

Towards a Linked Infrastructure for Networked Cultural Scholarship

Susan Brown and Kim Martin
University of Guelph

CSDH 2020

License: CC-BY-NC





Roadmap

- Why a *linked* infrastructure?
- The Semantic Web and Linked Open Data
- Towards a linked data infrastructure

A network graph visualization on a black background. The graph consists of numerous nodes, represented by small colored circles in shades of blue, green, yellow, orange, and purple. These nodes are interconnected by a dense web of thin, light-colored lines representing edges. The overall structure is complex and interconnected, with some nodes having higher degrees of connectivity than others. The text "Why a *linked* infrastructure?" is overlaid on the graph in a white, sans-serif font.

Why a *linked* infrastructure?

(Digital) humanities data

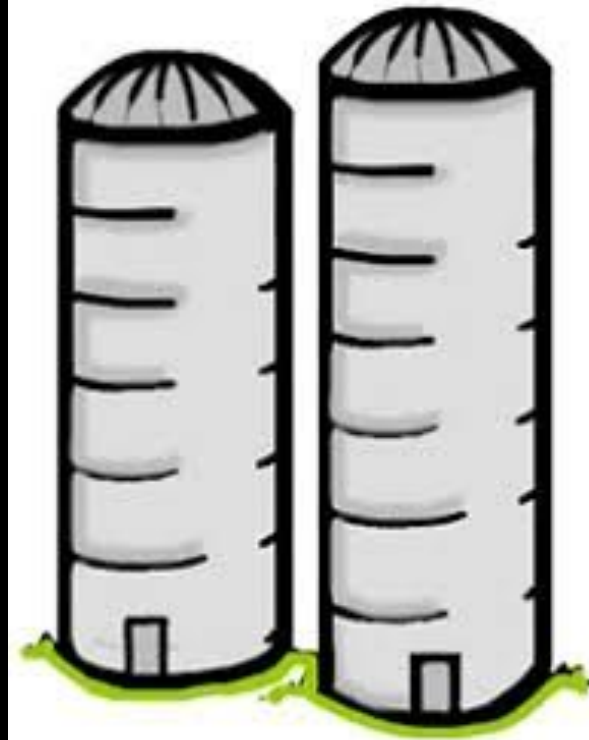
- rich, researched, meticulous
 - reliable sources of knowledge
 - with high potential for reuse
- and*
- hard to find
 - with heterogeneous formats
 - “siloes” *for both good & bad reasons*

Therefore:

hard to share data or work at scale

“What kinds of humanistic phenomena appear only at scale?”

Alan Liu
“The State of the Digital Humanities: A Report and Critique” (2012)



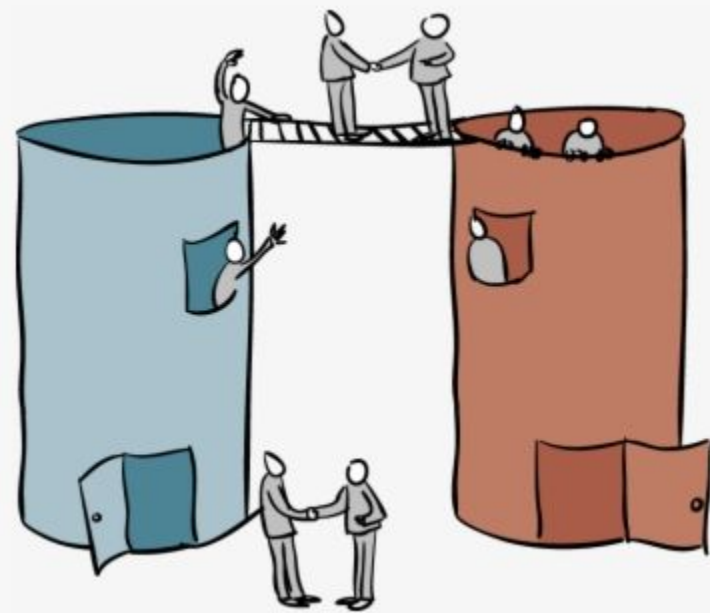
Scaling up links: networks and graphs

To overcome silos, we can link them.

Put together lots of links and you have a *network* (or a Web) ... or a *graph*.

What's the difference?

Partly just nomenclature, but let's explore some specifics.



This structure is very simple, consisting of a number of dots that are called vertices in graph terminology and called nodes in network terminology. ... a graph is made up of vertices connected by edges, while a network is made up of nodes connected by links.

Jesse Johnson
“[Graphs and networks](#)” (2013)

A complex network diagram with numerous nodes and edges. The nodes are represented by small colored circles in various colors including blue, green, yellow, orange, pink, and purple. The edges are thin, light-colored lines connecting the nodes, creating a dense web of connections. The overall structure is horizontal and spans the width of the image.

A network: the Web

the Web

- Uses http:// protocol:
 - We request HTML documents from web servers
 - Our web browsers (Firefox, Chrome, Explorer, Safari) then format HTML files so we can read them
- 5.7 billion pages (as of May 6, 2020)

