



Social Media Data: Practical and Ethical Collection, Visualization, and Management

Thursday, March 19, 2026

1:00pm – 2:30pm (**Online**)

 **Sherman
Centre**
for Digital Scholarship

Social Media Data:

Practical and Ethical Collection,
Visualization, and Management

In order of appearance:

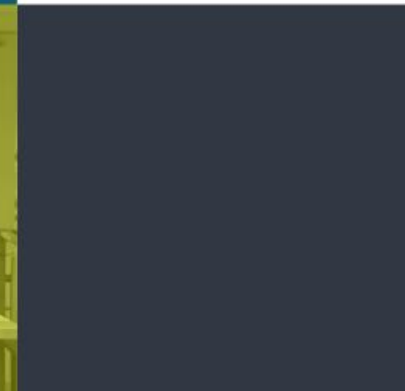
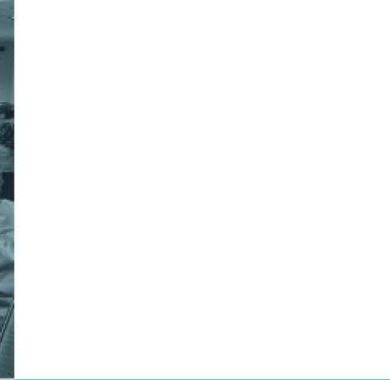
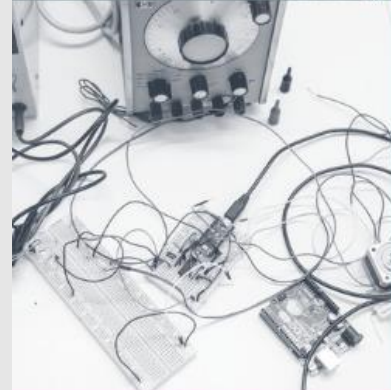
Alexis-Carlota Cochrane, Digital Scholarship Coordinator

John Fink, Digital Scholarship Librarian

Subhanya Sivajothy, Data Analysis and Visualization Librarian

Danica Evering, Research Data Management Specialist

March 19, 2026





 **Land Acknowledgement**

McMaster University is located on the traditional territories of the Mississauga and Haudenosaunee Nations. Settlers have responsibilities under the Silver Covenant Chain Wampum, part of the 1764 Treaty of Niagara.

Georgia Kirkos, "Spring Bloom Campus 2025," April 29, 2025, McMaster University, Hamilton, Ontario, Canada
<https://mcmaster.assetbank.app/assetbank-mcmaster/action/viewDownloadSharedAsset?shareId=9965400e-e49f-42d9-83cf-7db87db6029b>

Social Media Data

Practical and Ethical Collection, Visualization, and Management

Today's Facilitators:



Alexis-Carlota Cochrane,
MA



John Fink,
MLIS



Subhanya Sivajothy,
MI



Danica Evering,
MA

Outline



Setting the Table [10 mins]



1. How to collect social media data [30 mins]



2. How to analyze social media data [20 mins]



3. How to manage and care for social media data [20 mins]



Discussion [10 mins]



Code of Conduct, Session Recording, & Privacy

Code of Conduct: The Sherman Centre and the McMaster University Libraries are committed to fostering a supportive and inclusive environment for its presenters and participants. Please refer to our code of conduct webpage for more information: scds.ca/events/code-of-conduct

Session Recording: This session is being recorded with the intention of being shared publicly via the web for future audiences. In respect of your privacy, participant lists will not be shared outside of this session, nor will question or chat transcripts. Questions asked via the chat box will be read by the facilitator without identifying you. Note that you may be identifiable when asking a question during the session in an audio or visual format.



Certificate Programs

The Sherman Centre for Digital Scholarship Certificate of Attendance

The Sherman Centre's certificate program recognizes attendance at our workshops. It complements degree training, supports the development of critical competencies in data analysis, research data management, and digital scholarship, and formalizes core skills fostered by our workshops.

Participants are invited to collect **seven** workshop points to receive a certificate of attendance. To verify your participation in today's workshop, we will provide a code and additional instructions at the end of the session.

You can learn more about the certificate program at scds.ca/certificate-program

The Canadian Certificate for Digital Humanities

This workshop is also eligible for the Canadian Certificate for Digital Humanities. To learn more about the certificate, visit ccdhhn.ca. You can also contact local liaison Alexis-Carlota Cochrane at scds@mcmaster.ca.

The Sherman Centre for **Digital Scholarship**

The Sherman Centre for Digital Scholarship (SCDS) is a research centre located in McMaster University's Mills Memorial Library offering consulting, training, and technical support for all aspects of digital scholarship. SCDS staff support students, faculty, staff, and community members on and beyond campus who are engaged in digital scholarship projects.

Core service areas include:

- Data Discovery and Collection
- Statistical and Geospatial Data Analyses
- Text Mining and Natural Language Processing
- Data Visualization
- Digital Storytelling
- Podcasting
- 3D Printing
- Large Language Models
- Open Scholarship
- Bibliometrics and Research Impact Analysis
- Web Archiving
- Social Media Data Collection
- Digital Research Ethics
- Research Data Management

