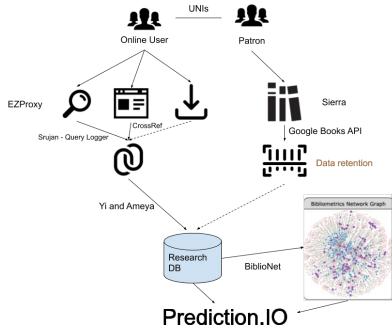
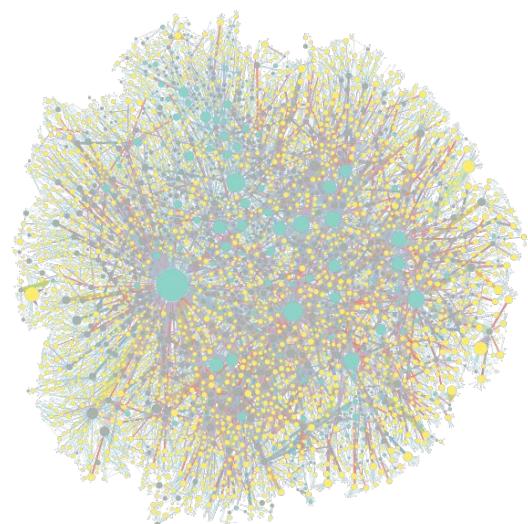


Networks, Library Data and Recommendations A Data Science Journey



Henry Williams - 07/24/2019



The Problem: Recommending Educational Content

- In Yi's talk - he discussed the challenges of building recommendation systems for education:
 - No perfect system
 - Need a lot of data - NLT did not have enough
 - Want to take user behavior into account
- In my talk:
 - How do we go about solving some of those problems?
 - What does (real-life) data science problem solving look like?
 - How can we understand and draw meaningful insights from our data?
 - What is the path forward?

Traditional Recommendation Systems

- Use mathematical methods to determine correlation between users and content
- Are content-agnostic (they don't care what they are recommending)
- Maximize whatever parameter you give them (usually user-rating)

SVD Algorithm

- Singular value decomposition (SVD) is a linear algebra technique
- **Basically: it takes a matrix with users across the top and items along the sides, with each cell containing some kind of rating, and decomposes it into smaller matrices which represent the data based on latent (hidden) factors**

$$\begin{array}{cccc} \text{M} & = & \text{U} & \Sigma & \text{V}^* \\ m \times n & & m \times m & m \times n & n \times n \\ \\ \text{U} & & \text{U}^* & = & \mathbf{I}_m \\ \\ \text{V} & & \text{V}^* & = & \mathbf{I}_n \end{array}$$

The diagram illustrates the Singular Value Decomposition (SVD) of a matrix M . It shows a large grey rectangular matrix M at the top left, labeled $m \times n$. To its right is the equation $M = U \Sigma V^*$, where U is an $m \times m$ matrix with vertical stripes of different colors (blue, green, yellow, orange), Σ is an $m \times n$ diagonal matrix with colored squares (orange, yellow, green, blue) on the diagonal, and V^* is an $n \times n$ matrix with horizontal stripes of different colors (purple, pink, blue, green). Below this, the equation $U U^* = I_m$ is shown, with U and U^* being the same as above, and I_m being an identity matrix with ones on the diagonal. Finally, the equation $V V^* = I_n$ is shown, with V and V^* being the same as above, and I_n being an identity matrix with ones on the diagonal.

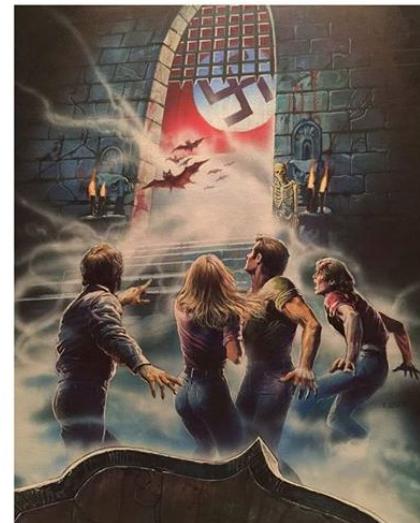
But...

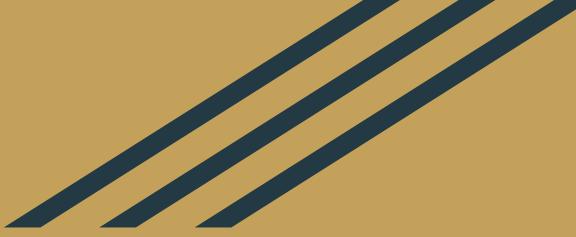
These methods don't care who your users are or what you're
recommending to them

Pitfalls of traditional methods

- Methods like collaborative filtering do not take into account any metadata about what is being recommended
- Most recommendation systems were developed in the tech industry, with the goal of:
 - Maximizing use-time
 - Making users happy
 - Increasing ad conversions
- This leads to moral hazards, i.e. the YouTube algorithm recommending increasingly extreme content
- Data science should be about people and their needs - tools like Deep Learning obfuscate what the computer is doing
- **For education, we want to recommend content that will actually help users *learn*: need to know what we are recommending and why.**

**YOUTUBE RECOMMENDATIONS
AFTER WATCHING BASICALLY
ANYTHING**





AUDIENCE PARTICIPATION SEGMENT:

What would you try to improve on these systems?



So, we need more data

Landscape of Library Data



Online User



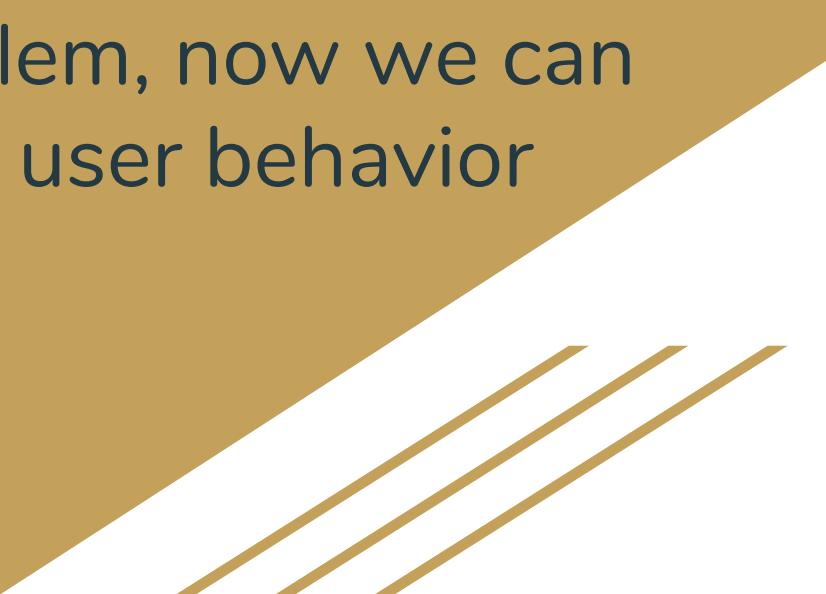
Patron

EZProxy



Sierra

- EZProxy is the library's proxy server
 - It gives users access to copyrighted content that Columbia has institutional rights to
 - When you see a "Columbia e-link" it comes from EZProxy
- Library Sierra holds the records for physical items



Perfect.
Solved that problem, now we can
start studying user behavior

Nope.