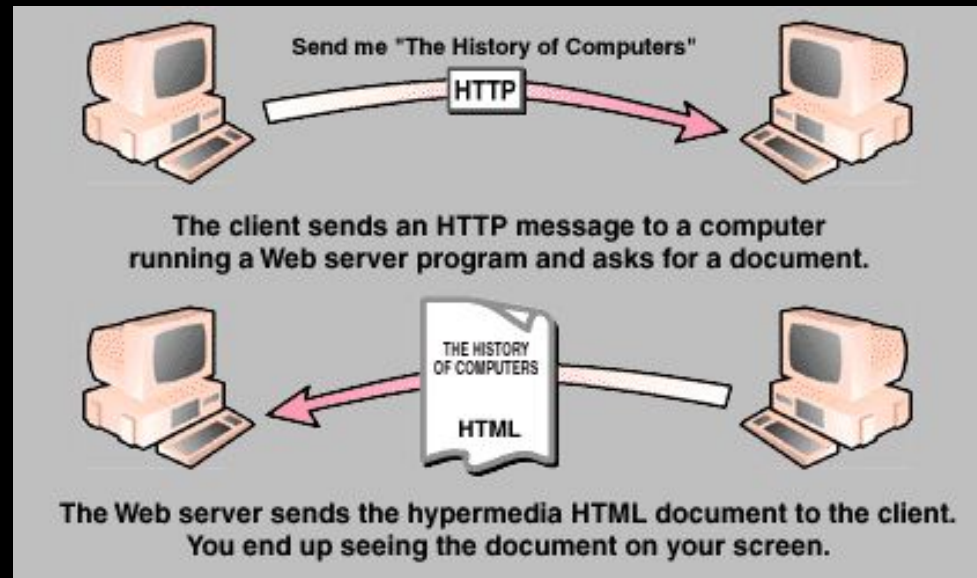


Image from Kevin Cove ["How does the Web work?"](#) (1994)

HTML = Hypertext Markup Language

the Web

- HTML (Hypertext Markup Language) documents
 - Format for structure not content
 - No “meaning”
- Hyperlinks
 - Addresses of files on machines
 - No “meaning” built into the links
- humans can read the contents of Web documents
- machines cannot (except through extra programs/steps)
- links between scholarly resources usually created manually or by error-prone algorithms (cf. ResearchGate author misattributions)



Emphasizing relationships: graphs

Facebook Social Graph

“the largest social network ever
analyzed”

Johan Ugander et al
"The anatomy of the facebook social graph" (2011)

Not just links between people but
between people and photos, Spotify
choices, events, videos, websites.



Google Knowledge Graph

Produces structured, organized information to complement and contextualize the results of the string-based search algorithm

- Blackbox knowledge graph
- No provenance information
- Uses (and appropriates user-created) linked data

“The Freebase knowledge base was a major Semantic Web and linked data technology that was acquired by Google in 2010 to support the Google Knowledge Graph, the backend for Google search results that include structured answers to queries instead of a series of links to external resources.”

Niel Chah

"OK Google, What Is Your Ontology? Or: Exploring Freebase Classification to Understand Google's Knowledge Graph" (2018)

The screenshot shows a Google search for 'toni morrison'. The search results include a Knowledge Graph panel on the right for 'Toni Morrison', an American novelist. The panel displays a grid of images, a brief biography, birth and death dates, awards, books, quotes, and a 'People also search for' section with images of other authors like Maya Angelou and Alice Walker. On the left, search results from various sources like Wikipedia, NYTimes, and Britannica are visible. A red arrow points from the text above to the Knowledge Graph panel, and another red arrow points from the text above to the 'People also ask' section.

Humanities data

➤ typically unlinked

Linked data reflecting humanities domains includes [Linked Jazz](#) (Christine Patuelli). Institutions in the Galleries, Libraries, Archives and Museums sector are embracing LOD, e.g. [Europeana](#), and [Linked Art](#). And there are explorations of alternative forms and methods of creating LOD, such as [HuNI](#) (Deb Verhoeven). However, linked data has had little impact on how humanities scholars, including many in the digital humanities, pursue their research or manage the resources they gather or create for it.

Web-scale linked data

➤ proprietary

➤ blackboxed

➤ inadequate for humanities scholarship

The humanities need Linked Open Data



A network graph visualization on a black background. The graph consists of numerous nodes, represented by small colored circles in shades of blue, green, yellow, orange, and pink. These nodes are interconnected by thin, light-colored lines representing edges. The connections form a complex, web-like structure with several clusters and long-range links. The overall appearance is that of a large, interconnected network, possibly representing data relationships or a semantic web.

The Semantic Web & Linked Open Data

What is the Semantic Web?

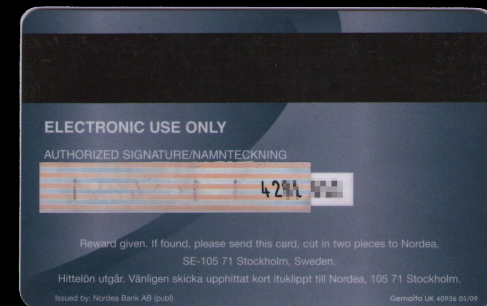
- vision of the Web (Tim Berners-Lee et al)
 - web links have meaning (semantic content)
 - machines can process links / relationships