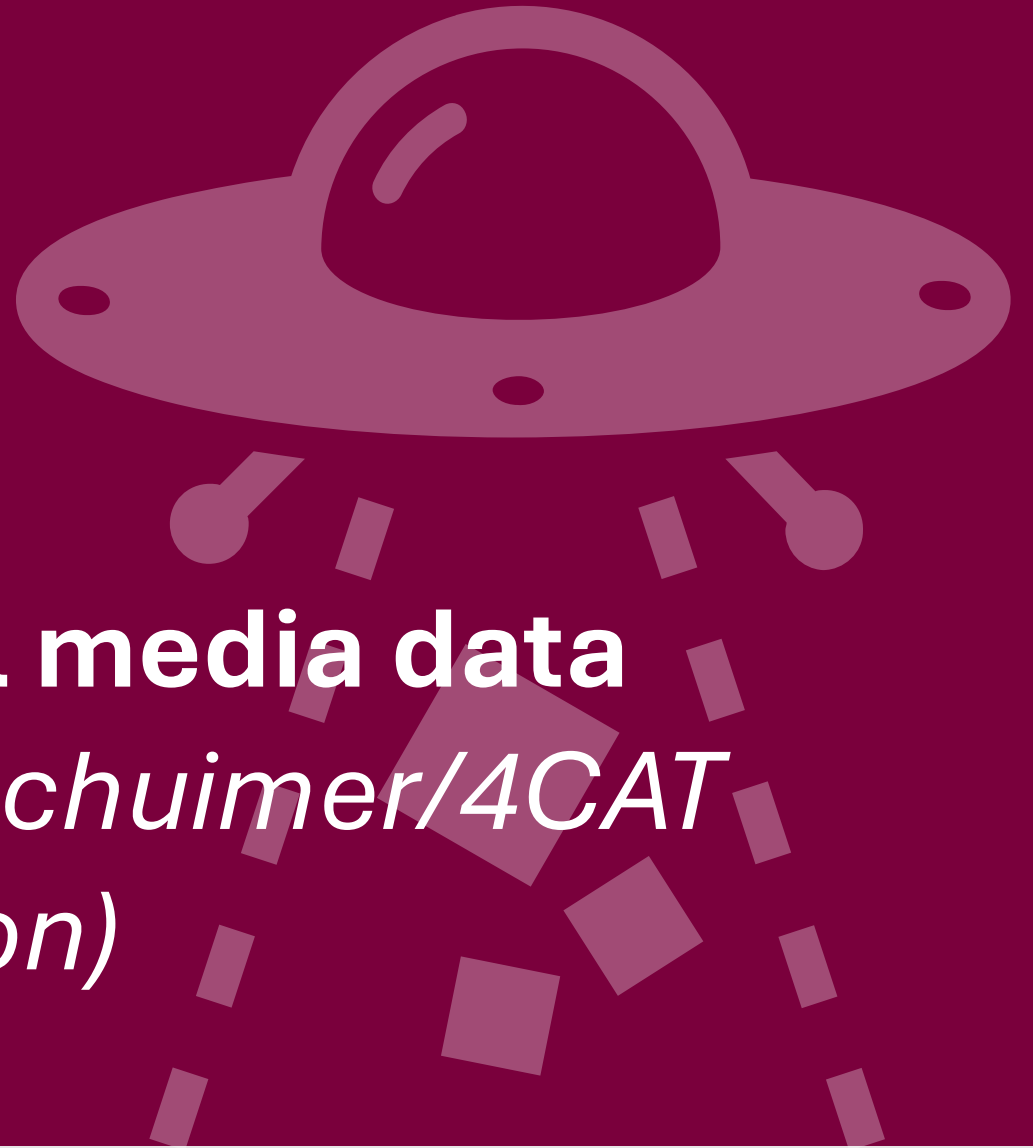
Learn more at scds.ca or email scds@mcmaster.ca



Learning Objectives

By the end of this workshop, you will be able to:

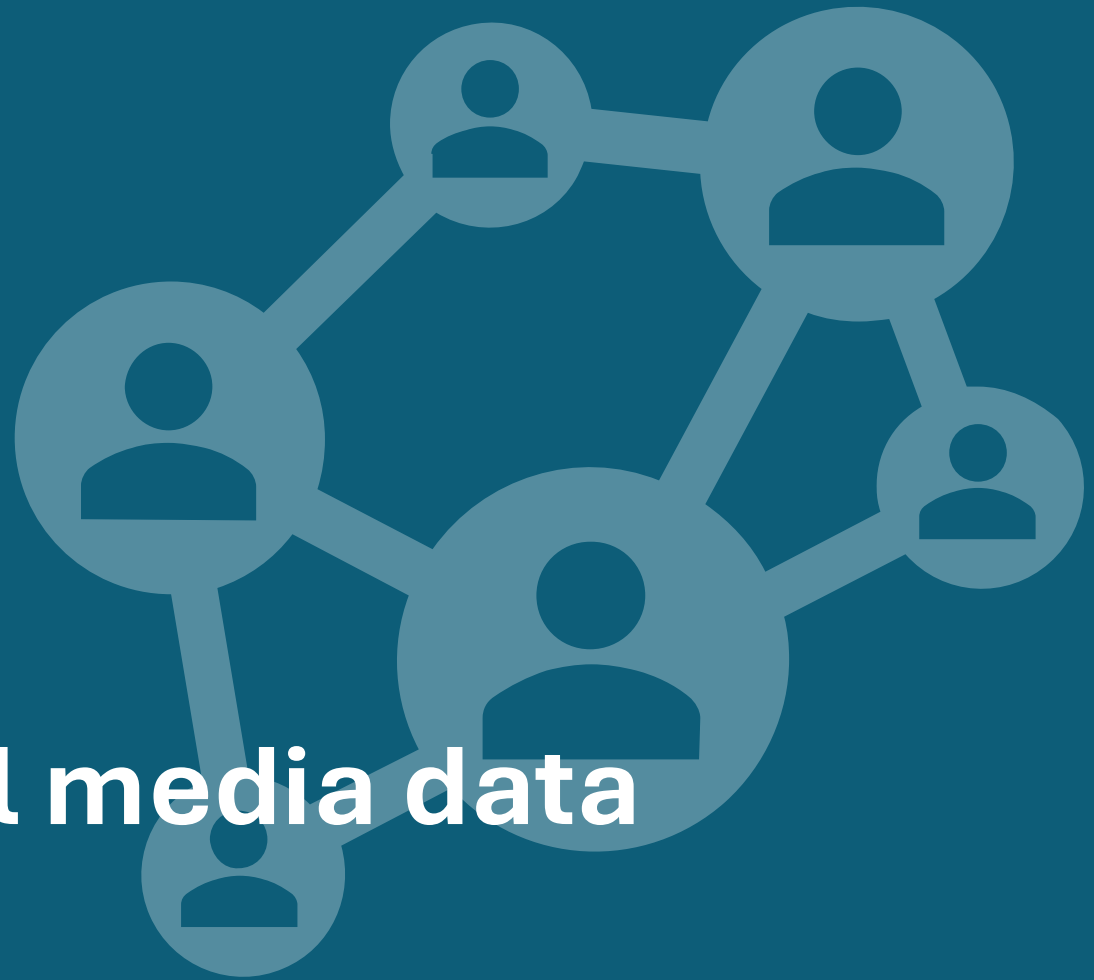
- Collect social media data using both manual and automated scraping tools like Zeeschuimer and 4CAT.
- Analyze datasets to identify key trends and insights with introductory data exploration techniques.
- Organize and manage social media data effectively by applying recommended best practices and tools.
- Understand and address ethical challenges in social media research, including protecting privacy and minimizing harm at all stages.