EZProxy Server Data

ID	Datetime	IP	Session	Web	Address
2	2019-01-02 00:01:06	141.193.49.114	dN4qZby9yWKXVv6	ebookcentral.proquest.com	http://ebookcentral.proquest.com/lib/teacherscollege-ebooks/detail.action?docID=1886731
3	2019-01-02 00:07:20	141.193.49.114	dN4qZby9yWKXVv6	ebookcentral.proquest.com	http://ebookcentral.proquest.com/lib/teacherscollege-ebooks/detail.action?docID=1886731
4	2019-01-02 00:35:26	69.115.20.150	IGFYstQGbjwG5my	search.proquest.com	http://search.proquest.com/pqdtfglobal?accountid=14258
5	2019-01-02 00:51:33	68.202.170.182	NnU1d3YpCxh9YMI	ebookcentral.proquest.com	https://ebookcentral.proquest.com/lib/teacherscollege-ebooks/reader.action?docID=5153204&ppg=35
6	2019-01-02 01:22:56	128.59.82.245	D90KMGWNI41iEiw	ebookcentral.proquest.com	https://ebookcentral.proquest.com/lib/teacherscollege-ebooks/reader.action?docID=5224777&ppg=4
7	2019-01-02 01:23:48	128.59.82.245	D90KMGWNI41iEiw	search.proquest.com	https://search.proquest.com/pqdtglobal/advanced
8	2019-01-02 01:36:09	220.248.96.74	WYN2ZbxWvbYdiKw	psycnet.apa.org	http://psycnet.apa.org/doi/10.1037/0033-2909.134.3.460
9	2019-01-02 01:36:11	220.248.96.74	WYN2ZbxWvbYdiKw	tc.summon.serialssolutions.com	http://tc.summon.serialssolutions.com/search?ho=t&fv=ContentType,Newspaper%20Article,f%7CContentType,Journal%20Article,f%7CContentType,Dissertation,f...
10	2019-01-02 01:38:29	220.248.96.74	WYN2ZbxWvbYdiKw	journals.sagepub.com	http://journals.sagepub.com/doi/10.1177/2056305116672885
11	2019-01-02 01:38:37	220.248.96.74	WYN2ZbxWvbYdiKw	search.ebscohost.com	http://search.ebscohost.com/login.aspx?direct=true&authtype=ip,uid&bQuery=IS+0250-6807+AND+TI+%28%22THE+MEDIA+INFLUENCE+ON+BODY+IMAGE+OF+BRAZILIAN+TEA...
12	2019-01-02 01:38:44	220.248.96.74	WYN2ZbxWvbYdiKw	psycnet.apa.org	http://psycnet.apa.org/doi/10.1037/men0000030
13	2019-01-02 01:39:09	104.185.30.151	ZwrsOFp6oxix40	infotrac.galegroup.com	http://infotrac.galegroup.com/itweb/new30429?db=PPDS
14	2019-01-02 01:40:15	220.248.96.74	WYN2ZbxWvbYdiKw	find.galegroup.com	http://find.galegroup.com/openurl/openurl?url_ver=Z39.88-2004&ctx_enc=info:ofi:enc:UTF-8&rft_val_fmt=info:ofi:fmt:kev:mtx:journal&url_ctx_fmt=info:of...
15	2019-01-02 01:40:22	220.248.96.74	WYN2ZbxWvbYdiKw	journals.sagepub.com	https://journals.sagepub.com/doi/full/10.1177/1461448187710837utm_source=summon&utm_medium=discovery-provider
16	2019-01-02 01:41:57	220.248.96.74	WYN2ZbxWvbYdiKw	journals.sagepub.com	https://journals.sagepub.com/doi/full/10.1177/1557988317698627utm_source=summon&utm_medium=discovery-provider
17	2019-01-02 01:43:14	220.248.96.74	WYN2ZbxWvbYdiKw	journals.sagepub.com	http://journals.sagepub.com/doi/10.1177/0272431618770809
18	2019-01-02 01:48:47	220.248.96.74	WYN2ZbxWvbYdiKw	search.proquest.com	https://search.proquest.com/docview/20764492247pq-origsite=summon&accountid=14258
19	2019-01-02 01:51:31	220.248.96.74	WYN2ZbxWvbYdiKw	link.galegroup.com	http://link.galegroup.com/apps/doc/A525346367/AONE?sid=summon&u=new30429
20	2019-01-02 01:55:12	220.248.96.74	WYN2ZbxWvbYdiKw	journals.sagepub.com	http://journals.sagepub.com/doi/10.1177/0361684317725311
21	2019-01-02 02:52:42	220.75.236.206	czGuHyqeReRDyW8	www.tandfonline.com	http://www.tandfonline.com/doi/abs/10.1080/14767724.2011.605328
22	2019-01-02 04:25:18	183.98.90.212	lbw6FbNldce3flw	openurl.ebscohost.com	http://openurl.ebscohost.com/linksvc/linking.aspx?genre=article&issn=0094-0771&title=Middle+school+journal&date=2012&volume=43&issue=5&page=40&atitle...
23	2019-01-02 04:26:32	183.98.90.212	lbw6FbNldce3flw	openurl.ebscohost.com	http://openurl.ebscohost.com/linksvc/linking.aspx?genre=article&issn=0094-0771&title=Middle+school+journal&date=2012&volume=43&issue=5&page=40&atitle...
24	2019-01-02 04:27:19	183.98.90.212	lbw6FbNldce3flw	openurl.ebscohost.com	http://openurl.ebscohost.com/linksvc/linking.aspx?genre=article&issn=0094-0771&title=Middle+school+journal&date=2012&volume=43&issue=5&page=40&atitle...
25	2019-01-02 04:28:48	183.98.90.212	lbw6FbNldce3flw	openurl.ebscohost.com	http://openurl.ebscohost.com/linksvc/linking.aspx?genre=article&issn=0094-0771&title=Middle+school+journal&date=2012&volume=43&issue=5&page=40&atitle...
26	2019-01-02 04:31:52	183.98.90.212	lbw6FbNldce3flw	www.ncte.org	http://www.ncte.org/journals/la/issues
27	2019-01-02 04:32:19	183.98.90.212	lbw6FbNldce3flw	find.galegroup.com	http://find.galegroup.com/openurl/openurl?url_ver=Z39.88-2004&ctx_enc=info:ofi:enc:UTF-8&rft_val_fmt=info:ofi:fmt:kev:mtx:journal&url_ctx_fmt=info:of...
28	2019-01-02 04:34:38	183.98.90.212	lbw6FbNldce3flw	search.proquest.com	https://search.proquest.com/docview/1816204529pq-origsite=summon&accountid=14258
29	2019-01-02 04:36:17	183.98.90.212	lbw6FbNldce3flw	find.galegroup.com	http://find.galegroup.com/openurl/openurl?url_ver=Z39.88-2004&ctx_enc=info:ofi:enc:UTF-8&rft_val_fmt=info:ofi:fmt:kev:mtx:journal&url_ctx_fmt=info:of...

Also no.

