	54	11	31	8	4	1
3 Good Health and Well Being	5,845	175	427	222	68	17
4 Quality Education	446	18	27	39	15	25
5 Gender Equality	54	2	6	34	10	3
6 Clean Water and Sanitation	9	28	29	3	1	
7 Affordable and Clean Energy	9	725	89		3	
8 Decent Work and Economic Growth	10	1	9	35	11	1
9 Industry, Innovation and Infrastructure	1	31	1		19	
10 Reduced Inequalities	10	1	9	19	4	1
11 Sustainable Cities and Communities	22	25	16	5	7	
12 Responsible Consumption & Production	2	19	2		8	
13 Climate Action	8	111	115	8	5	
14 Life Below Water	5	4	42	2	2	
15 Life on Land	2	22	97	4		2
16 Peace, Justice and Strong Institutions	97	4	12	31	14	22

1 - SDG publications: interdisciplinarity

	Matched in RIMS	Single Faculty			Two+ Faculty		
SDG		Pubs.	Cites	FCR	Pubs.	Cites	FCR
1 No Poverty	10	6	5	4.5	4	10	4.8
2 Zero Hunger	103	97	25.2	11.9	6	2.2	0.4
3 Good Health and Well Being	6,420	6,064	23.2	12.3	356	14.3	6.9
4 Quality Education	540	503	11.7	6.8	37	5.2	4.3
5 Gender Equality	106	99	18	15.2	7	7.9	3.3
6 Clean Water and Sanitation	66	61	13.3	5.1	5	2.8	0.7
7 Affordable and Clean Energy	815	787	19.3	6.8	28	21.6	5.5
8 Decent Work and Economic Growth	63	59	9.7	5.4	4	4.5	5.7
9 Industry, Innovation and Infrastructure	52	50	31.8	20.8	2	12	9.6
10 Reduced Inequalities	43	41	14.6	10.2	2	5	4.8
11 Sustainable Cities and Communities	66	56	10.9	6.8	10	11	6.8
12 Responsible Consumption and Production	29	27	29.4	3.7	2	2	0.3
13 Climate Action	238	223	24.4	6.4	15	30.3	5.8
14 Life Below Water	55	55	15	6			
15 Life on Land	123	119	26.7	8	4	9.3	8.7
16 Peace, Justice and Strong Institutions	178	175	7.7	5.5	3	2.3	
17 Partnerships for the Goals	12	10	19.7	10.9	2	15.5	8.6
Total	8,919	8,432			487		
Average			18	8.6		9.7	5.1

2 – Matching research across Faculties

Not counting SDG 3

Have NOT already co-authored

Matching on SDG + ANZSRC

8,571 cross-Faculty pairings



2 - Matching research across Faculties

```
SELECT DISTINCT ME1.User FULLname, ME1.Faculty, ME1.Department, ME2.User FULLname,
ME2.Faculty, ME1.Department, ME1.SDG, ME1.ANZSRCfield
       FROM MatchedExpanded AS ME1, MatchedExpanded AS ME2
       WHERE ME1.Article ID != ME2.Article ID
       AND ME1.Faculty != ME2.Faculty
       AND ME1.User FULLname != ME2.User FULLname
       AND ME1.SDG == ME2.SDG
       AND ME1.ANZSRCfield == ME2.ANZSRCfield
       AND ME1.ANZSRCcode > 1000
       AND ME1.SDG != "3 Good Health and Well Being"
       AND NOT EXISTS (SELECT *
              FROM CoAuthoredPreviously
              WHERE (ME1.User FULLname == CoAuthoredPreviously.CoAuthor1
              AND ME2.User FULLname == CoAuthoredPreviously.CoAuthor2)
              OR (ME1.User FULLname == CoAuthoredPreviously.CoAuthor2
              AND ME2.User FULLname == CoAuthoredPreviously.CoAuthor1))
              ORDER BY ME1.User FULLname
              LIMIT 60000;
```

2 - Identification of potential collaborations



Note $A \rightarrow B \rightarrow C$ linkage

3 – Expert input – test case

New Frontiers in Research fund announces grant on "sustainable transportation"

Our dataset contains 8,571 pairs of authors (*without SDG 3*).

The computational part is done, now *human judgement* is required.

Filter by:

- **SDG 7** "("Affordable and Clean Energy")
- **SDG 11** ("Sustainable Cities and Communities")

Add more granularity: ANZSRC fields of research:

- o 3304 Urban and Regional Planning
- 3509 Transportation, Logistics and Supply Chains
- 4011 Environmental Engineering

Select two Faculites:

- Engineering
- Science



3 - Expert input



19 from the Faculty of Engineering9 from the Faculty of Science

Bibliographic coupling of their previous articles arranges people by their citing similarity.