1. How to collect social media data

(John – 30 minutes; Zeeschuimer/4CAT demo + manual collection)

2. How to analyze social media data



scds.ca
scds@mcmaster.ca

A joint initiative of McMaster's
Faculty of Humanities and
McMaster University Libraries



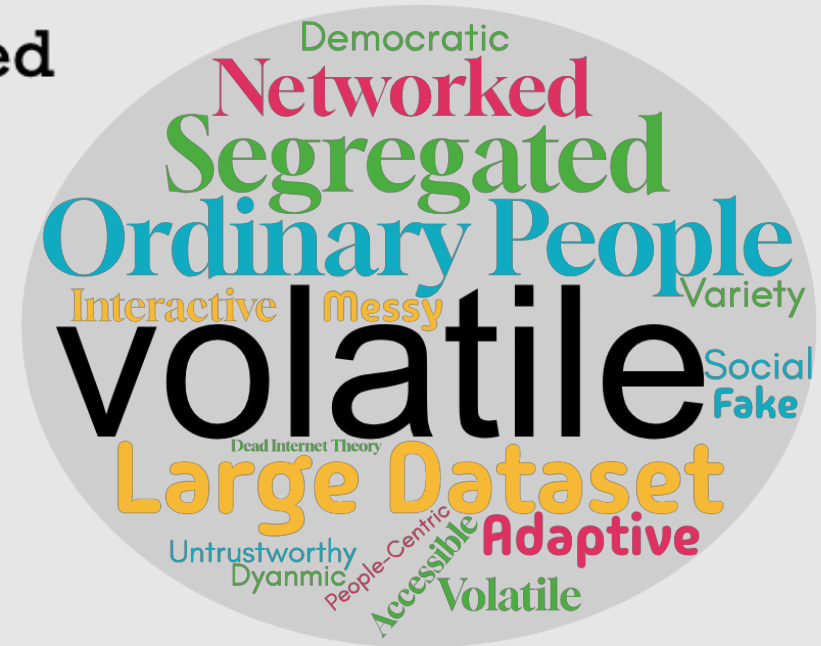
Stage	What to Focus on?	Reflect Upon
First Stage: Role of technology	Place for data collection	How the platform affects the data (the possibilities that any object, within a social media platform, offers for action) and social affordances (how dominant groups shape the kind of practices that can be observed)
	Phenomenon per se	
Second Stage: Collecting the digital data	Platform (e.g., Instagram)	On what can (or cannot) be known with the data that will be generated from each of these initial approaches (e.g., what Facebook friendship means? Will it be possible to know something about friendship offline using data only from a FB group?)
	Social phenomenon (e.g., friendship)	
	Participants (e.g., 10 feminist activists)	
	Instrumental case study (e.g., #metoo on Twitter)	
Third Stage: Analytical model	Communication (What ideas are people sharing?)	What data qualities the project is interested in? What are the underpinnings of the research and what is the relationship between the data generated and the focus of the project. Explore the options for reconciling what people say that they do with what they actually do.
	Interaction (How are they interacting with other people?)	
	Experience (How do users feel on this app?)	

Characteristics of Social Media Data



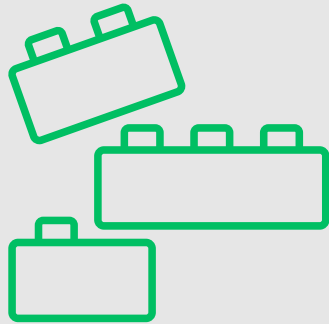
Giovanni Boldini - Conversation at the Cafe.jpg

- Heterogeneous
- Large Scale and Messy
- Network Based



Data Qualities	Examples in Social Media Data	Typical Approach(es)	Typical Analysis (Examples)
Verbal (and textual) data	Expressions, ideas, opinions, attitudes, and so on Online conversations (e.g., Facebook posts and or comments, YouTube videos; comments, Tweets)	Social semiotics	Discourse analysis Thematic analysis Content analysis (including sentiment analysis, semantic analysis)
Visual data	Photos, emoticons, and other visual productions such as maps, graphs, GIFs, and so on (e.g., selfies, maps)	Social semiotics	Thematic analysis Visual analysis Discourse analysis
Audiovisual data	Videos, screen recordings, video games recordings, and so on (e.g., YouTube or Vimeo videos)	Social semiotics	Thematic analysis Content analysis Visual analysis Multimodal discourse analysis
Digital actions	Gestures or concurrent actions. (e.g., like, Retweet, Favouriting, Endorsing)	Ethnomethodology	Conversational analysis Interaction order Membership categorization
Meta-data	Data which are not always visible from the interface (e.g., geo-location, logs, frequencies, interactions)	Ethnomethodology	Interaction order Content analysis
Sensors data or self-tracking	Data from wearables, mobile apps, GPS, self-tracking technologies (e.g., miles running, calories burnt, movements)	Ethnomethodology	Interaction order Content analysis
Interfaces	Design of social media or mobile apps	Social semiotics	Discourse analysis Thematic analysis Content analysis Visual analysis
Social network data	Information about the links and relationships in social media (e.g., Follow/Followers, Friends, Connexions)	Ethnomethodology	Interaction order Qualitative social network analysis (SNA)