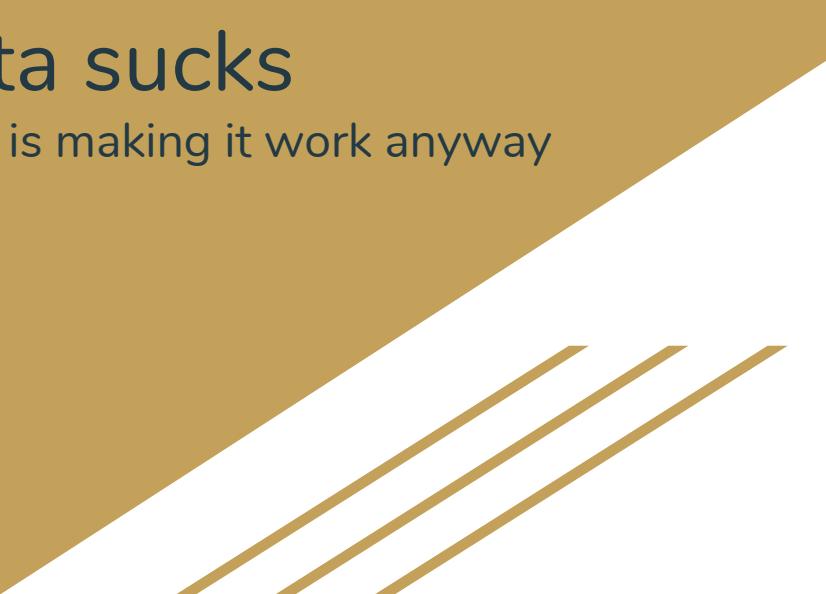
Library Sierra Data

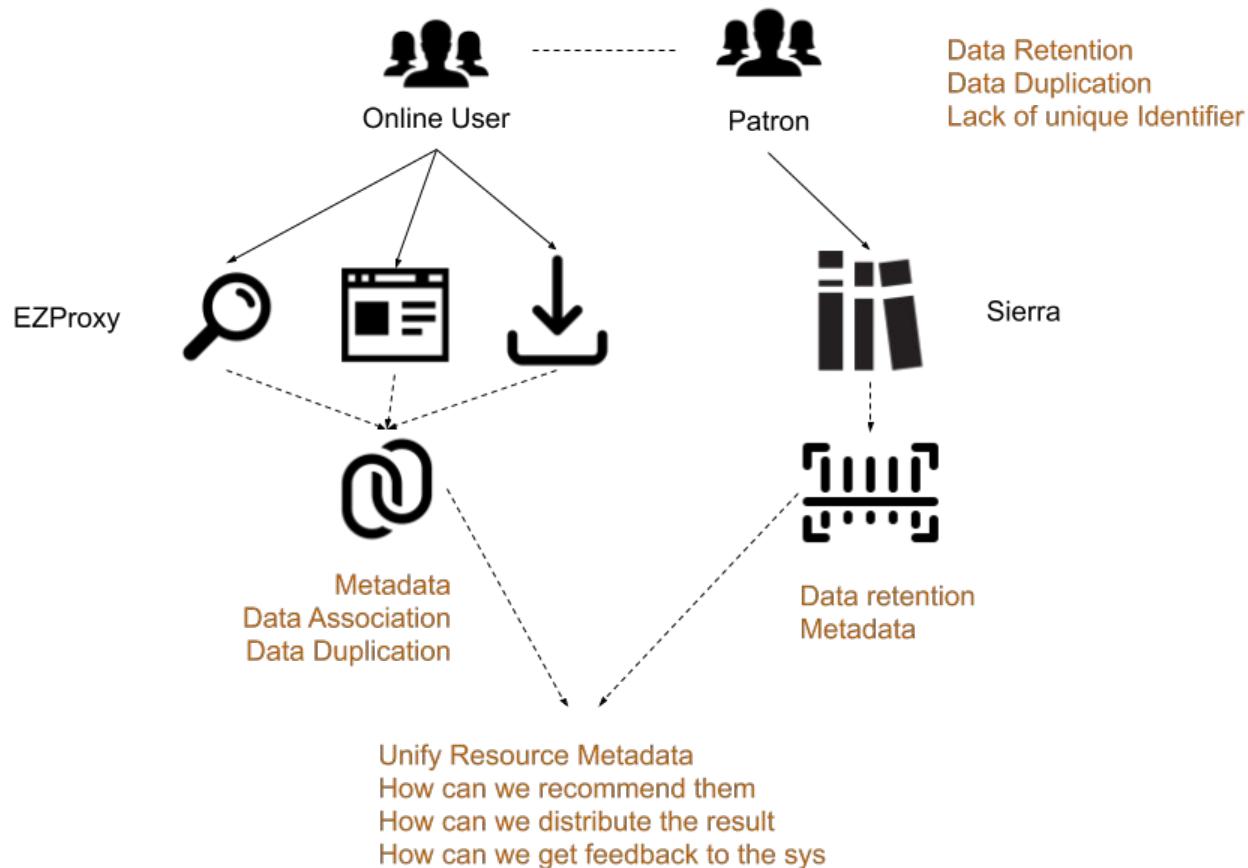
Id	Author	updatedDate	createdDate	Language	Title	PublishYear	CatalogDate	MaterialType	BibLevel	Country
1000002	American Association of Collegiate Registrars and Admissi...	2009-03-26	1992-08-21	English	Afghanistan, a guide to the academic placement of student...	1961	1996-10-14	PRINTED MATL	MONOGRAPH	New York
1000003		2009-03-26	1992-08-21	English	Accommodating the spectrum of individual abilities	1983	1994-06-08	PRINTED MATL	MONOGRAPH	New York
1000004	Simons, William E	2012-10-02	1992-08-21	English	Liberal education in the service academies	1965	1992-08-21	PRINTED MATL	MONOGRAPH	New York
1000006	Ripple, Richard E	2009-03-26	1992-08-21	English	The relationship of anxiety, creativity, and intelligence to s...	1966	1992-08-21	PRINTED MATL	MONOGRAPH	New York
1000007	Christen, Robert J	2009-03-26	1992-08-21	English	Monothemism and Moses,	1969	1992-08-21	PRINTED MATL	MONOGRAPH	Massachusetts
1000008	Harris, David P. (David Payne), 1925-	2009-03-26	1992-08-21	English	Testing English as a second language	1969	1992-08-21	PRINTED MATL	MONOGRAPH	New York
1000009	Madian, Jon	2017-12-13	1992-08-21	English	Beautiful junk; a story of the Watts Towers.	1968	1992-08-21	PRINTED MATL	MONOGRAPH	Massachusetts
1000010	Stuckenschmidt, Hans Heinz, 1901-1988	2009-03-26	1992-08-21	English	Twentieth century music.	1969	1992-08-21	PRINTED MATL	MONOGRAPH	New York
1000011	Stewart, Daniel K., 1925-	2009-03-26	1992-08-21	English	The psychology of communication	1968	1992-08-21	PRINTED MATL	MONOGRAPH	New York
1000012	D'Souza, Anthony	2012-10-02	1992-08-21	English	How people live in India.	1968	1992-08-21	PRINTED MATL	MONOGRAPH	New York
1000016	Jewish Theological Seminary of America. Institute for Religi...	2009-03-26	1992-08-21	English	American education and religion: the problem of religion i...	1952	1992-08-21	PRINTED MATL	MONOGRAPH	New York
1000017	McFarland, Andrew S., 1940-	2009-06-28	1992-08-21	English	Power and leadership in pluralist systems	1969	1992-08-21	PRINTED MATL	MONOGRAPH	New York
1000018	Surge, Frank	2012-10-02	1992-08-21	English	Singers of the blues	1969	1992-08-21	PRINTED MATL	MONOGRAPH	New York
1000019	Crane, Walter, 1845-1915	2012-10-02	1992-08-21	English	Of the decorative illustration of books old and new.	1968	1992-08-21	PRINTED MATL	MONOGRAPH	Michigan
1000020	Duhl, Leonard J	2009-03-26	1992-08-21	English	Mental health and urban social policy; a casebook of comm...	1968	1992-08-21	PRINTED MATL	MONOGRAPH	New York
1000021	Fisk University. Social Science Institute	2009-03-26	1992-08-21	English	Unwritten history of slavery; autobiographical accounts of...	1968	1992-08-21	PRINTED MATL	MONOGRAPH	District of Columbia
1000022	Noble, Iris	2017-08-10	1992-08-21	English	Firebrand for justice: a biography of Louis Dembitz Brandeis	1969	1992-08-21	PRINTED MATL	MONOGRAPH	New York
1000024	Wilson, Forrest, 1918-	2012-10-02	1992-08-21	English	What it feels like to be a building	1969	1992-08-21	PRINTED MATL	MONOGRAPH	New York
1000026	Monk, Kathleen	2012-10-02	1992-08-21	English	Fun with fabric printing	1969	1992-08-21	PRINTED MATL	MONOGRAPH	New York
1000027	Kübler-Ross, Elisabeth	2009-03-26	1992-08-21	English	On death and dying	1969	1992-08-21	PRINTED MATL	MONOGRAPH	New York
1000029	Allendoerfer, Carl B. (Carl Barnett), 1911-1974	2012-10-02	1992-08-21	English	Principles of mathematics	1969	1992-08-21	PRINTED MATL	MONOGRAPH	New York
1000031	Carse, Robert, 1902-1971	2012-10-02	1992-08-21	English	Keepers of the lights; a history of American lighthouses	1969	1992-08-21	PRINTED MATL	MONOGRAPH	New York
1000032	Dustin, David S., 1936-	2009-03-26	1992-08-21	English	How psychologists do research; the example of anxiety	1969	1992-08-21	PRINTED MATL	MONOGRAPH	New York
1000033	Zagoren, Ruby	2012-10-02	1992-08-21	English	Venture for freedom; the true story of an African Yankee.	1969	1992-08-21	PRINTED MATL	MONOGRAPH	Ohio
1000034	Helfman, Harry Carmozin, 1910-	2017-05-23	1992-08-21	English	Making pictures move	1969	1992-08-21	PRINTED MATL	MONOGRAPH	New York
1000036	Nolte, M. Chester (Mervin Chester), 1911-	2009-03-26	1992-08-21	English	Guide to school law	1969	1992-08-21	PRINTED MATL	MONOGRAPH	New York
1000037	Time-Life Books	2009-03-26	1992-08-21	English	This fabulous century.	1970	1992-08-21	PRINTED MATL	MONOGRAPH	New York
1000041	Woodford, Peggy	2012-10-02	1992-08-21	English	Schubert	1969	1992-08-21	PRINTED MATL	MONOGRAPH	New York
1000043	Duggins, Lydia A. (Lydia Adelia), 1913-	2009-03-26	1992-08-21	English	Developing children's perceptual skills in reading	1968	1992-08-21	PRINTED MATL	MONOGRAPH	Connecticut

