



GRAPHING YALE'S THEATRICAL HISTORY

Catherine DeRose
Yale University Library Digital Humanities Laboratory

Alex O'Keefe
Robert B. Haas Family Arts Library

Yale Digital Conference | June 14, 2019



Why Ensemble@Yale?

THE SUBJECT:

Yale Theatre History

- DRA 37 Archival Collection
- Department / School of Drama - 1925
- Yale Repertory Theatre - 1966
- 100+ world premieres
- Famous alumni
 - Meryl Streep, Wendy Wasserstein, Angela Bassett, William Ivey Long, Liev Schreiber, Sigourney Weaver, Michael Yeargan, Frances McDormand, John Guare, Lupita Nyong'o...

DESIRED OUTCOMES:

- Increase access to minimally-processed archives
- Generate database of Yale theater history
- Answer reference questions

Project Background



Inspired by NYPL Labs' Ensemble project

- Limits of OCR
- Importance of human judgment
- Semantic relationships
- Hard for computers to see
- Easy for humans to intuit

2013



Project Background

TASKS:

- Digitize programs
- Determine organization and presentation
- Adapt Scribe with customized workflows for theatrical data
 - **Mark**
 - **Transcribe**

LESSONS LEARNED:

- Usability testing before going live
- Fixing errors/preventing junk data
- How to decide when marking or transcribing is completed?

Launched May 2017



ENSEMBLE

[MARK](#)[TRANSCRIBE](#)[ABOUT](#)[LOG IN](#)

Help create a database
of Yale theater history

Click a button below to get
started:

[MARK](#)[TRANSCRIBE](#)

Browse 90 years of programs to see what's available to mark or transcribe:



Demo - Marking

OH
Lab

ENSEMBLE

MARK

TRANSCRIBE

ABOUT

HELLO ANO28 LOGOUT

164 TEMPLE STREET

Now Open Mondays

YALE UNIVERSITY SCHOOL OF THE FINE ARTS
THE DEPARTMENT OF DRAMA
presents

The Taming of the Shrew

By WILLIAM SHAKESPEARE

Directed by
FRANK McMULLAN

Special Coaching by
CONSTANCE WELCH

Costumes by
FRANK BEVAN

Setting by JACQUELINE BEYMER

Lighting by PETE HOWARD

Music by RICHARD GOFTON

YALE UNIVERSITY THEATRE

19, 20, 21, 22, 23 April 1955 at 8:30 P. M.

Student Matinee 23 April 1955 at 2:30 P. M.

Click on a tool below to mark fields that will subsequently be transcribed.

☐ Title (and Subtitle)

Draw one rectangle around the title of the play, then click "Playwright" to proceed.

☐ Playwright

☐ Director

☐ Dates

☐ Actor and Role

☐ Staff Member and Role


FLAG AN ISSUE

Is marking complete for this program?


☐ NO

SUBMIT

Demo - Transcribing

 ENSEMBLE

MARKTRANSCRIBEABOUTLOGIN



Yale School of Drama Presents

**THE
LOWER
DEPTHS**

by

M

A

Type the title of the play. ?


Title

The Lower Depths

Subtitle (Optional)

BAD REGION ILLEGIBLE? NEXT ENTRY

Demo - Transcribing

 ENSEMBLE

MARKTRANSCRIBEABOUTLOG IN

NATASHA, her sister
MEDVEDYEV, their uncle, a policeman
PEPEL, a thief
KLESHCH, a locksmith
ANNA, his wife
NASTYA, a streetwalker
KVASHNYA, a meat pie peddler
BUENOV, a capmaker
BARON

Tessie Hogan
Patricia Clarkson
Steven Blye
Joseph Urla
Don Harvey
Susan MacDonald
Melissa Weil
Andi Chapman
Dylan Baker
Steven J. Gefroh
Thomas Costello
Christopher Noth
Michael Rogers
Allen Evans
Robert Alford II
Tim Whitby
Oren Jacoby
Evan Yionoulis

The time is early

sa on February 4th,
On February 5th
n will perform the

Type the name of the actor and the name of the character. ?

Character

Bubnov

Character Description (Optional)

a capmaker

Actor

Dylan Baker

BAD REGION ILLEGIBLE? RETURN TO MARKING

Events






Dealing with the Data

Steps for Completed Programs


- Initial Cleaning [OpenRefine]
- Removing Duplicates [Excel]
- Program Check [student worker / Excel]
- Normalization to Improve Searching
 - Names and Roles
 - Roles example: shift from 'scenery' to 'set design'
 - Choosing name authorities
 - AusStage, IBDB, Playbill Vault, Library of Congress, etc.
- Final Check

Standardizing Data with OpenRefine

 2019-05-08-drama.csv [Permalink](#) Open... Export Help

Facet / Filter Undo / Redo 0

Using facets and filters



Use facets and filters to select subsets of your data to act on. Choose facet and filter methods from the menus at the top of each data column.