(A Few) Strategies for Analysis



Qualitative Analysis

Thematic Analysis

Discourse Analysis

Content Analysis

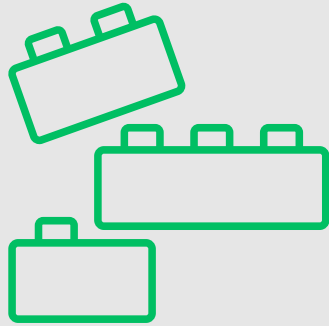


Network Analysis

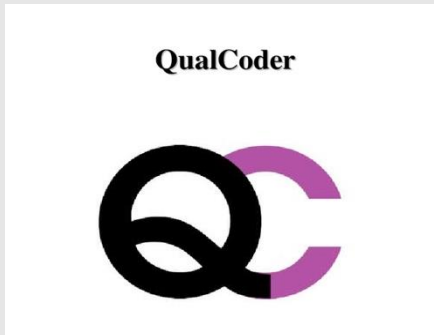


Textual Analysis

Strategies for Analysis



Qualitative Analysis

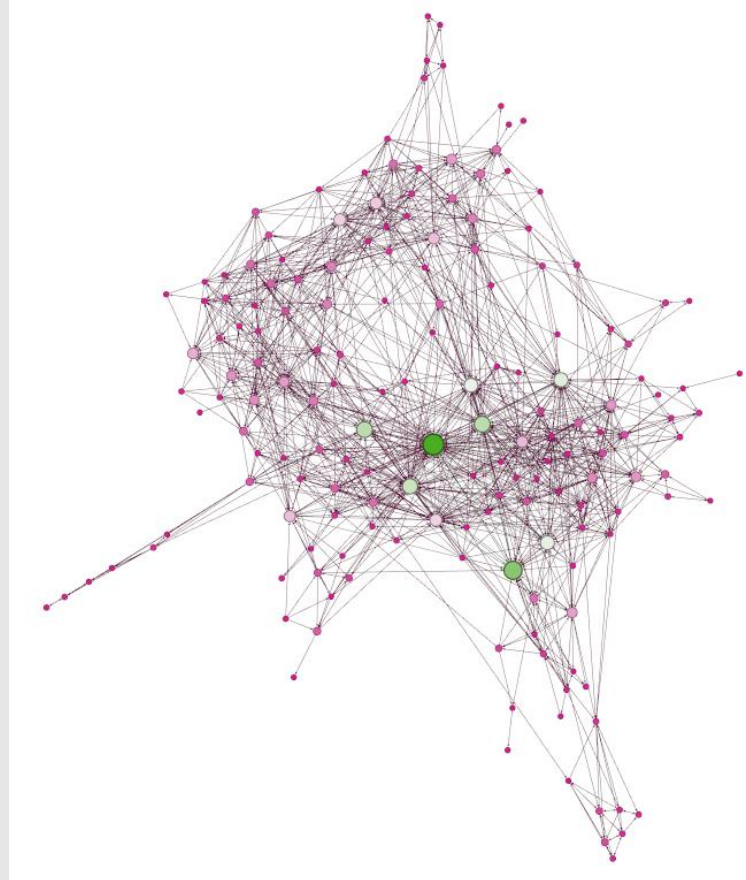


Network Analysis



Textual Analysis





scds.ca
scds@mcmaster.ca

A joint initiative of McMaster's
Faculty of Humanities and
McMaster University Libraries



Fig. 1

THE ALTERNATIVE INFLUENCE NETWORK ON YOUTUBE

GUEST APPEARANCES ON THE NETWORK FROM JANUARY 1, 2017 THROUGH APRIL 1, 2018

The graph is a partial representation of collaborative connections within the Alternative Influence Network (AIN)—a network of controversial academics, media pundits, and internet celebrities who use YouTube to promote a range of political positions from mainstream versions of libertarianism and conservatism to overt white nationalism. While collaborations can sometimes consist of debates and disagreements, they more frequently indicate social ties, endorsements, and advertisements for other influencers.

Each line indicates that two connected influencers appeared in the same Youtube video during the period of January 1, 2017 and April 1, 2018, serving as guests, hosts, or collaborators. The size of nodes are determined by the number of other influencers with whom they connect—demonstrating how much a given influencer serves as a conduit for viewers to other influencers in the AIN. The colors of nodes are determined by their total connectivity within the network, or how close the influencer is to all other influencers.