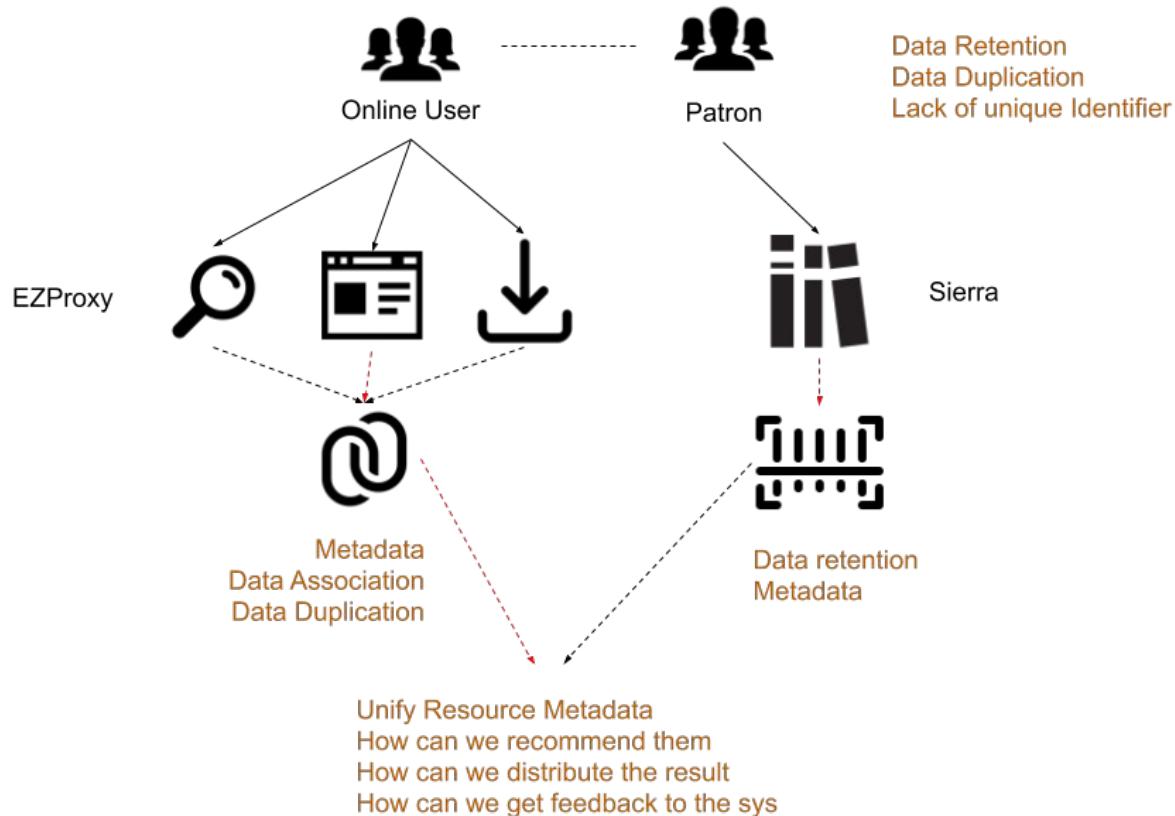


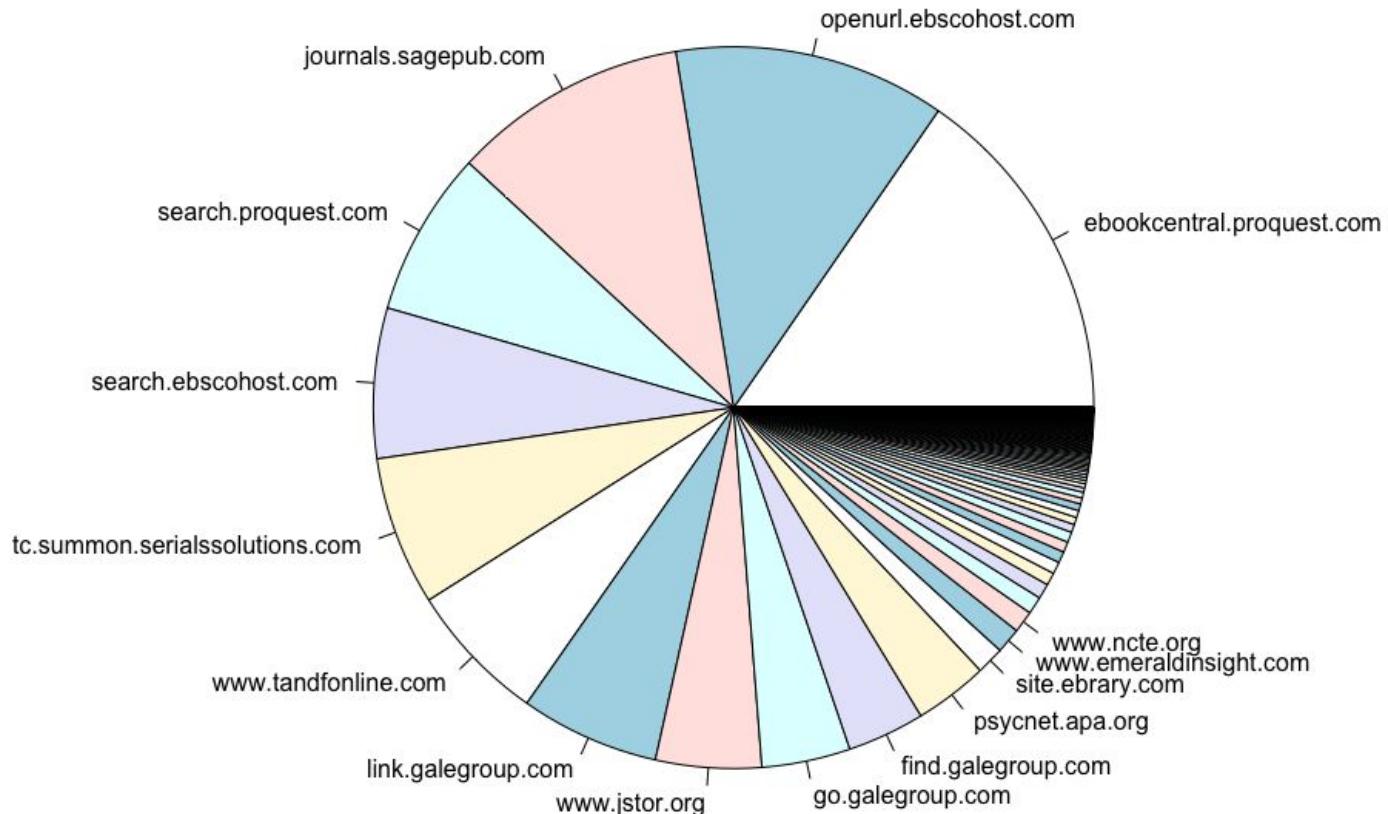
Lesson: “Real” data sucks

Primary job of data scientists is making it work anyway









Let's stare at links (this is what data scientists actually do)

What information do you see?

<http://journals.sagepub.com/doi/10.1177/0272431618770809>

<http://openurl.ebscohost.com/linksvc/linking.aspx?genre=article&issn=1538-6619&title=YC+young+children&date=2009&volume=64&issue=1&spage=32&atitle=When+Children+Have+Something+to+Say%2C+Writers+Are+Born&aulast=Stonier&aufirst=Francis>

<http://www.tandfonline.com/doi/abs/10.1080/15235882.2011.568831>

<http://www.emeraldinsight.com/10.1108/ET-02-2014-0006>

DOIs and OpenURL

- Turns out, other researchers have had the same problem
- Back in the day, no universal method for identifying or finding papers
- **Identifying:** DOIs - a unique number that identifies a particular article from some publication
- **Finding:** OpenURL - a standard format for requesting a paper from an academic database



Lesson:
Don't reinvent the wheel

What we wanted all along: useful metadata

- CrossRef (an open database of academic metadata) stores all the information we need.
- Armed with a DOI or OpenURL parameters, we can get this:

```
▼ {indexed: {date-parts: [[2019, 2, 13]], date-time: "2019-02-13T07:34:04Z", timestamp: 1550043244187,...}
  doi: "10.1155/2013/564864"
  ▶ ISSN: ["1687-725X", "1687-7268"]
  URL: "http://dx.doi.org/10.1155/2013/564864"
  abstract: "<jats:p>This paper presents a system for an interactive multimodal environment able (i) to train the listening comprehension >alternative-id: ["564864", "564864"]
  ▶ author: [{ORCID: "http://orcid.org/0000-0001-7083-4615", authenticated-orcid: true, given: "Sergio",...},...]
  container-title: "Journal of Sensors"
  container-title-short: "Journal of Sensors"
  content-domain: {domain: [], crossmark-restriction: false}
  ▶ created: {date-parts: [[2013, 7, 14]], date-time: "2013-07-14T21:02:13Z", timestamp: 1373835733000}