Web becomes both

human-readable

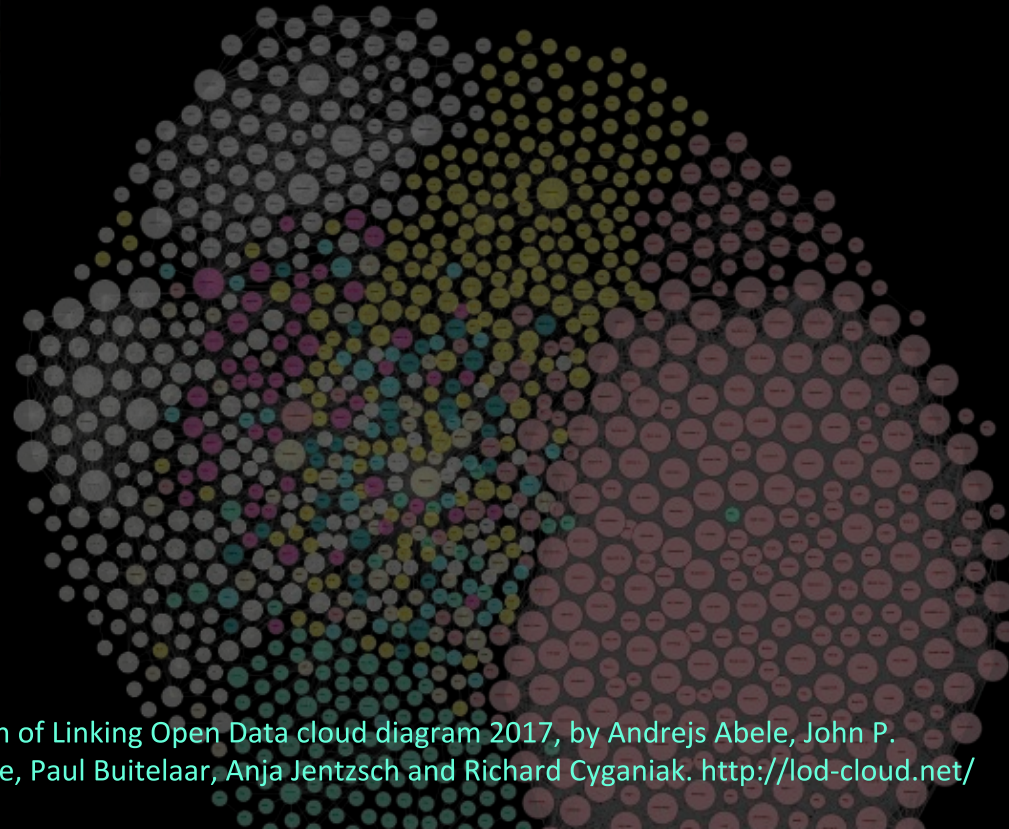
&


machine readable



Linked Open Data

- extension of Web standards for data format and exchange
- from “strings” to “things”:
making web (links) meaningful (“semantic”)
- promotes
 - sharing
 - interoperability
 - re-use
 - discoverability
 - inferencing
 - Serendipity





Linked Open Data (LOD): “Semantic web done right”

Tim Berners-Lee
“[Linked Open Data](#)” (2008)

5-star Linked Open Data

(Berners-Lee
“[Linked Data - Design Issues](#)” (2006)

★ Available on the web (whatever format) *but with an open licence, to be Open Data*

★★ Available as machine-readable structured data (e.g. excel instead of image scan of a table)

★★★ as (2) plus non-proprietary format (e.g. CSV instead of excel)

★★★★ All the above plus, Use open standards from W3C (RDF and SPARQL) to identify things, so that people can point at your stuff

★★★★★ All the above, plus: Link your data to other people’s data to provide context

Humanities LOD needs infrastructure

- 'Heavy' technology stack
- requires platforms for storing, searching, retrieving LOD, plus:
 - automated linked data generation
 - entity identification
 - entity disambiguation
 - vetting of automated results
 - relationship extraction
 - ontologies for humanities
 - domain-specific linked open vocabularies
 - discovery and exploration tools
 - enhancement and augmentation tools
 - people with expertise

... it is politically and ethically crucial to recognize the vital role of infrastructure in the 'built moral environment.'
Seemingly technical issues like how to name things and how to store data in fact constitute much of human interaction and much of what we come to know as natural.

Geoffrey Bowker and Susan Leigh Star
Sorting Things Out (1999)

About 98,400 results (0.24 seconds)

Joy Kogawa - Wikipedia, the free encyclopedia
en.wikipedia.org/wiki/Joy_Kogawa -
Joy Nozomi Kogawa, CM, OBC (born June 6, 1935) is a Canadian poet and ... Born Joy Nozomi Nakayama in Vancouver, British Columbia, she was sent with ...
Life - Recognition - Campaign to save Kogawa House - Bibliography

Joy Kogawa - About Joy
www.joykogawa.ca/ -
Joy Kogawa was born in Vancouver in 1935 to Japanese-Canadian parents. During WWII, Joy and her family were forced to move to Slocan, British Columbia, ...

Welcome to Historic Joy Kogawa House
www.kogawahouse.com/ -
The writer-in-residence program at Historic Joy Kogawa House welcomes writers and readers to join our community of people who are passionate about books ...

Obasan: Joy Kogawa: 9780140067774: Books - Amazon.ca
www.amazon.ca/ ... Literature & Fiction > Canadian > Asian Canadian -
A powerful and passionate novel, Obasan tells, through the eyes of a child, the moving story of Japanese Canadians during the Second World War. Naomi is a ...

Joy Kogawa - Brock University
www.brocku.ca/canadianwomenpoets/Kogawa.htm -
WORKS. POETRY. A Garden of Anchors: Selected Poems. Oakville, ON: Mosaic, 2003. A Song of Lilith. Vancouver: Polestar, 2000. Woman in the Woods.

Canadian Poetry Online | University of Toronto Libraries | Joy...
www.library.utoronto.ca/canpoetry/kogawa/ -
Joy Kogawa is a poet and playwright, an author in many genres. She was born June 6, 1935 in Vancouver, British Columbia, Canada. She has one son, Gordon; ...


Where there's a Wall - University of Toronto Libraries
www.library.utoronto.ca/canpoetry/kogawa/poem4.htm -
Where there's a Wall. Joy Kogawa. where there's a wall there's a way around, over, or through there's a gate maybe a ladder a door a sentinel who sometimes ...

Joy Kogawa - The Canadian Encyclopedia
www.thecanadianencyclopedia.com/articles/joy-kogawa -
Mar 7, 2014 - Joy Nozomi Kogawa, née Nakayama, poet, novelist (b at Vancouver 6 Jun 1935). Joy Kogawa spent her early childhood in VANCOUVER.



Joy Kogawa
Poet
Joy Nozomi Kogawa, CM, OBC is a Canadian poet and novelist of Japanese descent. Wikipedia
Born: June 6, 1935 (age 78), Vancouver
Awards: American Book Awards
Education: University of Toronto, University of Alberta

Books

 Obasan 1981	 The rain ascends 1995	 Naomi's road 1986	 Itsuka 1962	 Naomi's Tree 2008
---	---	---	---	---

People also search for

 Michael Ondaatje	 Hiromi Goto	 Roy Miki	 Evelyn Lau	 Wayson Choy
---	--	---	---	--

Feedback

A network graph visualization on a black background. It features numerous small, multi-colored nodes (blue, green, yellow, orange, pink, purple) connected by thin, light-colored lines. The nodes are arranged in a complex, interconnected pattern, with some nodes having multiple connections, suggesting a highly linked network structure.