Not sure how to get started?
[Watch these screencasts](#)

2430 rows Extensions:

Show as: **rows** records Show: 5 10 25 50 rows « first < previous 1 - 50 next > last »

<input type="checkbox"/> All	<input type="checkbox"/> Set Key	<input type="checkbox"/> Title	<input type="checkbox"/> Year	<input type="checkbox"/> Opening	<input type="checkbox"/> Closing	<input type="checkbox"/> Person [Original]	<input type="checkbox"/> Person [Normalized]	<input type="checkbox"/> Role [Original]	<input type="checkbox"/> Role [Normalized]	<input type="checkbox"/> Role	
		1. dra037-s01-b001-f004-i01	The Singing Minute	1927	11/8/27	11/12/27	Homer Abegglen	Abegglen, Homer	Mr. Snell	Mr. Snell	Cast
		2. dra037-s01-b001-f005-i02	What's A Fixer For?	1928	5/1/28	5/5/28	Homer Abegglen	Abegglen, Homer	Stage Carpenter	Scenic Artist	staff
		3. dra037-s01-b001-f005-i02	All People	1928	5/1/28	5/5/28	Homer Abegglen	Abegglen, Homer	Stage Carpenter	Scenic Artist	staff
		4. dra037-s01-b001-f005-i02	The Crab	1928	5/1/28	5/5/28	Homer Abegglen	Abegglen, Homer	Stage Carpenter	Scenic Artist	staff
		5. dra037-s01-b001-f005-i02	Minnie Field	1928	5/1/28	5/5/28	Homer Abegglen	Abegglen, Homer	Stage Carpenter	Scenic Artist	staff
		6. dra037-s01-b001-f005-i03	Brand	1928	5/28/28	5/29/28	Homer Abegglen	Abegglen, Homer	Properties	Properties Master / Mistress	Staff
		7. dra037-s01-b001-f013-i12	It's Morning	1940	2/5/40	2/7/40	Herbert Abrahams	Abrahams, Herbert	Slaves	Slaves	cast
		8. dra037-s01-b001-f003-i01	Machiavelly	1927	4/30/27	4/30/27	Theodore C. Achilles	Achilles, Theodore C.	First Soldier	First Soldier	Cast
		9. dra037-s01-b001-f003-i01	Machiavelly	1927	4/30/27	4/30/27	Theodore C. Achilles	Achilles, Theodore C.	Lepers	Lepers	Cast
		10. dra037-s01-b001-f003-i01	Machiavelly	1927	4/30/27	4/30/27	Theodore C. Achilles	Achilles, Theodore C.	Soldiers	Soldiers	Cast
		11. dra037-s01-b001-f004-i04	Finished	1928	2/13/28	2/14/28	T. C. Achilles	Achilles, Theodore C.	People on Train	People on Train	Cast
		12. dra037-s01-b001-f004-i05	Pueblo	1928	2/29/28	3/1/28	Theodore C. Achilles	Achilles, Theodore C.	Shamans	Shamans	Cast
		13. dra037-s01-b001-f004-i07	Finished	1928	3/15/28	3/16/28	T. C. Achilles	Achilles, Theodore C.	People on Train	People on Train	Cast
		14. dra037-s01-b001-f005-i03	Brand	1928	5/28/28	5/29/28	T. C. Achilles	Achilles, Theodore C.	People of the Hamlet	People of the Hamlet	Cast
		15. dra037-s01-b001-f020-i05	All The King's Men	1949	12/14/49	12/16/49	Robert Ackart	Ackart, Robert	Stage Crew		Staff
		16. dra037-s01-b001-f006-i06	Welcome Home	1929	5/8/29	5/11/29	Tony Acri	Acri, Tony	Cora's Sons	Cora's Sons	cast
		17. dra037-s01-b001-f002-i04	Lazy	1927	1/31/27	2/3/27	Thomas J. Ahearn	Ahearn, Thomas J.	Moses Newbolt	Moses Newbolt	Cast
		18. dra037-s01-b001-f002-i06	Lazy	1927	2/22/27	2/22/27	Thomas J. Ahearn	Ahearn, Thomas J.	Moses Newbolt	Moses Newbolt	Cast
		19. dra037-s01-b001-f002-i07	Twelve Before Three	1927	3/16/27	3/19/27	Thomas J. Ahearn	Ahearn, Thomas J.	Playwright	Playwright	playwright
		20. dra037-s01-b001-f003-i01	Machiavelly	1927	4/30/27	4/30/27	Thomas J. Ahearn	Ahearn, Thomas J.	Lepers	Lepers	Cast
		21. dra037-s01-b001-	Machiavelly	1927	4/30/27	4/30/27	Thomas J. Ahearn	Ahearn, Thomas J.	Soldiers	Soldiers	Cast