  ▶ deposited: {date-parts: [[2017, 6, 21]], date-time: "2017-06-21T16:15:05Z", timestamp: 1498061705000}
  ▶ funder: [{name: "Friuli Venezia Giulia Region (Italy)", award: []}]
  ▶ indexed: {date-parts: [[2019, 2, 13]], date-time: "2019-02-13T07:34:04Z", timestamp: 1550043244187}
  is-referenced-by-count: 2
  ▶ issued: {date-parts: [[2013]]}
  language: "en"
  ▶ license: [{URL: "http://creativecommons.org/licenses/by/3.0/",...}]
  ▶ link: [...]
  member: "98"
  original-title: []
  page: "1-12"
  prefix: "10.1155"
  ▶ published-print: {date-parts: [[2013]]}
  publisher: "Hindawi Limited"
  ▶ reference: [{key: "1", DOI: "10.1007/s10648-007-9047-2", doi-asserted-by: "publisher"}, {key: "3", year: "1986"},...]
  reference-count: 23
  references-count: 23
  ▶ relation: {cites: []}
  score: 1
  short-title: []
  source: "Crossref"
  subtitle: []
  title: "A Multimodal Learning System for Individuals with Sensorial, Neuropsychological, and Relational Impairments"
  type: "article-journal"
  volume: "2013"
```

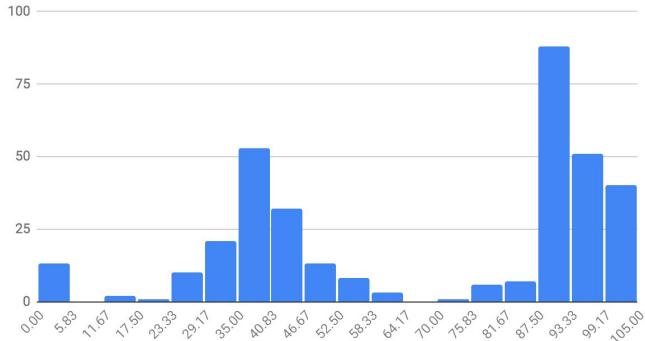
Side note: same goes for Library Sierra

- Similarly, we don't want to reinvent the wheel
- Use the information we have (title, author) to get the information we don't have (category, ISBN, metadata)
- Google Books API

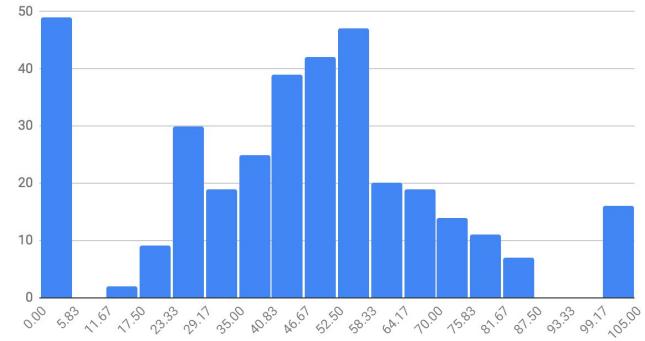
```
{  
  "kind": "books#volumes",  
  "totalItems": 967,  
  "items": [  
    {  
      "kind": "books#volume",  
      "id": "9g@6rpJ0eHkC",  
      "etag": "VAlsaDdaNV8",  
      "selfLink": "https://www.googleapis.com/books/v1/volumes/9g@6rpJ0eHkC",  
      "volumeInfo": {  
        "title": "Pinocchio",  
        "publishedDate": "1973-10",  
        "industryIdentifiers": [  
          {  
            "type": "ISBN_10",  
            "identifier": "0394826264"  
          },  
          {  
            "type": "ISBN_13",  
            "identifier": "9780394826264"  
          }  
        ],  
        "pageCount": 40,  
        "printType": "BOOK",  
        "categories": [  
          "Crafts & Hobbies"  
        ]  
      }  
    }  
  ]  
}
```