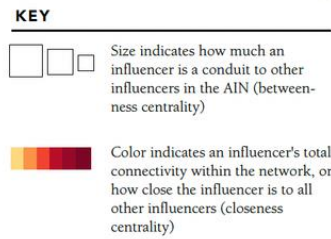
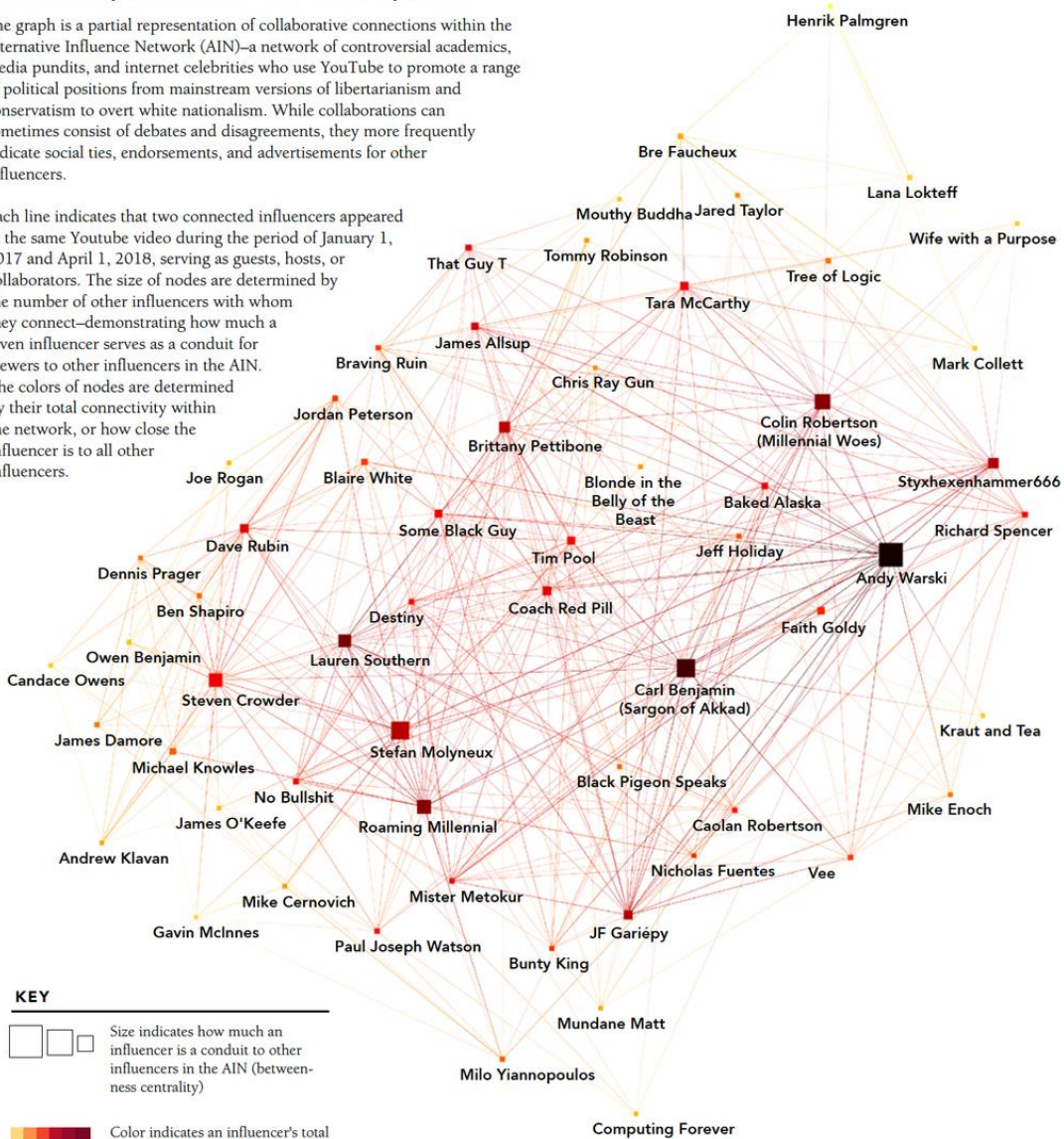


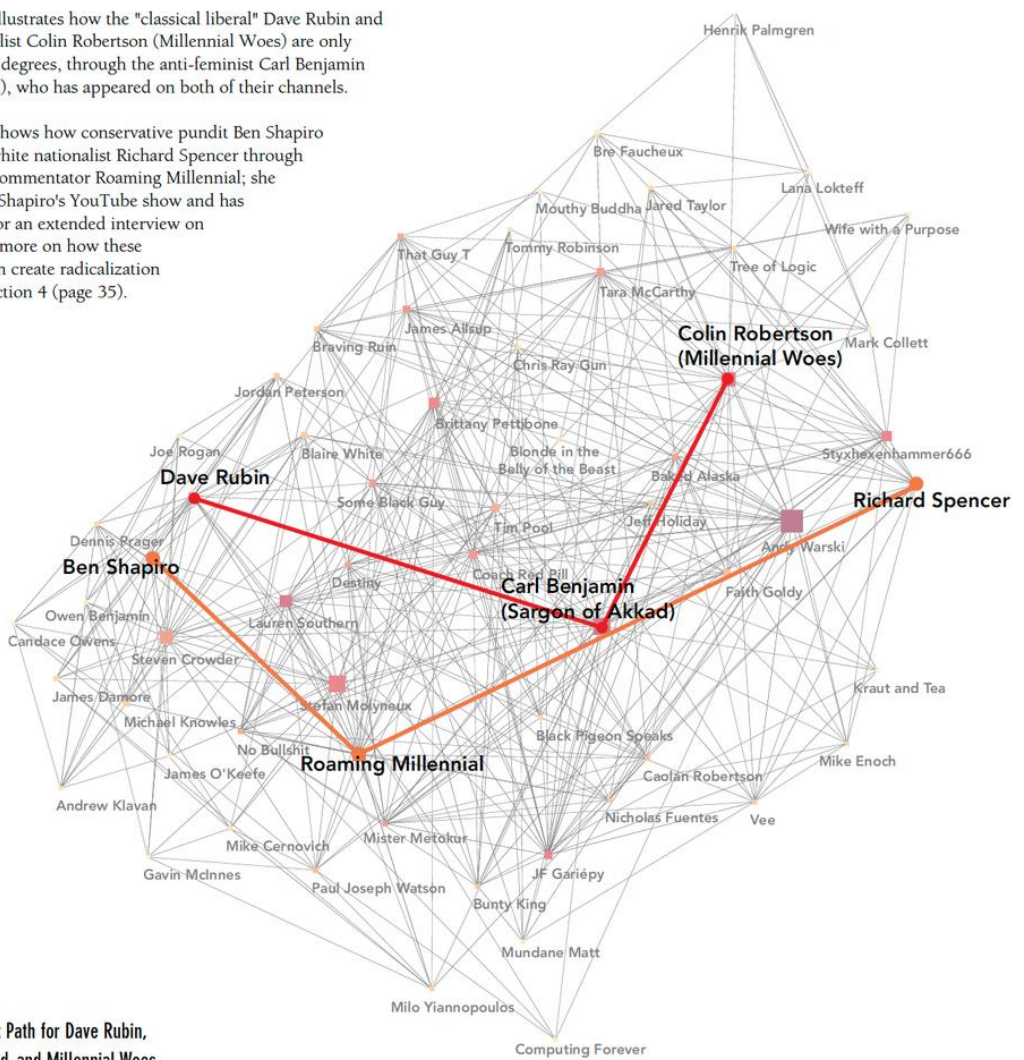
Fig. 1 (detail)

AIN NETWORK PATHS

These graphs show examples of collaborative connections between influencers of differing ideologies and how these collaborations can create pathways to radicalization.

