

Serendip

Topic Model-Driven Visual Exploration of Text Corpora

Eric Alexander, *Department of Computer Science, University of Wisconsin-Madison*

Joe Kohlmann, *Department of Computer Science, University of Wisconsin-Madison*

Robin Valenza, *Department of English, University of Wisconsin-Madison*

Michael Witmore, *Folger Shakespeare Library in Washington, D.C.*

Michael Gleicher, Member, IEEE, *Department of Computer Science, University of Wisconsin-Madison*



Topic controls

Sorting

Selection

Advanced Select

No topics selected

Document controls

Sorting

Aggregation

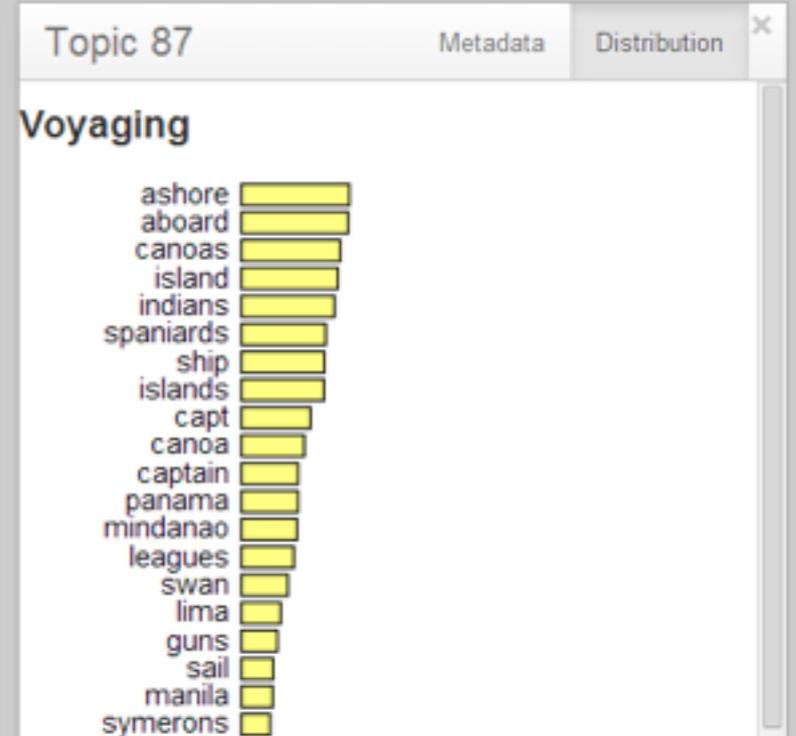
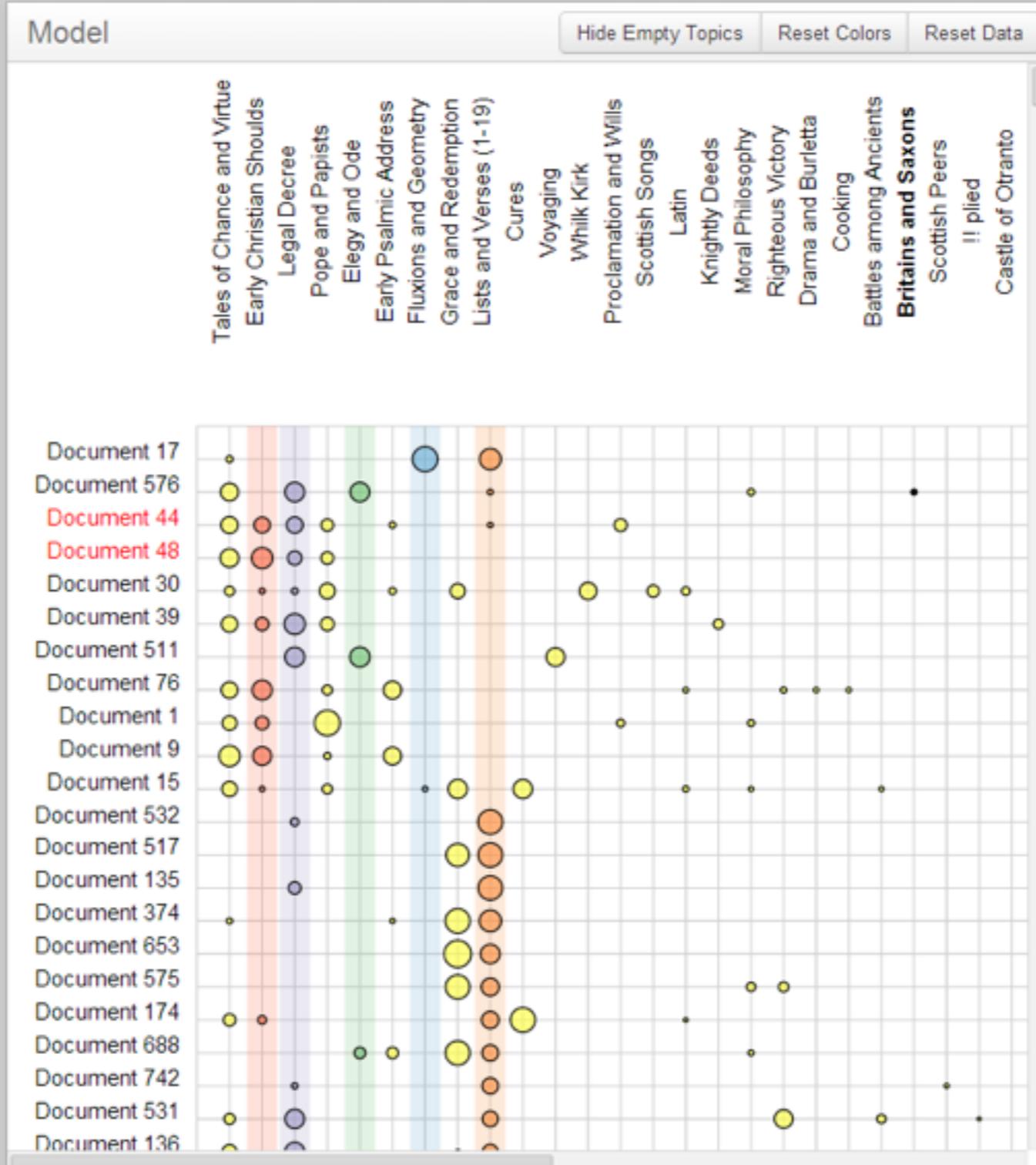
Filtering