Side note: refining results

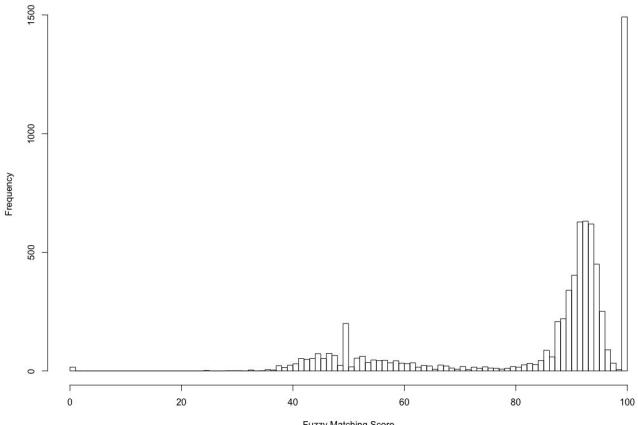
Histogram of Title Fuzzy Matching Scores TEST 1 (n=349)



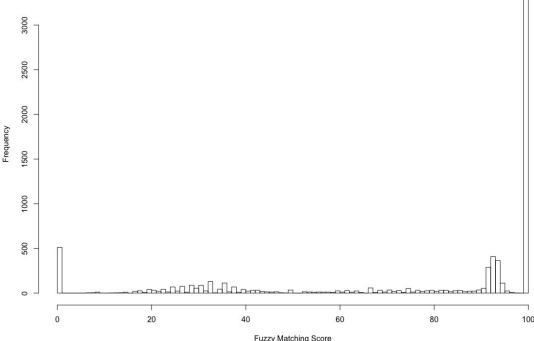
Histogram of Author Fuzzy Matching Scores TEST 1 (n=349)



Histogram of Title Fuzzy Matching Scores TEST 6 (n = 7164)



Histogram of Author Fuzzy Matching Scores TEST 6 (n = 7164)

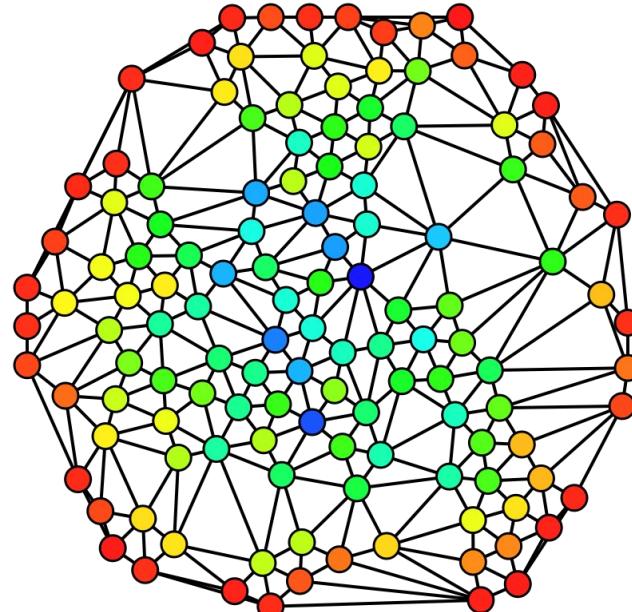




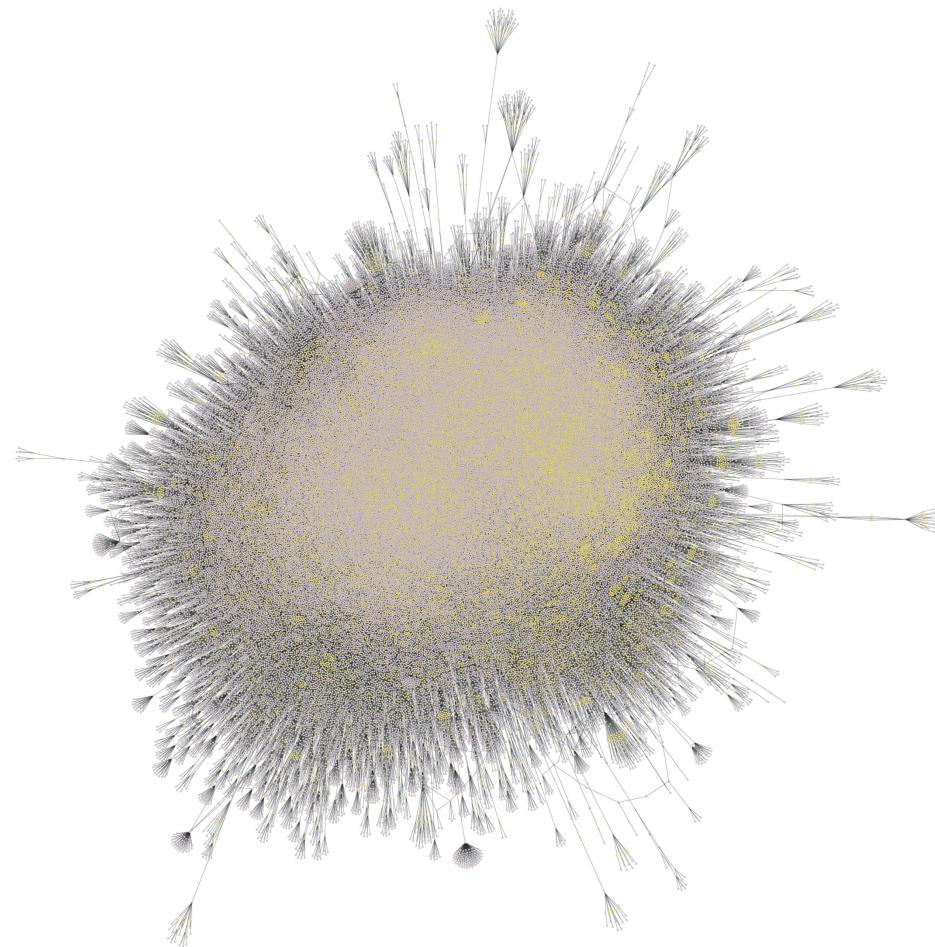
AUDIENCE PARTICIPATION SEGMENT: What's the best way to represent this data now that we have it?