Standardizing Names

Refine

2019-05-08-drama.csv Permalink

Facet / Filter Undo / Redo 0

Refresh Reset All Remove All

Person [Original] change

1101 choices Sort by: name count Cluster

A. Nicholas Vardac 2
Ada Ward 1
Addison Powell 2
Adelaide C. Whitman 2
Adrian Sayre Harris 1
Agnes Bane 1
Agnes Birney 1
Al Bennett 1
Alan Baxter 1
Alan Ferry 5
Alan Harper 1
Alan L. Ferry 3

2430 rows

Show as: rows records Show: 5 10

All Set Key Title

1.	dra037-s01-b001-f004-i01	The Singin
2.	dra037-s01-b001-f005-i02	What's A F
3.	dra037-s01-b001-f005-i02	All People
4.	dra037-s01-b001-f005-i02	The Crab
5.	dra037-s01-b001-f005-i02	Minnie Fie
6.	dra037-s01-b001-f005-i03	Brand
7.	dra037-s01-b001-f013-i12	It's Mornin
8.	dra037-s01-b001-f003-i01	Machiavell
9.	dra037-s01-b001-f003-i01	Machiavell
10.	dra037-s01-b001-f003-i01	Machiavell
11.	dra037-s01-b001-f004-i04	Finished
12.	dra037-s01-b001-f004-i05	Pueblo
13.	dra037-s01-b001-f004-i07	Finished
14.	dra037-s01-b001-f005-i03	Brand
15.	dra037-s01-b001-f020-i05	All The Kir
16.	dra037-s01-b001-f006-i06	Welcome t
17.	dra037-s01-b001-f002-i04	Lazy
18.	dra037-s01-b001-f002-i06	Lazy
19.	dra037-s01-b001-f002-i07	Twelve Before Three
20.	dra037-s01-b001-f002-i07	Machiavelly

Cluster & Edit column "Person [Original]"

This feature helps you find groups of different cell values that might be alternative representations of the same thing. For example, the two strings "New York" and "new york" are very likely to refer to the same concept and just have capitalization differences, and "Gödel" and "Godel" probably refer to the same person. [Find out more ...](#)

Method key collision

Keying Function fingerprint

7 clusters found

Cluster Size	Row Count	Values in Cluster	Merge?	New Cell Value
2	8	<ul style="list-style-type: none">Herschel V. Williams Jr. (5 rows)Herschel V. Williams Jr (3 rows)	<input checked="" type="checkbox"/>	Herschel V. Williams Jr.
2	3	<ul style="list-style-type: none">Robert MacKaye (2 rows)Robert Mackaye (1 rows)	<input checked="" type="checkbox"/>	Robert MacKaye
2	2	<ul style="list-style-type: none">R. J. H. Powel Jr (1 rows)R. J. H. Powel, Jr. (1 rows)	<input checked="" type="checkbox"/>	R. J. H. Powel Jr
2	7	<ul style="list-style-type: none">Keith Mackaye (4 rows)Keith MacKaye (3 rows)	<input checked="" type="checkbox"/>	Keith Mackaye
2	4	<ul style="list-style-type: none">William J. Macmillan (3 rows)William J. MacMillan (1 rows)	<input checked="" type="checkbox"/>	William J. Macmillan
2	4	<ul style="list-style-type: none">Richard H. Diggs Jr (3 rows)Richard H. Diggs Jr. (1 rows)	<input checked="" type="checkbox"/>	Richard H. Diggs Jr
2	2	<ul style="list-style-type: none">Edwin Bradley Jr (1 rows)Edwin Bradley, Jr. (1 rows)	<input checked="" type="checkbox"/>	Edwin Bradley Jr

Select All Unselect All

Export Clusters

Merge Selected & Re-Cluster

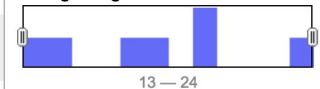
Merge Selected & Close

Close

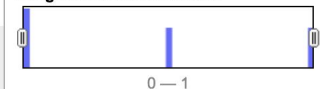
Rows in Cluster



Average Length of Choices



Length Variance of Choices



1927	2/22/27	2/22/27	Thomas J. Ahearn	Ahearn, Thomas J.	Moses Newbolt	Moses Newbolt	Cast
1927	3/16/27	3/19/27	Thomas J. Ahearn	Ahearn, Thomas J.	Playwright	Playwright	playwright
1927	4/30/27	4/30/27	Thomas J. Ahearn	Ahearn, Thomas J.	Lepers	Lepers	Cast