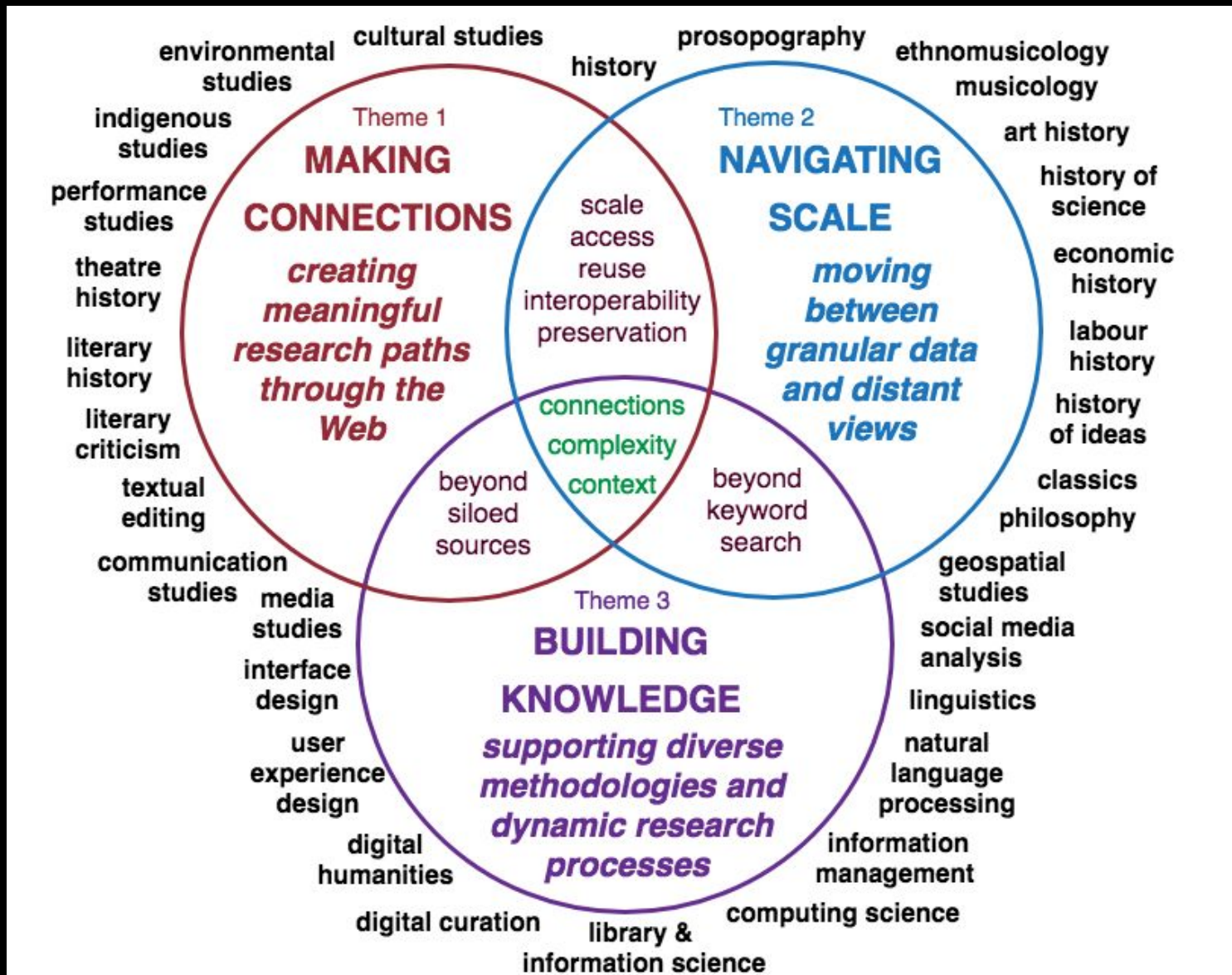
LINCS: towards a linked data infrastructure



What is the Linked Infrastructure for Networked Cultural Scholarship?

- CFI Cyberinfrastructure project
- 3 years (2020-2023)
- 50+ researchers and collaborators in the humanities, library and information science, computer science
- 14 university and other partners

LINCS will convert, enhance, mobilize, link, and make accessible existing datasets, and thereby enable research on the vast and heterogeneous body of data relevant to human culture, history, and knowledge, both in Canada and throughout the world.



LINCS research themes with domains represented by LINCS datasets and activities

Making Connections (led by Janelle Jenstad) | Navigating Scale (led by Stacy Allison-Cassin) | Building Knowledge (led by Jon Bath)

LINCS will connect and connect to large sets of digitized Cdn content

Over 10 million records; 30 million triples; 300 TB

Canadiana Héritage Collections; Libraries & Archives Canada; Canadian Internet Archive

Canadian

LINCS data will be linked to the global graph

10.5 billion semantic web assertions ("triples")

Global general knowledge base; Wikidata; DBpedia; WorldCat; VIAF

Linked

Over 60 million pages; 54,000 texts

Cultural

Project Gutenberg; Scholarly Archives; Advanced Research Consortium; HathiTrust;