Selection

Advanced Select

Document 44

Document 48



Document 48

Metadata Graph Counts

Document 48

Title: A breve cronycle of the Bysshope of Romes blessynge and of his prelates beneficiall and charitable rewardes. from the tyme of Kynge Heralde vnto this daye.[Breve cronycle of the Bysshope of Romes blessynge]Breve cronycle of the Bysshope of Romes blessynge

Author: Gybson, Thomas, d. 1562.

Genre: History

Publisher: By [R. Wyer for] Iohn Daye, dwellyng in Sepulchres parish at the signe of the Resurrection a lyle aboute Holbourne Conduite. Cum priuilegio regali, ad imprimendum solum,

Group: 1540

Token Count: 1757

Serendipity



- „The Three Princes of Serendip“
- Happy accidents
- Fate or extreme cleverness
- Research: How to make coincidence more likely?
- „The bohemian bookshelf“ by Thudt et al

Promoting Serendipity

A. Thudt, U. Hinrichs, and S. Carpendale

„The bohemian bookshelf: supporting serendipitous book discoveries through information visualization“

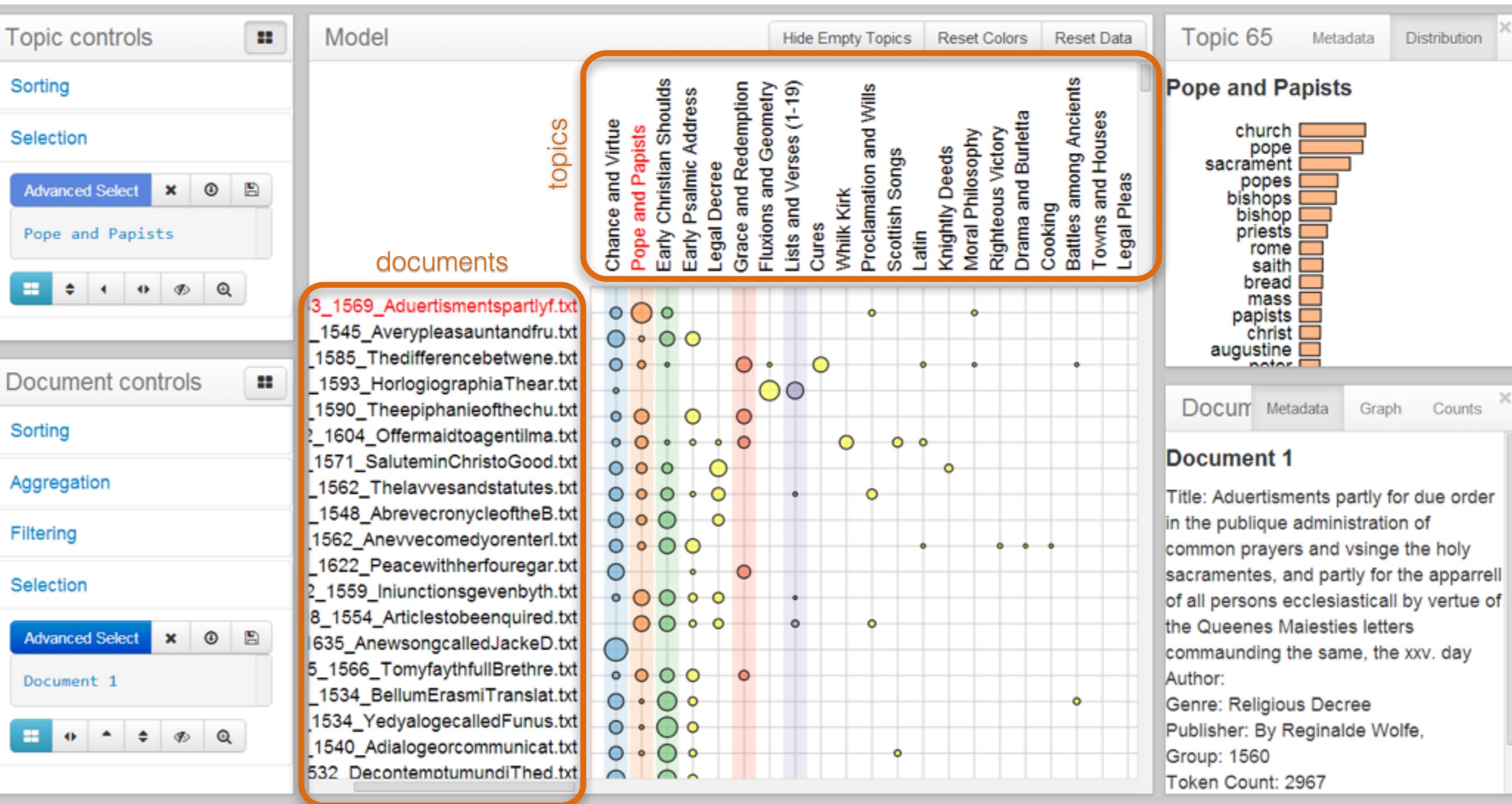
In Proc. ACM Human Factors in Computing Systems

- Providing **multiple access points**
- Highlighting **adjacencies**
- Offering **flexible pathways** for exploration
- Enticing **curiosity and playfulness**