Network Path 1 illustrates how the "classical liberal" Dave Rubin and the white nationalist Colin Robertson (Millennial Woes) are only separated by two degrees, through the anti-feminist Carl Benjamin (Sargon of Akkad), who has appeared on both of their channels.

Network Path 2 shows how conservative pundit Ben Shapiro is connected to white nationalist Richard Spencer through the vlogger and commentator Roaming Millennial; she has appeared on Shapiro's YouTube show and has hosted Spencer for an extended interview on her channel. For more on how these collaborations can create radicalization pathways, see Section 4 (page 35).



Focus: Network Path for Dave Rubin, Sargon of Akkad, and Millennial Woes

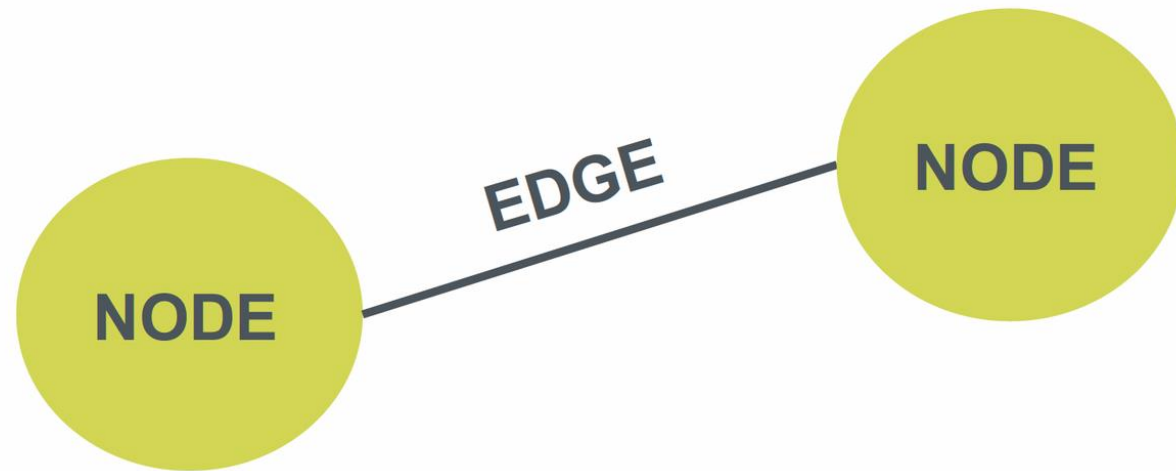
To read more about Network Path 1, see page 12.

Describing Relationships

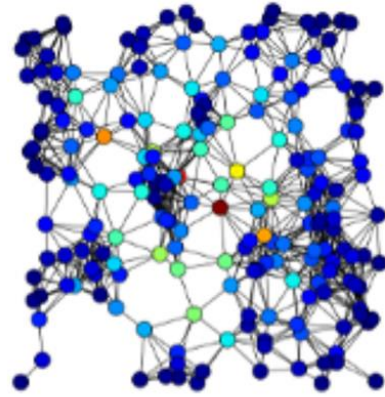
Node: the 'actor' in the network

Edge: the relationship connecting actors

Attribute: features of the node or edge

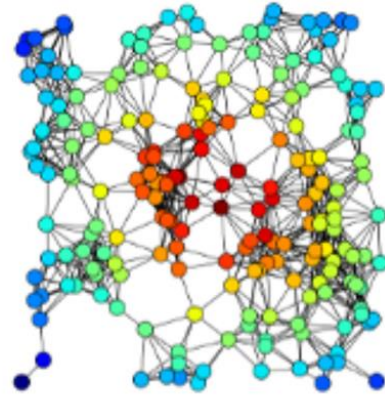


A. Betweenness



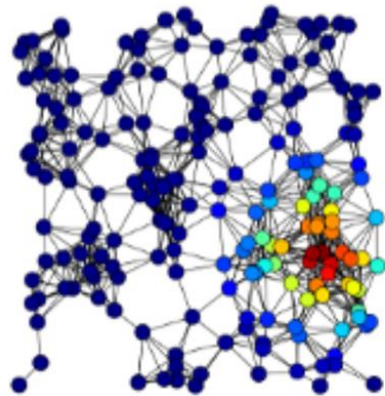
A

B. Closeness



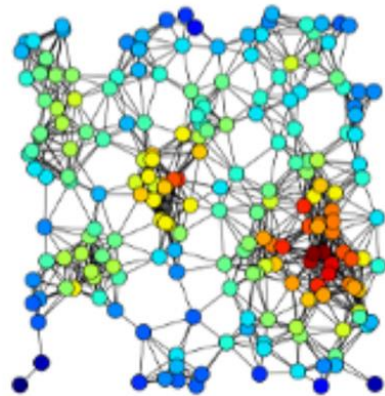
B

C. Eigenvector



C

D. Degree



D

Image: [Tapiocozzo](#) (CC 4.0 BY-SA)



Thematic Analysis Coding

I started home gardening because I wanted to grow my own fresh, organic produce and connect with nature right in my own neighborhood. In our community garden, we planted tomatoes, peppers, lettuce, herbs like basil and cilantro, and some **flowers to attract pollinators**. We ran into several issues including water shortages during **dry spells**, some plants getting damaged by pests, and coordinating watering schedules among different gardeners. Our community garden was about 50 feet by 30 feet, divided into individual plots for each family, plus a shared space for communal crops. My favorite part of community gardening was definitely the **sense of camaraderie** and shared knowledge - swapping gardening tips, exchanging **surplus produce**, and watching the garden transform from empty plots into a thriving green space where neighbors became friends over our shared love of growing things.