LOD will be extracted from large digitized datasets

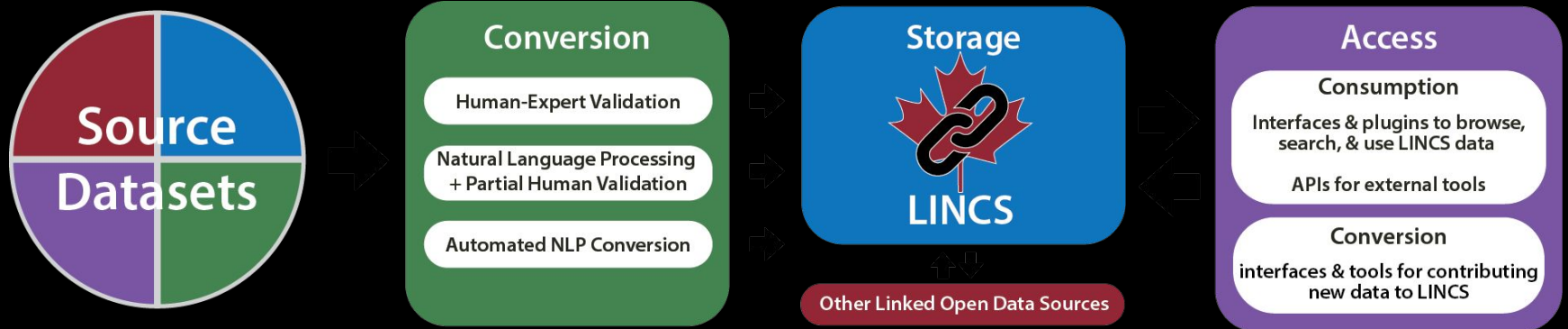
Research

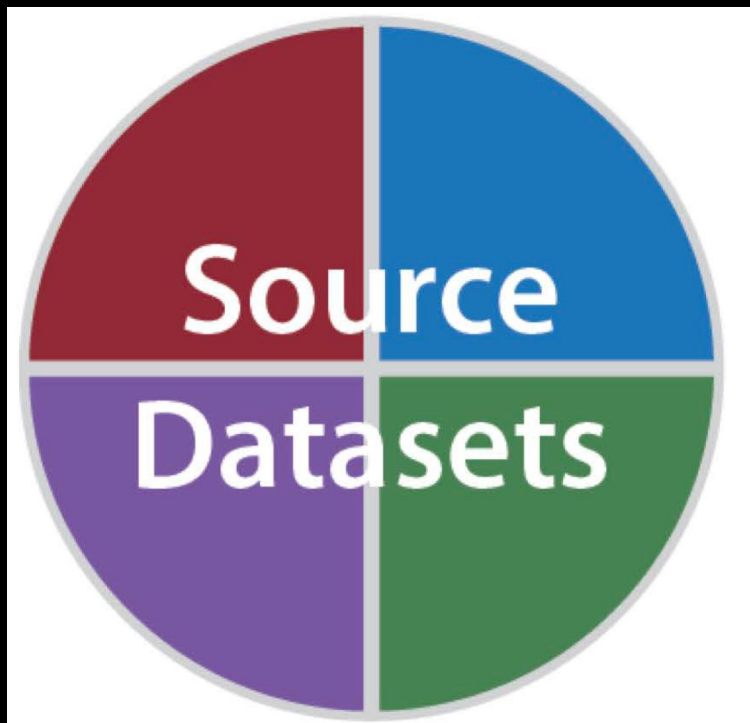
history; politics, literature, music; travel; Canadian francophonie; first nations; modernism; gender; sexuality + user-uploaded data

Over 500,000 records; 1.2 million triples

LINCS
Databases
Researcher datasets and needs are at the heart of LINCS

LINCS in a nutshell





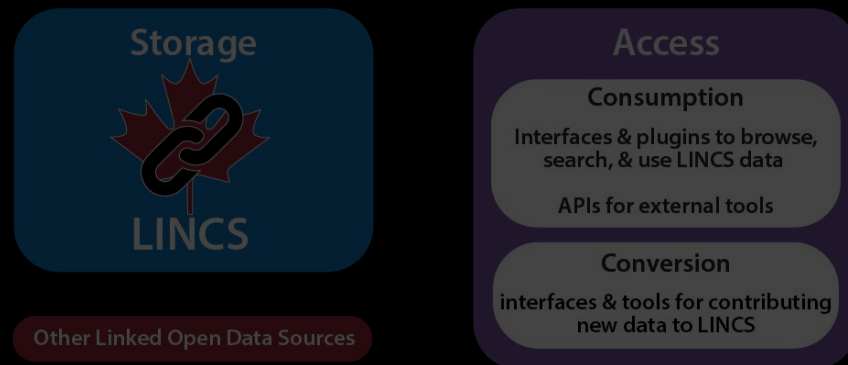
metadata

(titles, authors, dates of publication)

or

data

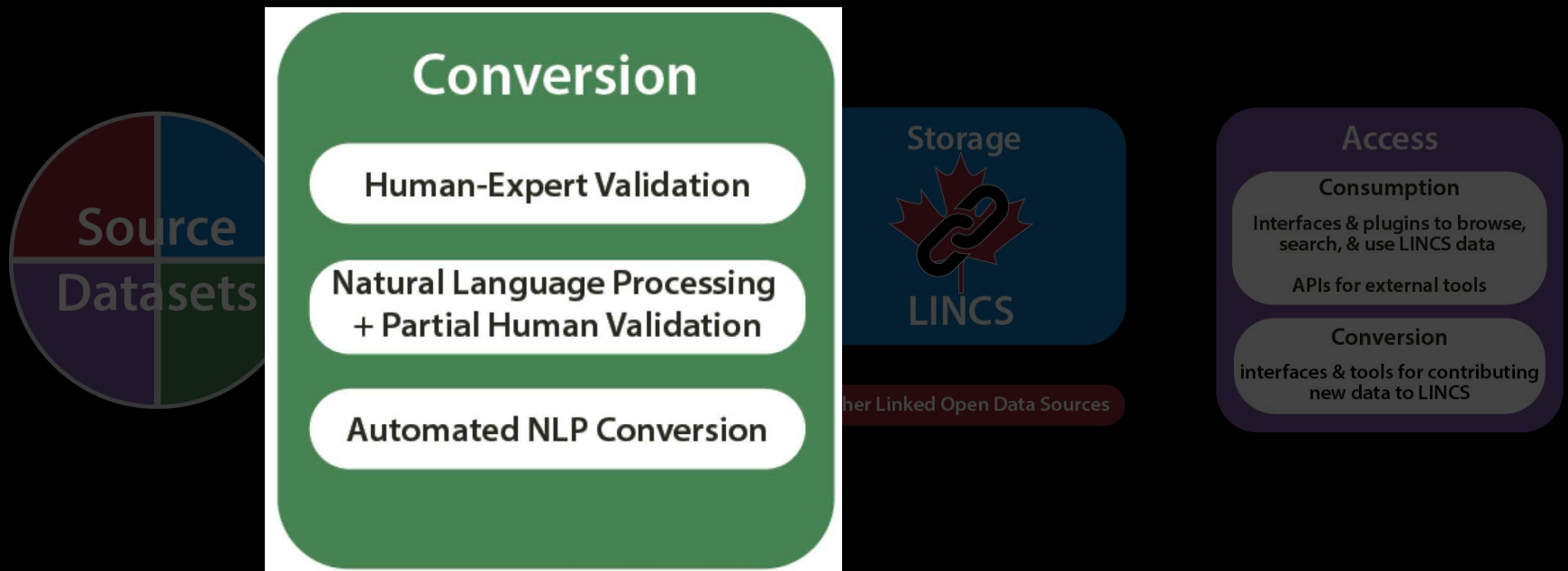
(full text of articles)