Processing Data: Creating Edges

```
import csv
import itertools
from collections import defaultdict

peopleList = []
x = 0

with open("2019-05-08-drama.csv", "r", encoding='UTF-8') as f:
    reader = csv.reader(f)
    writer = csv.writer(f)

    headers = next(reader)
    headers.append('id')
    writer.writerow(headers)

    for row in reader:
        # add unique ID for each person[normalized]
        person = row[6]
        if person not in peopleList:
            peopleList.append(person)
            x += 1
            #print(person, x)
            row.extend([str(x)])
            writer.writerow(row)
```

```
import csv
import itertools
from collections import defaultdict

d = defaultdict(list)
plays = {}

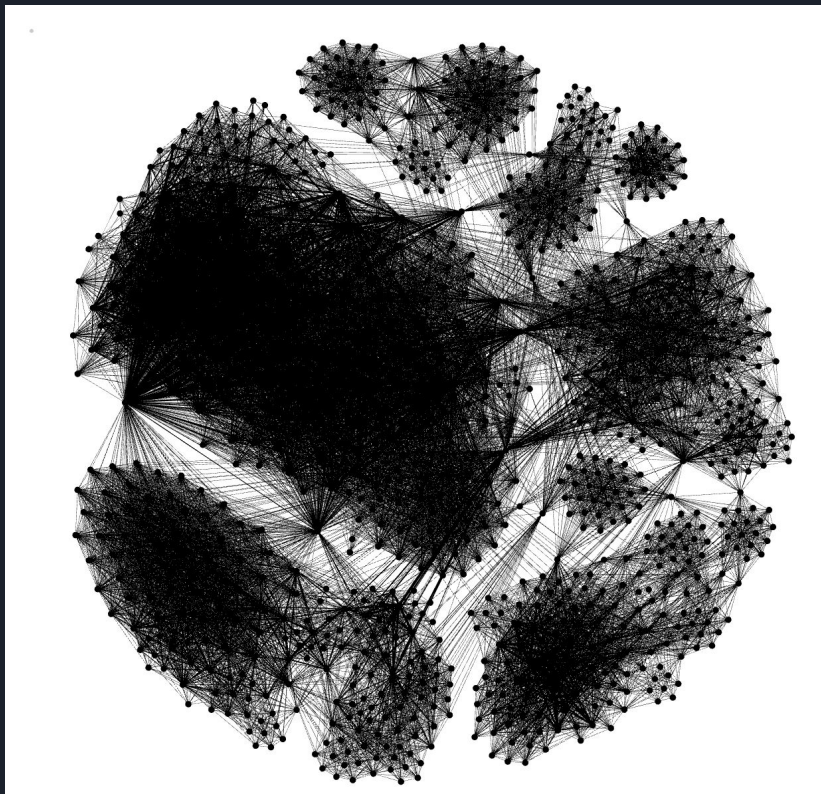
with open("2019-05-08-drama-ids.csv", "r", encoding='UTF-8') as f:
    reader = csv.reader(f)
    writer = csv.writer(f)

    output.write('source' + ',' + 'target' + ',' + 'label' + '\n')
    headers = next(reader)

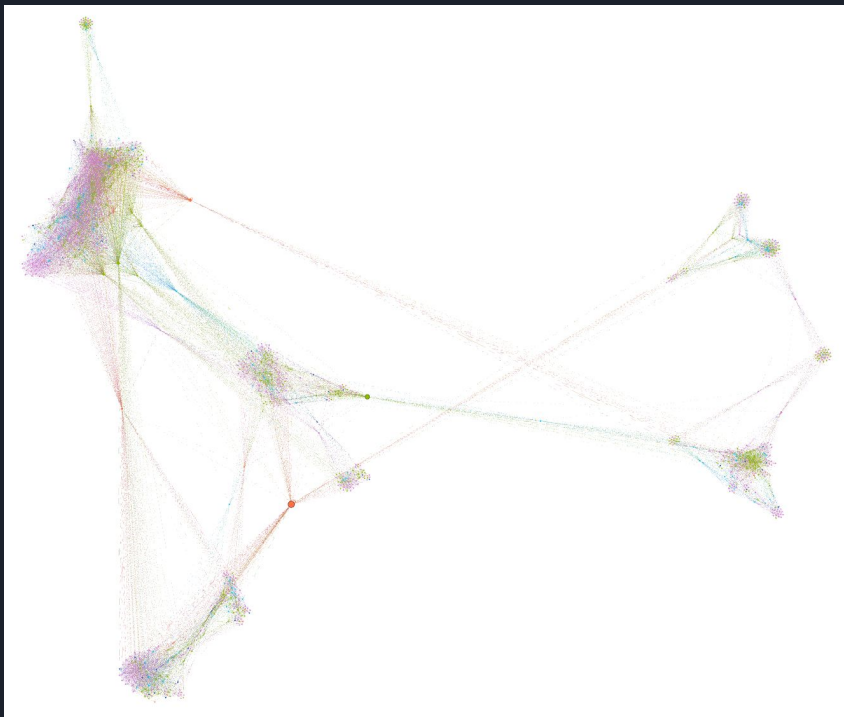
    # populate dictionary
    for row in reader:
        d[row[0]+'-'+row[1]+'-'+row[2]].append(row[10])

    # create combinatorial edges
    for key, value in d.items():
```