Name	Memo	Count
Ecosystem Restoration		2
Relationships		4
Transaction		1
Weather		3

We were at the movies on a Saturday night – watching *The Simpsons* movie with some friends. I remember sitting there feeling like I had sand in my left shoe. My toes felt gritty and weird. It was a weird sensation but I thought it would pass. My husband and I went home and when I woke up the next morning, the sensation was up to my waist. The sensation was like pins and needles – the type you get when you fall asleep on your arm and limit blood circulation. My skin was buzzing slightly. I was starting to freak out at this stage. I was wondering if I had a brain tumour. [Orientation] [Abstract]

I went to work on Monday morning and, by this stage, the sensation was quite strong and was in both my legs and my torso up to my chest. It was uncomfortable but not painful. It was alarming, though. I was working at a university and my boss was a general practitioner. I told her about my symptoms and she said it could be stress, and to have a bath and breath in and out into a paper bag. When I got home that night I tried doing what she told me, but it didn't help. I had made an appointment with my normal GP for the next day. When I saw her she thought it could be neurological and gave me a referral to see a private neurologist. She also asked me to get more bloods done. However, the earliest appointment was in 3 weeks. I was starting to feel like I was going crazy and didn't feel like I could wait that long. [Complicating action]

Anonymity By Design

"A business promises not to give up your identity unless forced to do so via court order. This is anonymity by policy. If a business doesn't have your identity, then there is nothing to divulge. This is anonymity by design."



Fabrication as Ethical Practice

"One practical method of data representation in contexts in which privacy protection is unstable is fabrication, involving creative, bricolage-style transfiguration of original data into composite accounts or representational interactions."

Markham, A. (2012). Fabrication As Ethical Practice. *Information, Communication & Society*, 15(3), 334–353. <https://doi.org/10.1080/1369118X.2011.641993>

3. How to manage and care for social media data



scds.ca
scds@mcmaster.ca

A joint initiative of McMaster's
Faculty of Humanities and
McMaster University Libraries



Research Data Management activities



Planning

- Creating a **data management plan**
- **Finding** existing data
- Obtaining systems and software for managing data



Data collection and analysis

- **De-identification** of information
- **Organizing** data efficiently
- **Documenting** data effectively
- **Storing** data securely



Knowledge Sharing

- Preparing data for **archival** and sharing
- **Publishing** data and providing access to participants and other researchers

Data Management Plans

A **Data Management Plan (DMP)** is a living document outlining how you are organizing your research data, both during the active research phase and after your project is complete.

- Identify potential issues and stumbling blocks ahead of time (ethical, legal, commercial)
 - *Terms of service! More on ethical things later.*
- Develop storage and backup strategies
- Figure out how you will organize your data before it gets out of control (including roles)
- Plan for data archival and sharing

DMP Assistant: <https://dmp-pgd.ca/>

What file formats will your data be collected in? Will these formats allow for data re-use, sharing and long-term access to the data?

The data will be in different formats, such as XLS, docx, PDF/A, sav, and MP3. These formats are easy to re-use as long as researchers are able to work with the software. In addition, the data will be available and accessible to the user.

What conventions and procedures will you use to structure, name and version-control your files to help you and others better understand how your data are organized?

At the beginning, we will add the final version for any type of the data, and there will be three other copies of the data, with one copy stored off site (external hard disk), with an access to certain staff members. In addition, the data will be grouped based on the nature of data as follows:

- The data that will be collected from the field, the file's name will include "Field data", with indication for type whether this data is questionnaire or guideline, audio, report, stories, or interview transcription.

Wael R (2017) Data Management Plan: HarassMap.
Research Ideas and Outcomes 3: e1513.

Other DMP Examples: McMaster Data Management Plan Database rdm.mcmaster.ca/dmps

Documentation and Metadata ▲

scds.ca

scds@mcmaster.ca

A joint initiative of McMaster's
Faculty of Humanities and
McMaster University Libraries



**Sherman
Centre**
for Digital Scholarship

Social Media as Research Data

- **Multimedia:** Videos, audio, images, extracted frames, speech-to-text, image analysis
- **Quantitative:** URLs, times and dates published, length, poster, views, reactions, content analysis
- **Qualitative:** emojis, observations, posts/content, tone, music
- **Networks:** likes by, follows from, appearances with other users

10videos_GV_Frames_Formulas

File Edit View Insert Format Data Tools Extensions Help

100% View only

	A	B	C	D	E	F	G	H	I	J	K	L	M	N	O
1	file_num	video_n	createTi	data/aut	data/des	hashtag	data/sta	data/sta	data/sta	data/sta	web enti	labels	best gue	top level	speech-to-te
2	002.mp4	002	22:36:06	antsants	w o a h	#deeptik	960400	94800	45400	3200000	Anime m	Fawn; G	kaguya s	pinterest	I climbed out
3	003.mp4	003	07:45:46	heybesti	it's alrea	#summe	931500	14600	31100	2500000	Violet; A	Purple; V	atmosph	pinterest	LOL. that part
4	004.mp4	004	15:26:45	famousg	mi pan	#foryou:#	785800	24000	20300	2800000	Face; Fo	Forehear	head	tiktok.cor	me pan su su
5	015.mp4	015	15:31:47	baddiesu	easy pee	#shaney	425700	1334	2482	2300000	Bowling	Photograp	game	tiktok.cor	people love th
6	016.mp4	016	01:35:01	justins_	#greensk	#greensk	412000	38700	26600	1800000	iFunny; T	Plant; P	atmosph	tiktok.cor	Mini bemedoc
7	021.mp4	021	12:25:12	skinnysu	it's jes	#beanz:#	310900	17200	16000	992300	TikTok; C	Purple; I	room;des	tiktok.cor	I climbed out
8	022.mp4	022	04:26:49	da.ketch	she didn	#fyp:#de	299900	1068	1620	1200000	Samsung	Gesture; l	hand;mo	tiktok.cor	how to be cre
9	043.mp4	043	21:54:09	wandana	wax	#alt:#alt	1600000	108500	53700	6300000	Design; C	Purple; F	light;des	tiktok.cor	brain is poison
10	083.mp4	083	21:26:42	deepfells	#deeptik	#deeptik	71300	1608	1666	175800	Product (Rectangl	orange;sky	compu	I climbed out
11	084.mp4	084	03:59:42	come_or	Я верну	#dre	70200	618	980	406800	Product (Violet; F	compute	gamejolt	this one here
12															
13															
14															
15															
16															
17															
18															
19															

"Z Spaces: TikTok Tactics and Cross-Platform Circulation of Russian Propaganda - 10videos_GV_Frames_Formulas" https://docs.google.com/spreadsheets/d/1aEX-sWQkC98CvOqXR6le5D7w8Jze2oC2n3Of7_VRqWA/edit#gid=836442687

What do I do with my data?