Formats:

- Table, database export
- Free text
- XML (Extensible Markup Language) - already has semantic structure

Source needs a stable URL

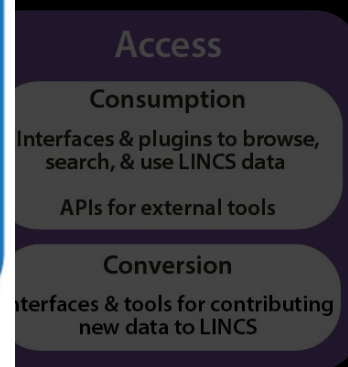
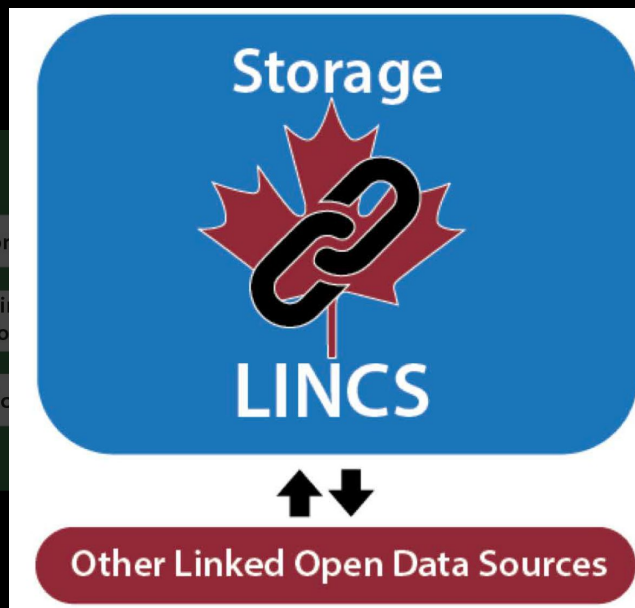


Conversion:

- identifying entities (people, places, organizations, titles)
- identifying relationships

Accuracy:

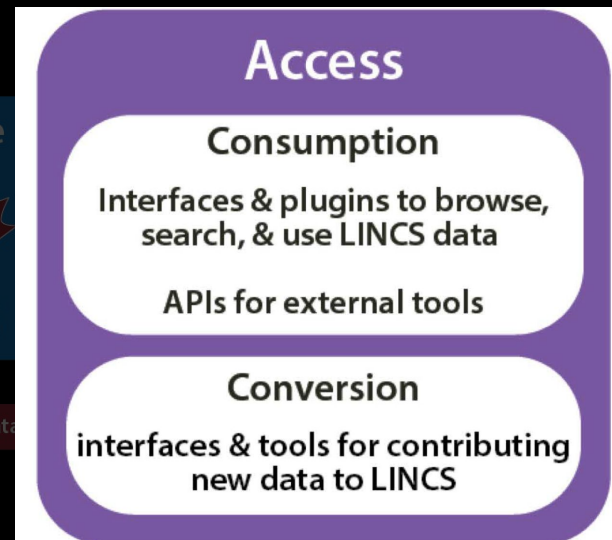
- machine results are imperfect
- human (researcher) refinement required
 - opens up dialogue
 - resource-intensive



“Triple store”

- LODatabase
- data aggregated but stored as separate collections (“graphs”)

- connected to other related LODsets (e.g. HumaNum)
- interoperable (shared ontologies / vocabularies)



User interface:

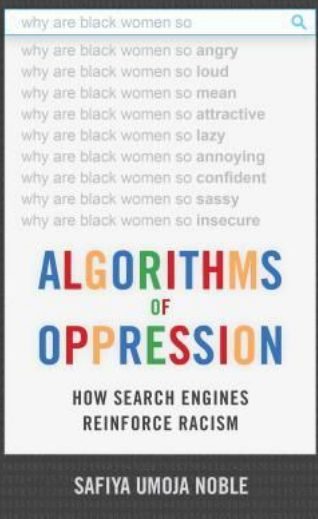
- search, browse
- contexts widget
(cf. Google knowledge graph)
- send results to other tools
([HuViz](#); [Voyant/Spyral Notebooks](#))
or download

Conversion tools

- User-friendly on-ramps
- Documented methods, tools, Jupyter notebooks for intermediate users
- Open and documented code repositories for advanced users/developers

A complex network graph visualization on a black background. The graph consists of numerous nodes, represented by small colored circles in shades of blue, green, yellow, orange, and pink. These nodes are interconnected by a dense web of thin, light-colored lines representing edges. The overall structure is intricate and somewhat chaotic, with many nodes having multiple connections. The word "Challenges" is centered over the graph in a large, white, sans-serif font.