Three Views

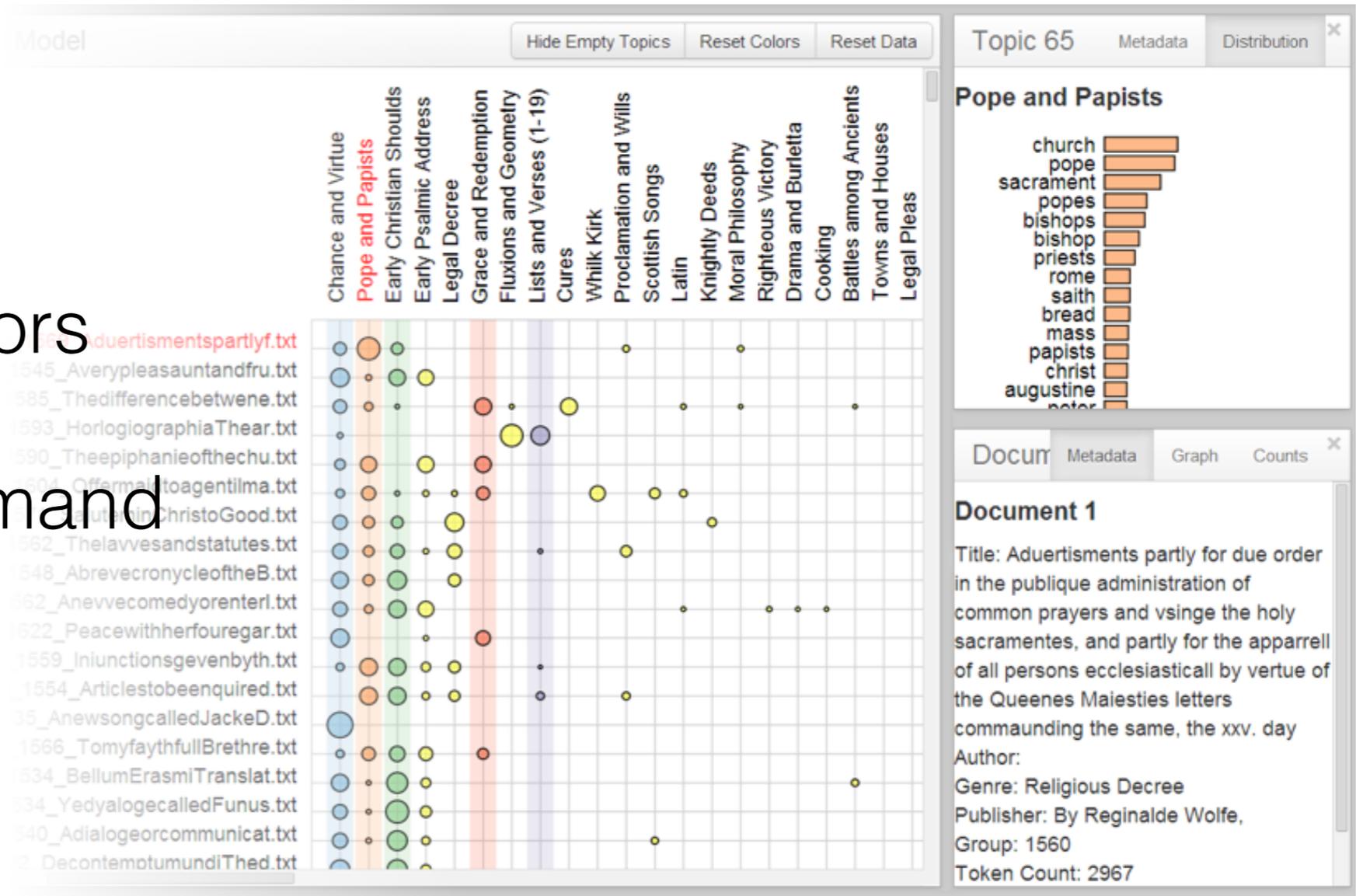
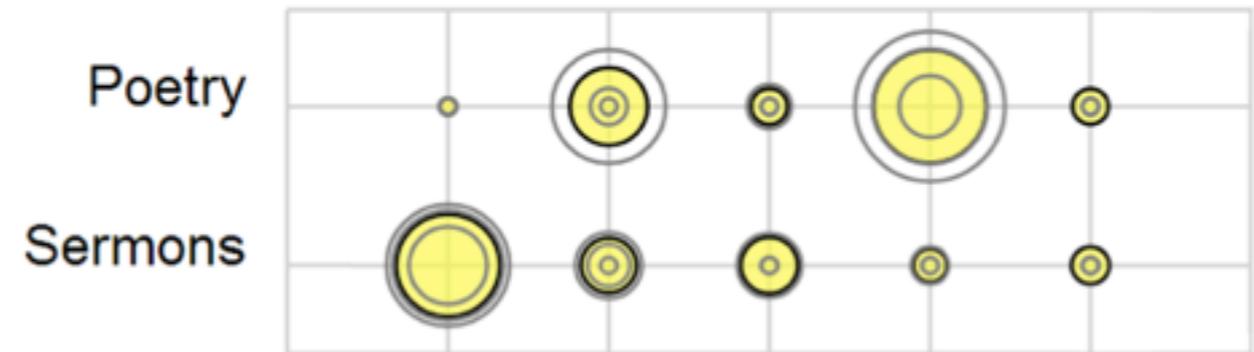
- **CorpusViewer**: Re-orderable matrix
- **TextViewer**: Examination within one document
- **RankViewer**: Examine specific words

CorpusViewer



Features combatting scale

- Ordering
- Aggregation
- Annotation
- Assigning colors
- Details on demand



TextViewer

Tags

Clear All

- Tales of Chance and Virtue
- Legal Decree
- Early Christian Shoulds**
- Proclamation and Wills
- Pope and Papists
- Early Psalmic Address
- Lists and Verses (1-19)