- Make sure you **back up** your data:
 - To find storage/backup options, check out the **McMaster Research Data Storage Finder**: <http://u.mcmaster.ca/storagefinder>
- Keep your data **organized**:
 - Use a consistent folder structure and file naming scheme – platforms; data types; collection dates
 - Tropy for images ([webinar](#)); survey software to organize coding, spreadsheet to organize URLs
 - For examples check out: <https://rdm.mcmaster.ca/organize>
- Make sure data are **documented**:
 - Take process notes - How did you approach data collection? What went wrong? Who collected the data and where is it kept?
 - Annotations for analysis code, data dictionaries, qualitative coding tags + approaches

Data Storage Finder

The tool is currently in **testing**, if you have any comments, please fill out our [feedback form](#). As McMaster data storage guidelines and practices change and providers listed here may change.

This interactive tool lists various data storage and backup providers recommended by the Research Data Management team at McMaster. If none of these providers meet your needs, contact us to set up a [consultation](#). To use the tool, just follow the following steps:

Step 1: Answer a few questions about your research data storage needs. Answering these questions will recommend specific options for data storage providers.

Step 2: Choose the data storage providers you would like to compare

Step 3: Explore the details of the providers you have chosen.

Step 1: Answer these questions to narrow down storage provider options.

Clear Answers

1. What risk level is your data? ⓘ

Low
 Medium
 High

2. What type of data storage are you looking for? ⓘ

Step 2: Select data storage providers you would like to compare **Select All**

Compute Canada Advanced research computing systems, storage and software	Compute Canada NextCloud Advanced research computing File hosting services	Dataverse Store, share, publish and discover research data	FR Fin Res
Github Distributed version control system for software code	MacDrive File Synchronization and Sharing solution	MacDrive (Encrypted Data) Encrypted File Synchronization and Sharing solution	Ma Wel tran

scds.ca
scds@mcmaster.ca

A joint initiative of McMaster's
Faculty of Humanities and
McMaster University Libraries



**Sherman
Centre**
for Digital Scholarship

Analysis of Social Media Discourse and Life Cycle Evolution

Principal Investigator(s): Wen Deng, Huazhong University of Science and Technology

Version: V1

Name	File Type	Size	Last Modified
Twitter-July-November.xlsx	application/vnd.openxmlformats-officedocument.spreadsheetml.sheet	3.2 MB	02/05/2021
Weibo-July-November.xlsx	application/vnd.openxmlformats-officedocument.spreadsheetml.sheet	5.5 MB	02/05/2021

Project Citation:

Deng, Wen. Analysis of Social Media Discourse and Life Cycle Evolution. Ann Arbor, MI: Inter-university Consortium for Political and Social Research [distributor], 2021-02-12. <https://doi.org/10.3886/E132341V1>

Project Description

Deng, Wen. Analysis of Social Media Discourse and Life Cycle Evolution. Ann Arbor, MI: Inter-university Consortium for Political and Social Research [distributor], 2021-02-12. <https://doi.org/10.3886/E132341V1>

Long-Term Data Management

- **Support reproducibility, verification, re-use, and avoid re-collection of data.**
- *Read the terms and conditions closely.*
- **Archive:** Tidy files and folders, delete miscellaneous files, keep de-identified files, documentation and process docs.
- **Consider depositing:** Qualitative analysis, metadata (titles, usernames, urls), methodologies, code used to analyze data.
- **Restrict access to:** De-identified data, raw files.



(A Few) Ethical Considerations When Collecting Social Media Data

Hanna Barakat & Cambridge Diversity Fund, “Data Lab Dialogue,” <https://betterimagesofai.org> - CC BY 4.0

Is social media data considered ‘public domain’?

“REB review is required in all human research using (Social Media) platforms, with one exception – **when participants have no reasonable expectation of privacy, and the information is in the public domain**” (TCPS 2, 2022). More at ethics.gc.ca/eng/reb-cer_social-sociaux.html.

- Just because an account is public, doesn’t mean users are consenting to **researcher use**.
- How might users conceptualize ‘public’?

What are the rediscovery risks of social media data?

Even after anonymization, social media data remains highly re-identifiable: platform search tools, timestamps, and keyword matching can trace paraphrased or de-identified posts back to their original users.

- Privacy as Contextual Integrity (Helen Nissenbaum) – when are information flows appropriate?



(A Few) Ethical Considerations When Collecting Social Media Data

Hanna Barakat & Cambridge Diversity Fund, "Data Lab Dialogue," <https://betterimagesofai.org> - CC BY 4.0

How does an intersectional ethics framework shift the researcher's responsibility?

- Harm unevenly distributed across different groups within digital spaces.
- Social media data can contain disclosures from groups whose risks are compounded by gender, sexuality, race, migration status, class, disability, or occupation.
- Consider differential vulnerabilities when thinking about participant risk.

How do platforms influence the research we can conduct?

- Platforms play a significant role in research, actively shaping the conditions under which data is produced, accessed, and interpreted.
- Researcher ↔ Participant ↔ Platform

Discussion and Reflections

- Are you currently working with social media data in your research?
 - What platforms, methods, or strategies have you found useful?
 - What challenges have you encountered when collecting, organizing, or analyzing it?
- Are there any tools, workflows, or ethical strategies you're taking away from today that you plan to try in your own research?
- Do you have any questions about the content we covered today?
- And anything else – we want to hear from you!

Jimmie Quick, "Cat stares at itself on computer monitor," Flickr January 18, 2009, <https://www.flickr.com/photos/jimmiehomeschoolmom/3050445842/> - CC-BY-2.0

SCDS Links

Send SCDS an Email:

scds@mcmaster.ca

Register for a Workshop:

<https://u.mcmaster.ca/scds-workshops>

Subscribe to our Newsletter:

<https://u.mcmaster.ca/sign-up>

Schedule a Consultation:

<https://libcal.mcmaster.ca/appointments>