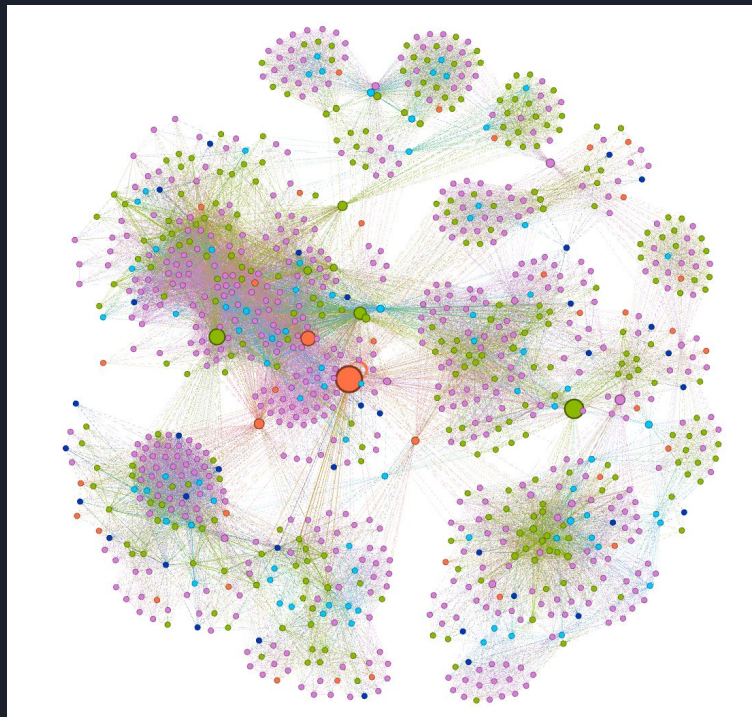
Experiments with Networks



Experiments with Networks: Layout

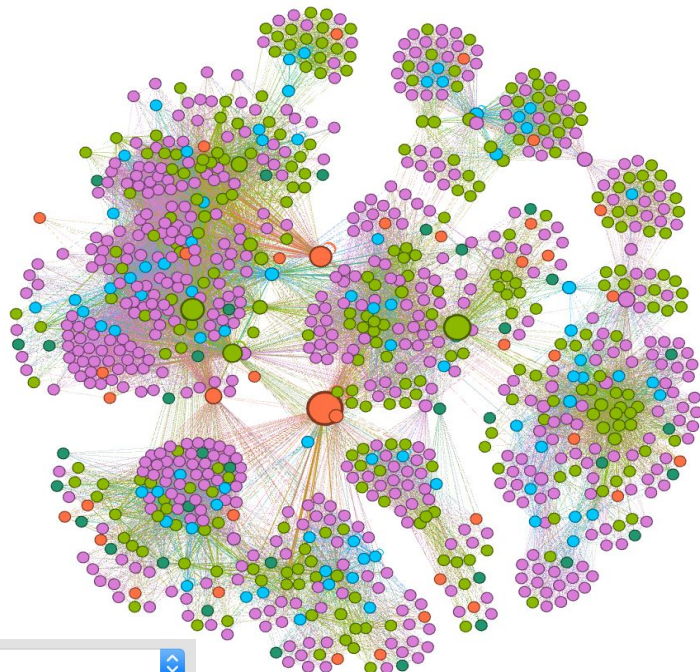


Force Atlas 2



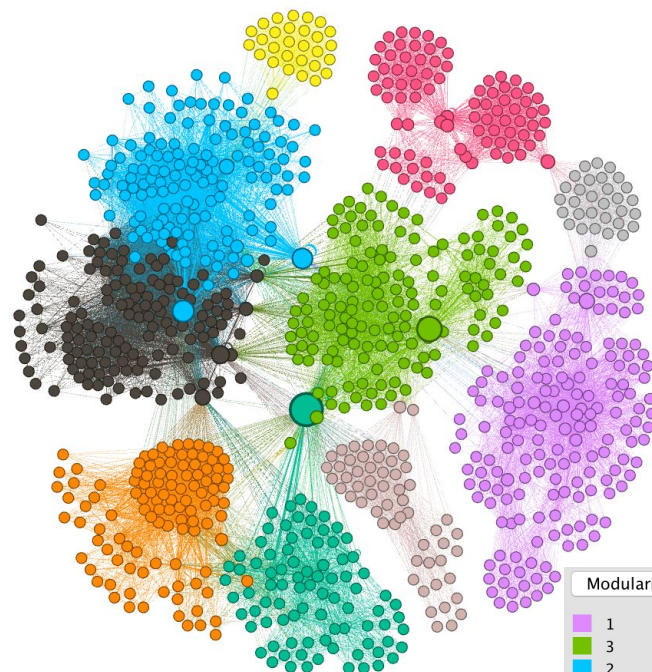
Fruchterman Reingold

Experiments with Networks: Color



role-majority