Challenges



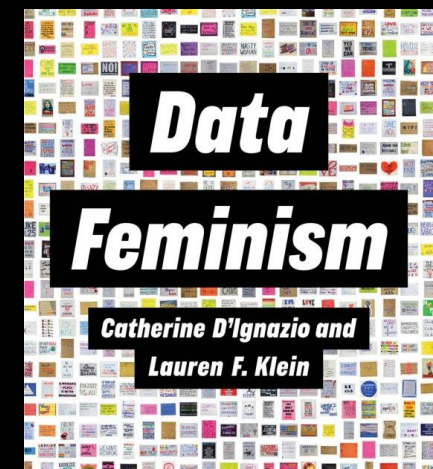
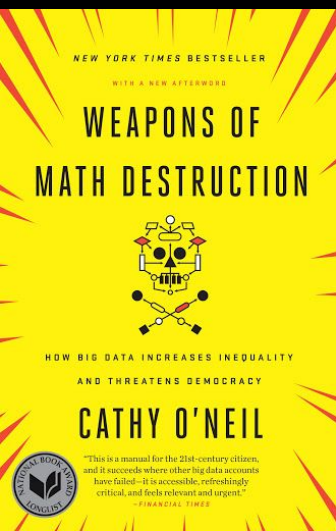
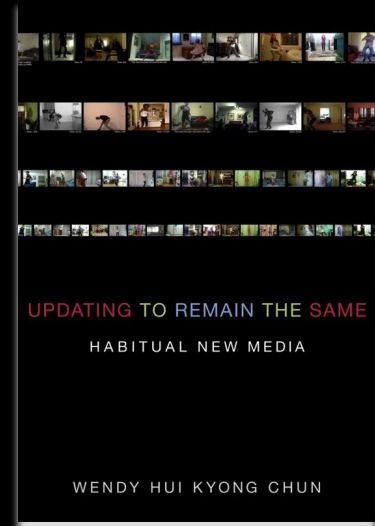
General challenges

“Even people with deep convictions about Linked Data confess that the systemic obstacles are daunting. We will require a great deal of cooperation and collaboration across institutional and national boundaries if Linked Data is ever to really achieve its potential.”

Lisa Goddard and Gillian Byrne, “[Linked Data Tools: Semantic Web for the Masses](#)” (2010)

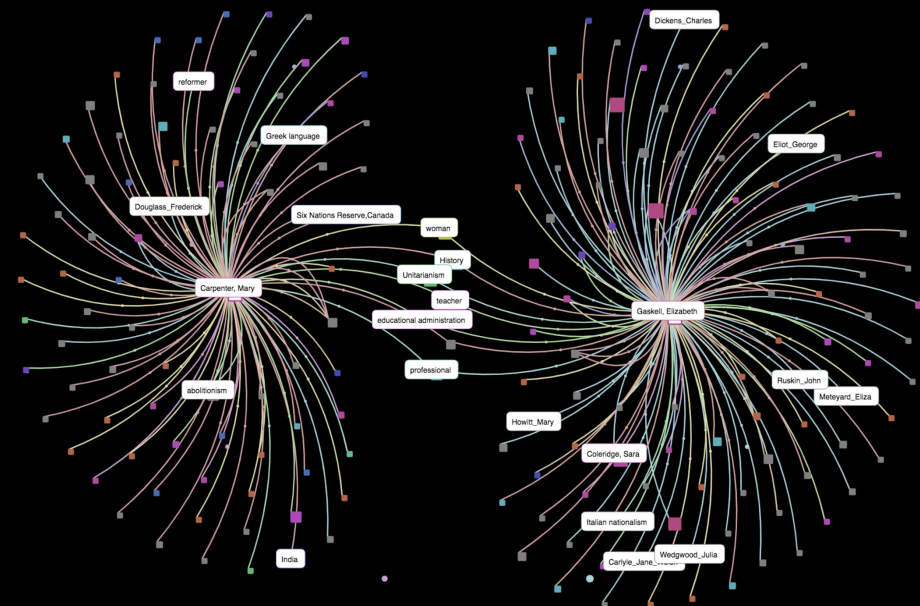
A LINC'S cornerstone must be recognition that engaging with LOD is all the harder for individuals and organizations representing marginalized communities—including Indigenous, Black and other POC, LGBTQ+, and women scholars—whose energies are necessarily absorbed and spread thin by other priorities, and who have a justified suspicion of big and algorithmic data.

([Noble 2018](#); [Chun 2016](#); [O’Neil 2018](#); [D’Ignazio and Klein 2020](#))



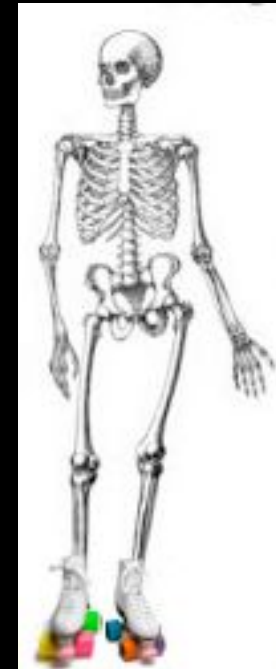
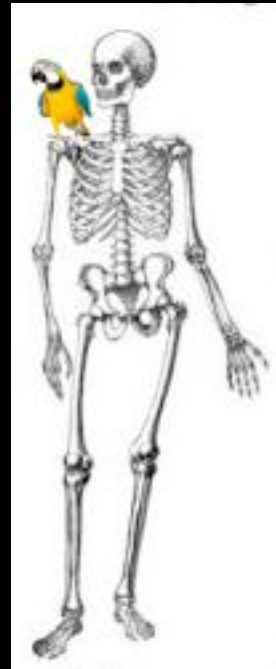
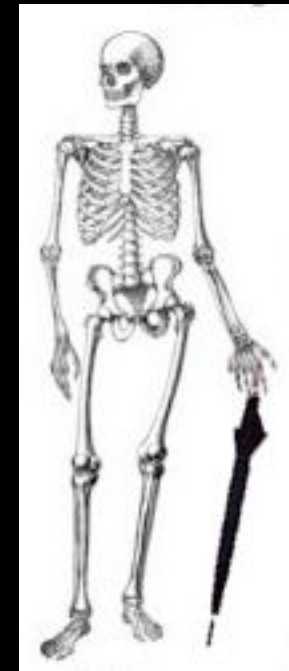
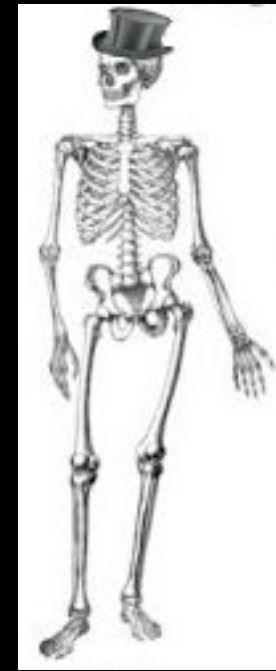
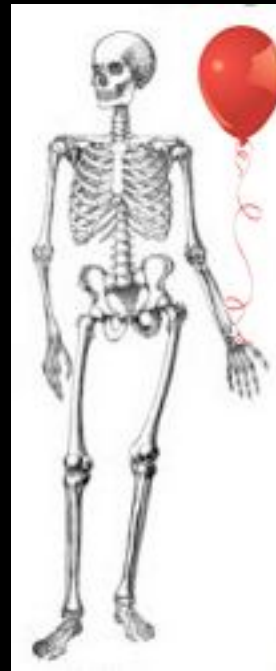
Challenges for humanities LOD infrastructure