Hanna Maoh's previous publications are on transportation:

"Battery electric vehicle acquisition timeframes in Canadian fleets" (*Transportation Planning and Technology*)

"Examining the Variability of Crossing Times for Canadian Trucks at the Three Major Canada–U.S. Border Crossings" (*Professional Geographer*)

Saiedeh Razavi is near Maoh. His research seems similar:

"Adoption patterns of autonomous technologies in Logistics: evidence for Niagara Region." (Transportation Letters)

"Transportation data visualization with a focus on freight: a literature review" (Transportation Planning and Technology)

PART 2 - Parallels with Linked Literature Analysis

No longer talking about SDGs and Faculties!

Parallels with Linked-Literature Analysis

Don Swanson



"Undiscovered Public Knowledge" in PubMed

- Fish Oil and Raynaud's Disease (1986)
- Migraine and magnesium
- Somatomedin C and arginine



Note $A \rightarrow B \rightarrow C$ linkage

Similar patterns can be applied to Schol Comm

Finding potential collaborators across Faculties is an instance of Swanson's LLA technique.

- So LLA is not restricted to PubMed or biomedicine.
- We can apply hypothesis-generation in Bibliometrics.

Further possibilities:

- **One-node A-B-C** (Swanson's original Linked-Literature Analysis)
- **Two-node A-B-C** (Smalheiser's ARROWSMITH tool)
- Multi-step paths (Baek et al., 2017; Hossain et al., 2012; Sebastian, Siew, & Orimaye, 2017)
- Ranking of **shared/implicit relationships** (Wren et al. 2004)

Can LLA be automated?

Demaine, J., **Martin**, J., & **De Bruijn**, B. (2003). Haystacks and hypotheses. *Proceedings of the American Society for Information Science and Technology*, *40*(1), 59–64. <u>https://doi.org/10.1002/meet.1450400107</u>

Pretend it is 1995. Search PubMed for all MeSH terms that do not co-occur ("A" and "C"), but which share a third co-occurring term ("B"). How many A+C co-occurrences happen after 1995?

Result = 8%

BITOLA - Biomedical Discovery Support System University of Ljubljana, Slovenia - circa 2005 https://ibmi3.mf.uni-lj.si/bitola/

ARROWSMITH

University of Illinois https://arrowsmith.psych.uic.edu/arrowsmith_uic/



LLA + LLMs

Don Swanson (ARROWSMITH, BITOLA) relied on the **MeSH** controlled vocabulary.

 \rightarrow LLA limited to BioMedical field

Now, LLMs can *infer semantic relationships* from raw text.

→ Hypothesis generation by <u>non-specialists</u> now possible *in any field*

DiscipLink: LLM tool that helps researchers identify potentially relevant topics in other (interdisciplinary) fields:

Zheng, C., Zhang, Y., Huang, Z., Shi, C., Xu, M., & Ma, X. (2024). **DiscipLink: Unfolding interdisciplinary information** seeking process via human-AI co-exploration. In *Proceedings of the 37th Annual ACM Symposium on User Interface Software and Technology*, 1–20. UIST '24. New York, NY, USA: Association for Computing Machinery. <u>https://doi.org/10.1145/3654777.3676366</u>.

Resurgence of research in UPK driven by AI:

Alkan, A. K., Sourav, S., Jablonska, M., Astarita, S., Chakrabarty, R., Garuda, N., Khetarpal, P., et al. (2025). A survey on hypothesis generation for scientific discovery in the era of Large Language Models. *arXiv*. https://doi.org/10.48550/arXiv.2504.05496.

Summary

Part 1: Mapping SDGs by Faculty

- 1. Basic SDG-by-Faculty analysis
- 2. Re-combine metadata to calculate all SDG + ANZSRC matches across faculties
- 3. Target a specific grant: filter dataset to find handful of potential collaborators

Part 2: Similarity with Linked Literature Analysis

- A. ...implies applicability of hypothesis-generation techniques in Bibliometrics
- B. Integration of LLA/UPK with AI enables hypothesis generation in any field.
- > 3 techniques you can use to analyze/identify collaborations.
- > **Predictive analytics** as new conceptual approach to Bibliometrics.

Mapping publications by SDG at the faculty level... (2024): <u>https://doi.org/10.1108/IJSHE-01-2024-0058</u>

• Methods & data (Figshare): <u>https://doi.org/10.6084/m9.figshare.25075727.v1</u>

Haystacks and Hypotheses (2003): https://doi.org/10.1002/meet.1450400107