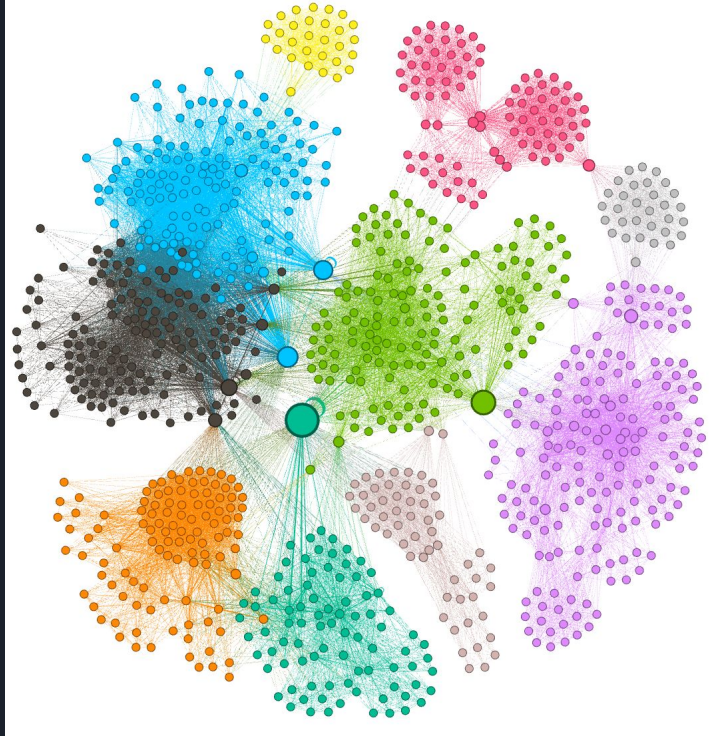
Cast	(55.4%)
Staff	(29.06%)
Multiple	(8.17%)
Director	(3.83%)
Playwright	(3.53%)



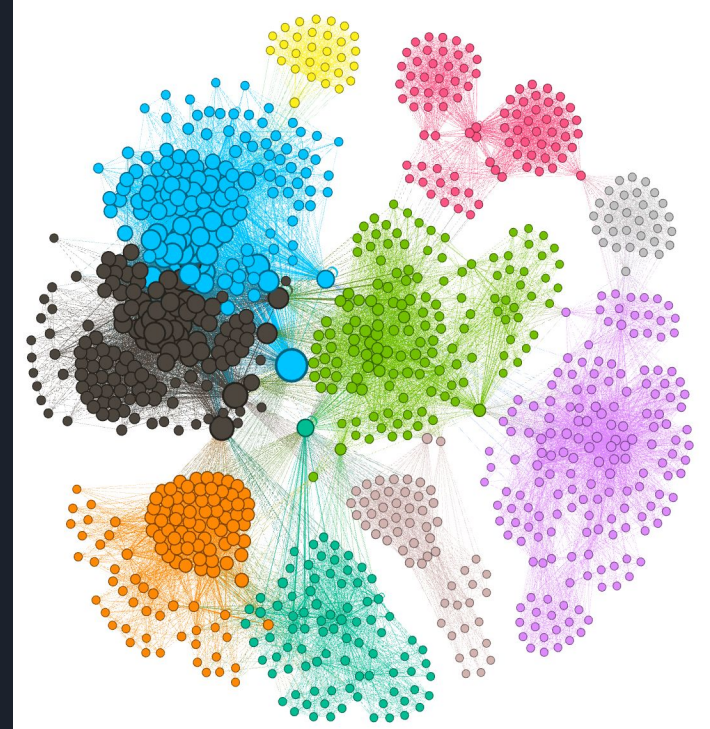
Modularity Class

1	(17.15%)
3	(14.53%)
2	(13.93%)
0	(13.82%)
8	(11.1%)
6	(8.88%)
7	(8.78%)
4	(5.75%)
9	(3.13%)
5	(2.93%)