- translating humanities knowledge into LOD
 - structuring disparate resources from many domains
- nuanced LOD representations
 - ontologies and contexts
- accessibility
 - LOD literacy
 - tools for Linked Open Usable Data ([Sanderson, 2018](#))
- sustainability



LOD for difference and diversity

- differences of provenance, representations of diverse cultural producers
- distinctions of particularity, specificity, and locality of meaning
- deviations from the norm: the marginalization and anomalousness associated with categories such as race, gender, sexuality, or nation
- key to understanding literature, culture, and ideas of the past and the present



“... it is precisely the ability to treat infrastructure not as a foundation, a given, but instead as a tactical medium that opens the possibility of critical infrastructure studies as a mode of cultural studies.”

Liu, Alan. "[Drafts for Against the Cultural Singularity](#)" (2016)

Thanks for your attention!

- Visit the LINCS website, and subscribe to the LINCS newsletter: <https://lincsproject.ca/>
- Follow us on Twitter [@lincsproject](https://twitter.com/lincsproject)
- Get involved! We welcome new collaborators and organizational partners.

✉ lincsproject@gmail.com

Susan Brown
@susanirenebrown
Project Lead

Kim Martin
@antimony27
Research Board Chair



LINCS News



Photo by [Markus Spiske](#) on [Unsplash](#)

Welcome to LINCS!

Welcome to the newsletter for the Linked Infrastructure for Networked Cultural Scholarship (LINCS). You're receiving this because you're in some way connected to or have expressed an interest in the LINCS project.

If you want to keep receiving this, then you need to sign up for our mailing list.

[Sign up for future newsletters!](#)

Institutional Partners



The University of Alberta



The University of Guelph



The University of Ottawa



The University of Saskatchewan



The University of Toronto



The University of Victoria



York University



Thanks to our partners and collaborators



Thanks to CFI and provincial funders

Works Cited

Abele, Andrejs, John P. McCrae, Paul Buitelaar, Anja Jentsch and Richard Cyganiak. Linking Open Data cloud diagram 2017. <http://lod-cloud.net/>

Berners-Lee, Tim. "Linked Open Data." W3C Consortium. 2008.
[https://www.w3.org/2008/Talks/0617-lod-tbl/#\(1\)](https://www.w3.org/2008/Talks/0617-lod-tbl/#(1))

---. "Linked Data." *Design Issues*. W3C Consortium. 2006.
<https://www.w3.org/DesignIssues/LinkedData.html>

Bowker, Geoffrey C., and Susan Leigh Star. *Sorting Things Out*. MIT Press, 1999.

Chah, Niel. "OK Google, What Is Your Ontology? Or: Exploring Freebase Classification to Understand Google's Knowledge Graph." *arXiv preprint*. 2018. arXiv:1805.03885.

Chun, Wendy Hui Kyong. *Updating to Remain the Same: Habitual New Media*. MIT press, 2016.

Cove, Kevin. "How does the Web work?" *Guide to Cyberspace*. 1994.
<https://www.kevcom.com/words/guide/guide.10.html>

D'Ignazio, Catherine, and Lauren F. Klein. *Data Feminism*. MIT Press, 2020.

Europeana. Dir. Harry Verwayen. <https://pro.europeana.eu/>

Humanities Networked Infrastructure. Dir. Deb Verhoeven. <https://huni.net.au/>

Jesse Johnson "Graphs and networks," *The Shape of Data*. 2013.

<https://shapeofdata.wordpress.com/2013/08/13/graphs-and-networks/>

Linked Art. <https://linked.art/>

Linked Infrastructure for Networked Cultural Scholarship. Dir. Susan Brown.

<https://lincsproject.ca/>

Linked Jazz. Dir. Christine Patuelli. <https://linkedjazz.org/>

Liu, Alan. "Drafts for Against the Cultural Singularity." (2016).

<https://hcommons.org/deposits/item/mia:699/>

---. "The state of the digital humanities: A report and a critique." *Arts and Humanities in Higher Education* 11.1-2 (2012): 8-41.

Noble, Safiya Umoja. *Algorithms of Oppression: How Search Engines Reinforce Racism*. NYU Press, 2018

O'Neil, Cathy. *Weapons of Math Destruction: How Big Data Increases Inequality and Threatens Democracy*. Broadway Books, 2016.

Ugander, Johan, et al. "The anatomy of the facebook social graph." 2011. *arXiv preprint arXiv:1111.4503*