Networks and Graphs

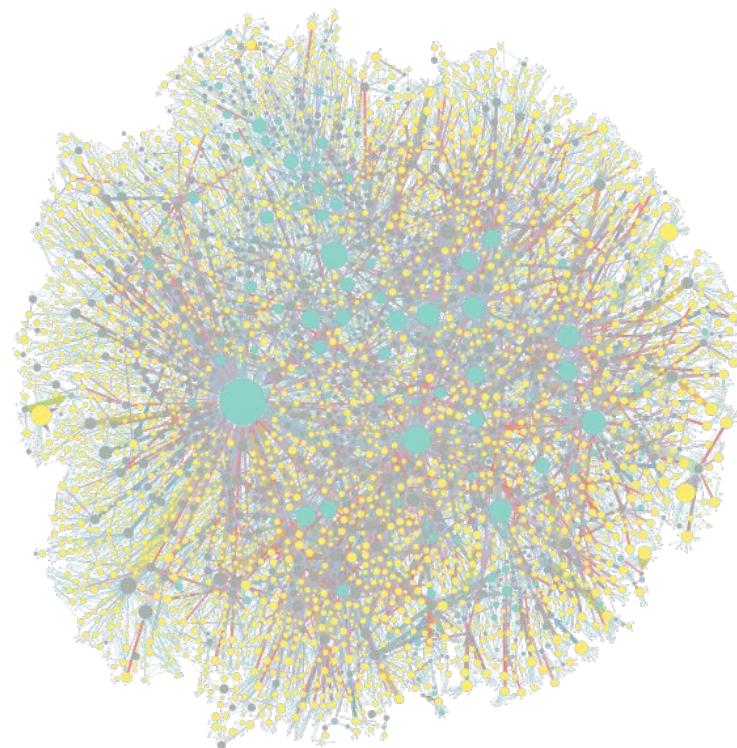
- Metadata is all the information which provides context for data
 - If our data is academic articles, the metadata situates that article in a context
- Academic data is inherently interconnected - how?
- Want to represent this data in some way which represents this inherent interconnectivity?
- Graphs are collections of nodes and edges



Everything users have ever looked at



Let's make it a bit clearer...



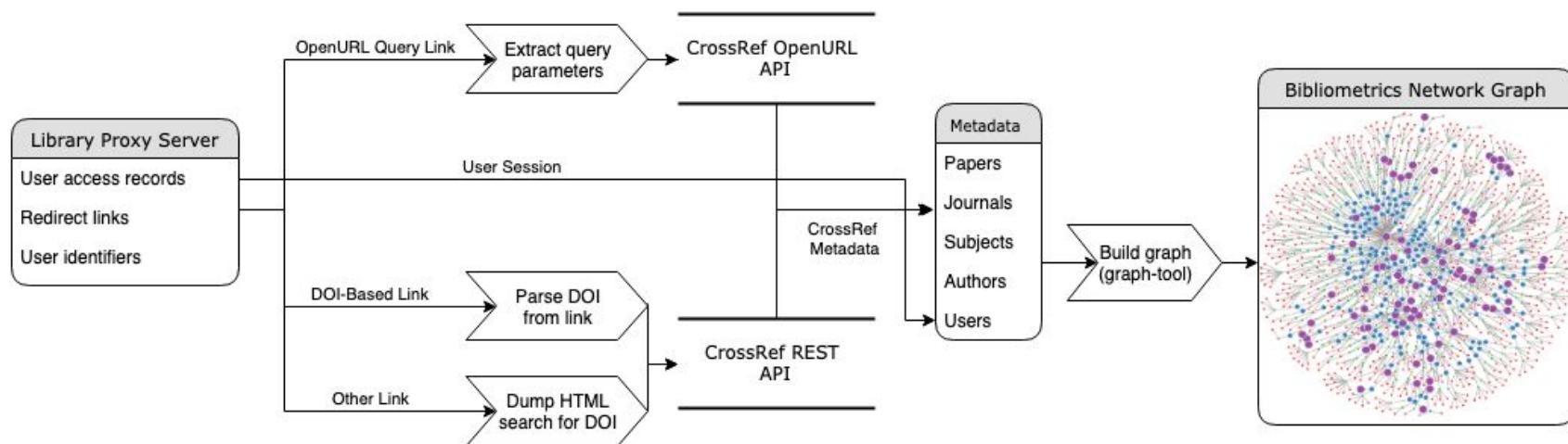
...what do you see?

Potential Graph Insights

- Node centrality
- Blockmodels (condense the graph)
- Inter-node distance
- Edge prediction - could form basis for recommendations
- Social networks and co-authorship



My EZProxy analysis pipeline - AERA Paper



BiblioNet - Open-Source Software for replicating this analysis

EdLabTC / Research

biblionet-ezproxy

Here's where you'll find this repository's source files. To give your users an idea of what they'll find here, [add a description to your repository](#).

master Filter files

/

Name	Size	Last commit	Message
↳ biblionet		5 days ago	stuff Former-commit-id: 4be6af802ccb05b20f75a1783...
↳ bin		5 days ago	stuff Former-commit-id: 4be6af802ccb05b20f75a1783...
↳ development		5 days ago	no node modules Former-commit-id: 772fe2739fe038bd...
↳ .gitignore	1.34 KB	5 days ago	ignore Former-commit-id: 94380fb45bf784270c28333...
↳ Dockerfile	686 B	2019-06-24	access records routine
↳ LICENSE	1.04 KB	2019-07-03	more packaging work
↳ MANIFEST.in	0 B	2019-07-03	more packaging work
↳ README.md	170 B	2019-06-17	update
↳ cloudbuild.yaml	158 B	2019-06-23	various cloud build things
↳ requirements.txt	197 B	2019-06-21	made Dockerfile
↳ savedrecs.txt	86.05 KB	2019-06-21	made Dockerfile
↳ setup.py	785 B	5 days ago	stuff Former-commit-id: 4be6af802ccb05b20f75a1783...

README.md

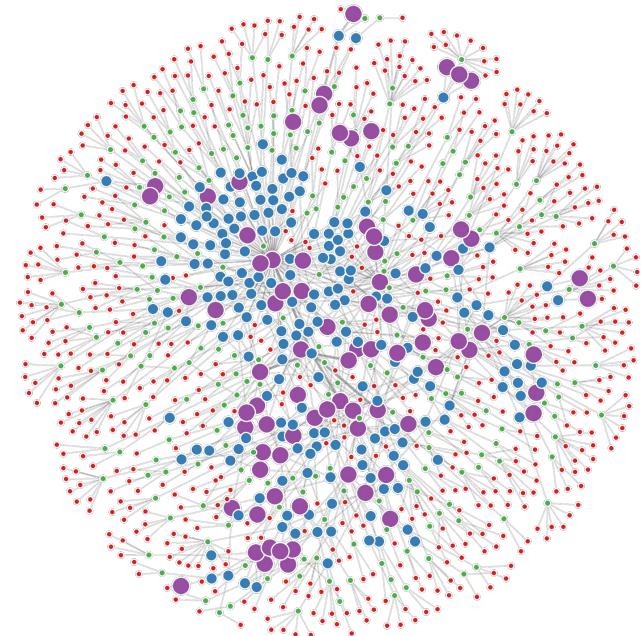
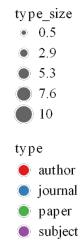
EdLab Search and Recommendation Research