Experiments with Networks: Size

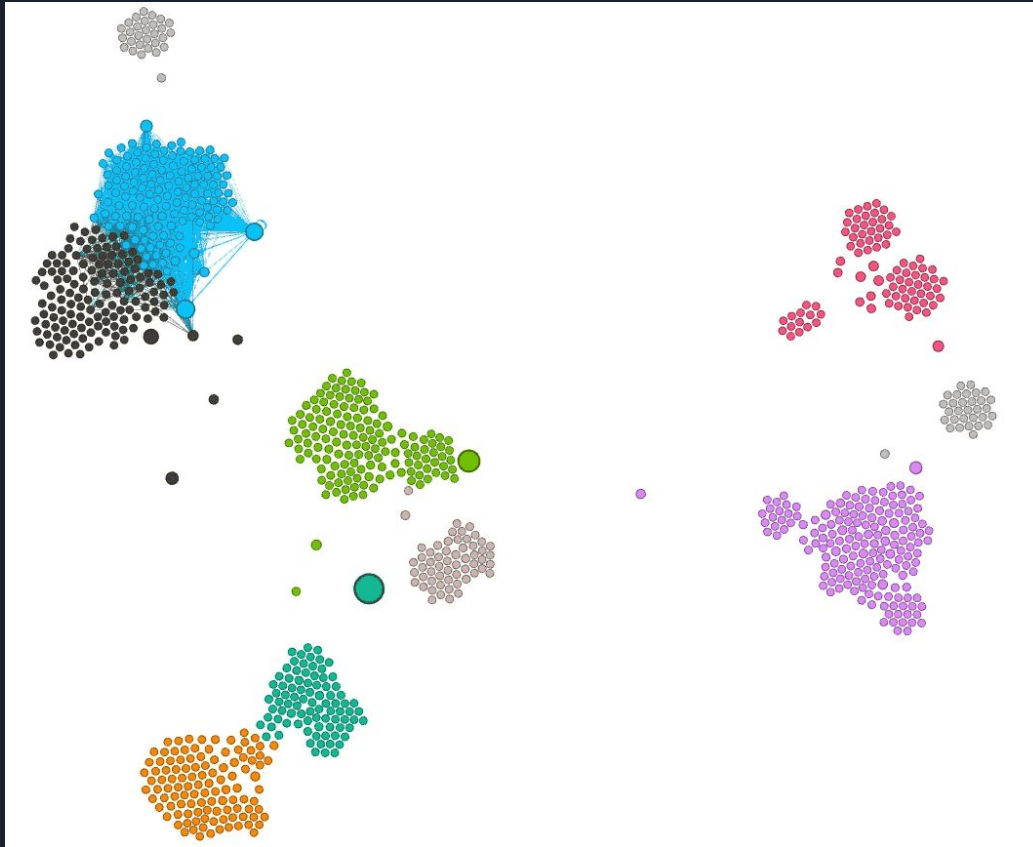


Betweenness Centrality



Eigenvector Centrality

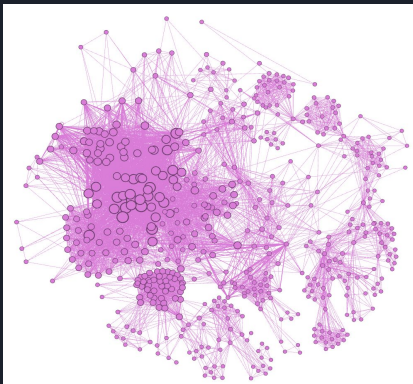
Experiments with Networks: Time



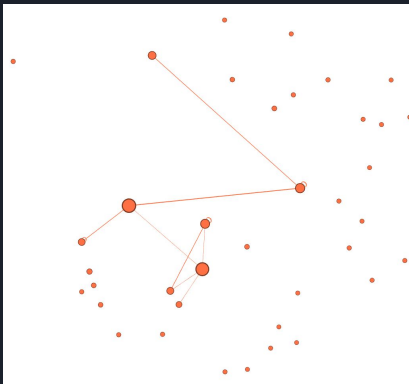
Animating
connections:
1925 to 1952

Experiments with Networks: Partitions

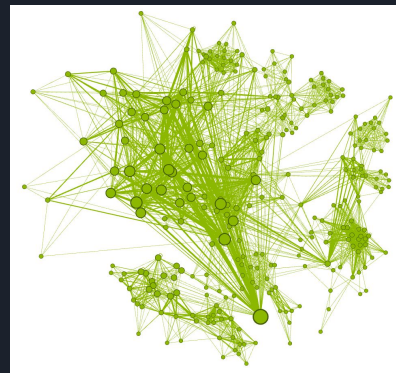
Cast



Directors



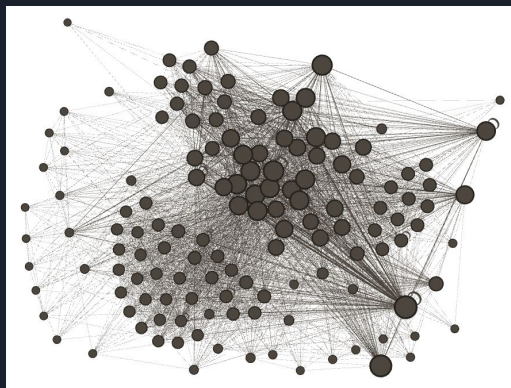
Staff



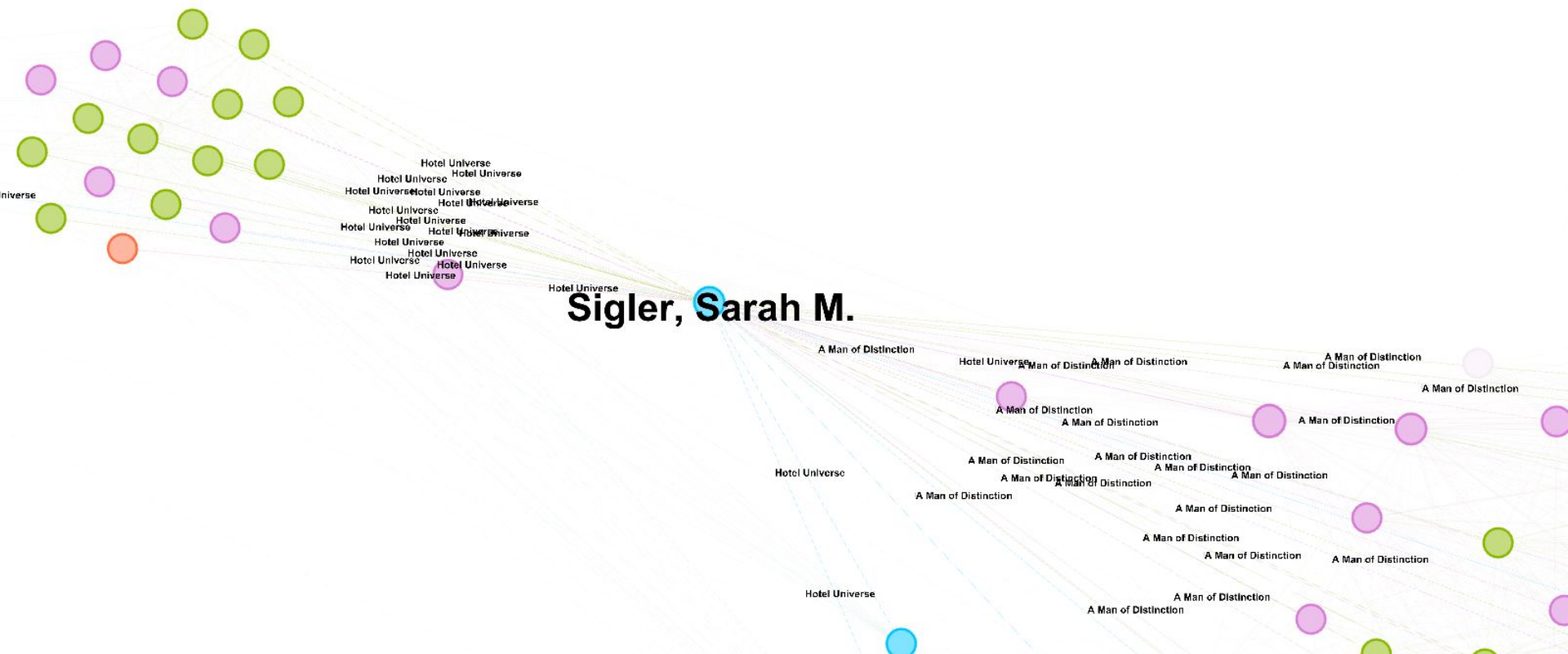
Modularity Group 2



Modularity Group 0



Finding Individual Networks



Thank You! Questions?

Catherine DeRose

Digital Humanities Lab Manager
catherine.derosé@yale.edu

Alex O'Keefe

Arts Digital Projects Librarian
alexandra.okeefe@yale.edu

Acknowledgments: Paul Beaudoin | Tess Colwell | Douglas Duhaime | Mauricio Giraldo | Lindsay King | Peter Leonard | Francesca Livermore | Steven Padla | Alexandra Provo | Danielle Reay | Monica Ong Reed | Doug Reside | Ben Vershbow | Maria Zapata

ensemble.yale.edu