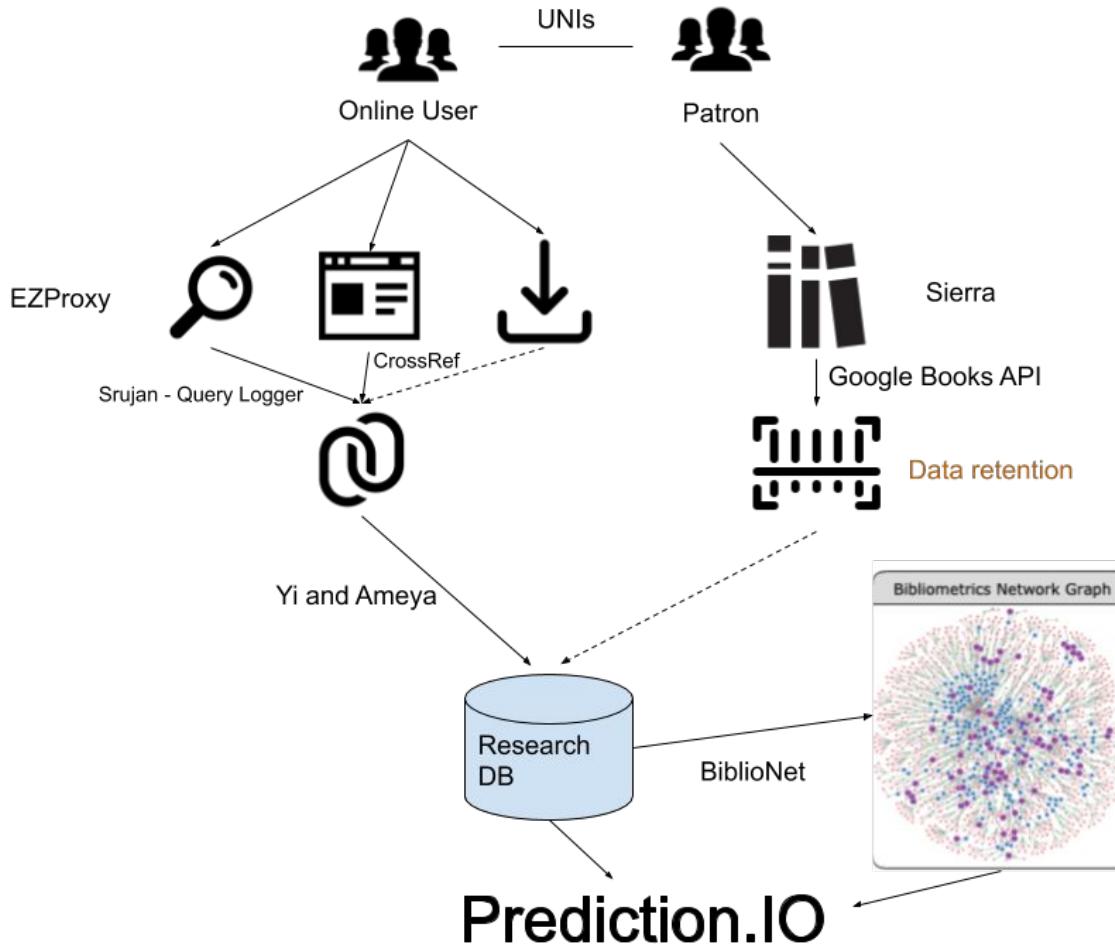
Research done as a DSI Scholar at EdLab, studying Search and Recommendation systems for education.

Henry Williams © 2019

Potential applications

- Live updating network of what library users are looking at right now
- Personal networks that users can explore to see works connected to what they've already seen
- Parameters like graph distance for Prediction.IO
- Cite-Finder: what papers should you be citing for your dissertation?
- Social network modeling (AMEYA)
- Comparing Library corpuses



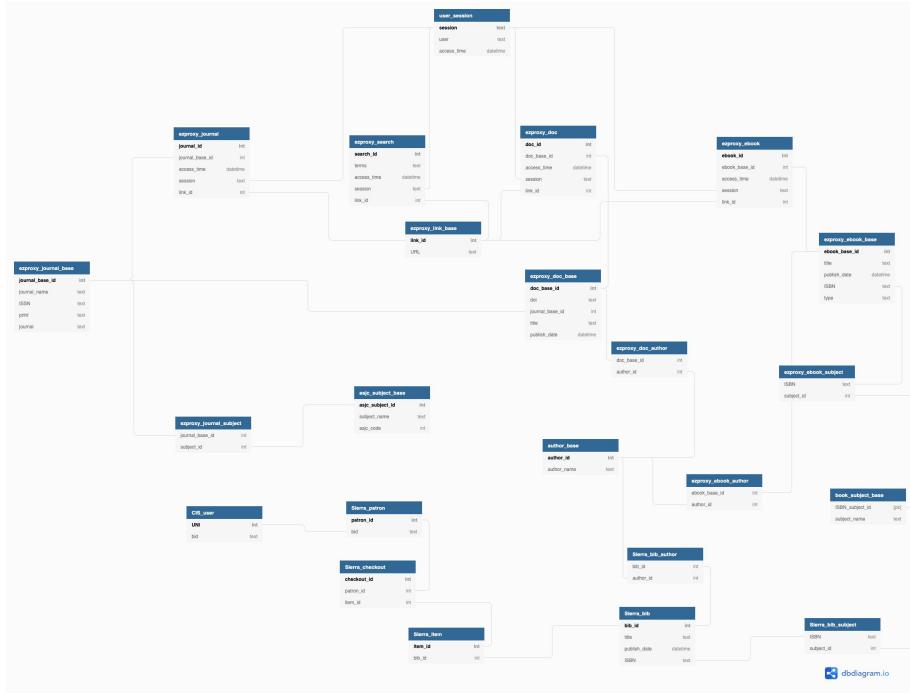


Lessons Learned

1. Education research requires development of new methods for recommendation - traditional methods have pitfalls
2. Real-world data is messy and difficult to use
3. #2 is often the MAIN BARRIER to research (this EZProxy data has been going unused - diamond in the rough)
4. Don't reinvent the wheel - rely on databases and techniques that already exist
5. Represent data according to its inherent structure
6. Try everything - fail often
 - a. Failed ideas: a webcrawler to download papers/circumvent JSTOR restrictions, using machine learning to parse pages or PDFs, using topic modeling

The Path Forward

1. Creating a research database including all available metadata
2. Using this as a basis for recommendations (Predictionio)
3. Doing social network analysis
4. Writing a journal paper on the method and open-sourcing the software



Curious?

The screenshot shows a user interface for the Rhizr platform. At the top, there is a navigation bar with the Rhizr logo, a search bar labeled "Search Rhizr", a "New Rhizr" button, and a user profile for "Henry". Below the header, the title "Henry Williams Projects" is displayed, followed by the subtext "This Rhizr has no description". The main content area is titled "Untitled Block View" and contains several project categories:

- Projects**: Biblionet Python Library, Analysis Pipeline.
- Research Papers**: Network Modeling Methods and Metadata Extraction for Library Access Records.
- Presentation & Seminar**: 07/23 Seminar Presentation.
- Others**: EdLab Bio, EdLab Blogs.
- Sierra Library Project**: Problem Statement, Google Books Integration, Topic Modeling and LDA.
- Network Analysis Project**: Graph Theory and Bibliometrics, BACKGROUND - Metaknowledge, Graph-Tool Python, Bibliometric Networks.
- EZProxy Project**: Problem Statement, BACKGROUND - DOIs, OpenURL, and link resolvers, Get Metadata from OpenURL, Alt Sources for Metadata, CrossRef REST API, EZProxy Data and Metadata.

At the bottom right of the main content area, there are icons for search, view, star, and message.

What else I've been working on

Today is my last day at the lab...

The New York Times Magazine

SUBSCRIBE Account ▾

≡ Q

FEATURE

Are These Teenagers Really Running a Presidential Campaign? Yes. (Maybe.)

The retired senator Mike Gravel gave two young fans his Twitter password and permission to campaign in his name. It might be a stunt — or the future of politics.

By JAMIE LAUREN KEILES

Tina Tyrrell for The New York Times