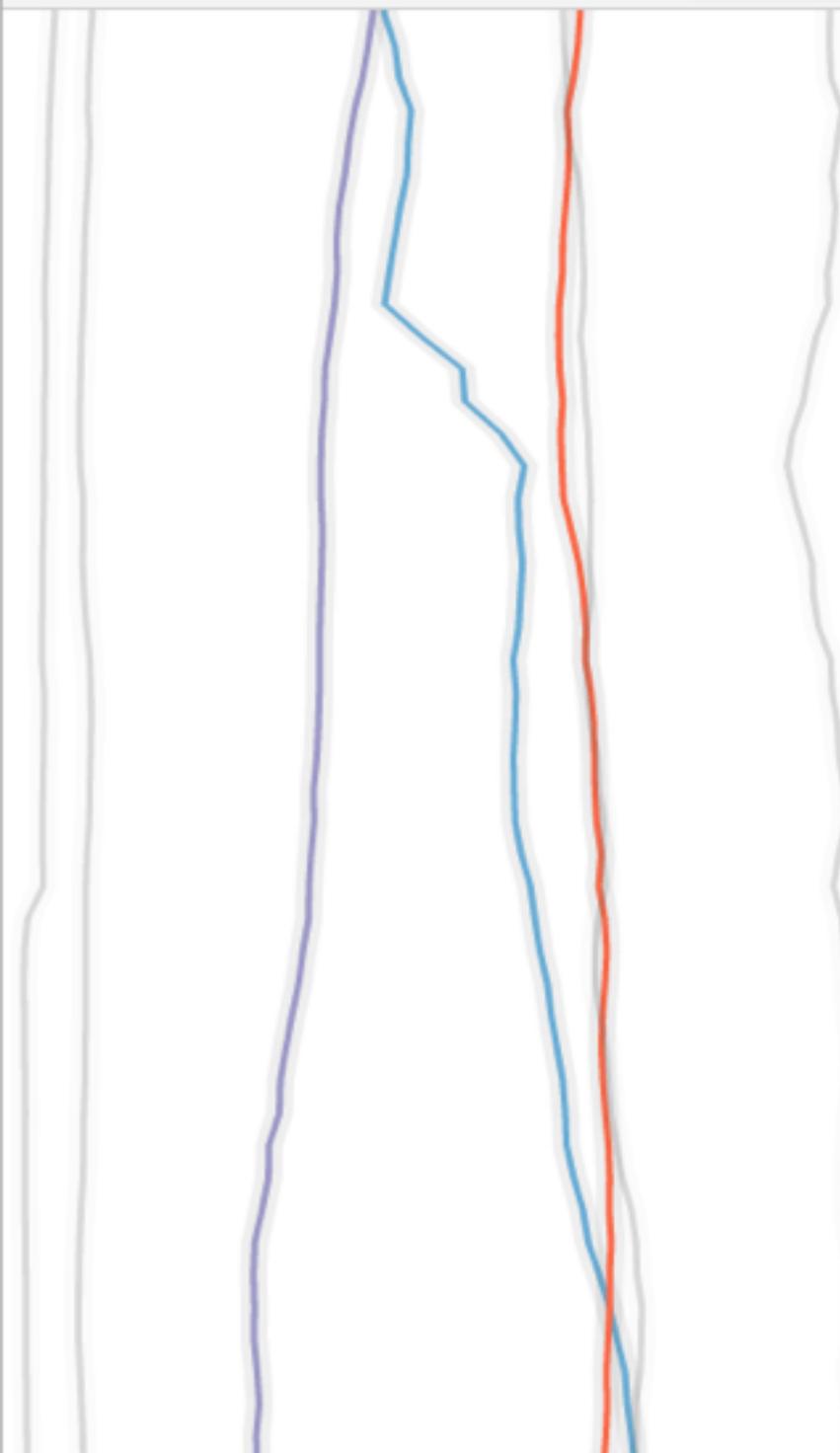
Text: A01594_1562_TheIavesandstatutes

✎

which the first inquisition pertained
to the consistory ecclesiasticall, let the
commissioners or elders with ye mi|nisters
take heed thereto: and if any be
convicted let them make their report
to the counsel with their advise and
Judgment, so that ye last Judgment
for the correction be always reserved
to the Seniory.
As concerning the offences which
ought to be corrected by simple admo|nitions,
let them therein proceed accor+ding
to ye order of our saviour christ,
so that the cause may be ended in the
ecclesiastical Judgment.
To maintain this discipline in his
estate, every three months let the mi+nisters
specially enquire if there be
any thing to be talked of amōge them
selves, and remedy it according to
reason.
•Of the number• place and, time of the
Sermons.
Upon the Sondayes there shall be
mornīg sermons at the churches

↑
↓
5
6
7
8
9
10
11
12
13
14
15
16
17
18
19
20
21
22
23

Topic Overview

Clear All

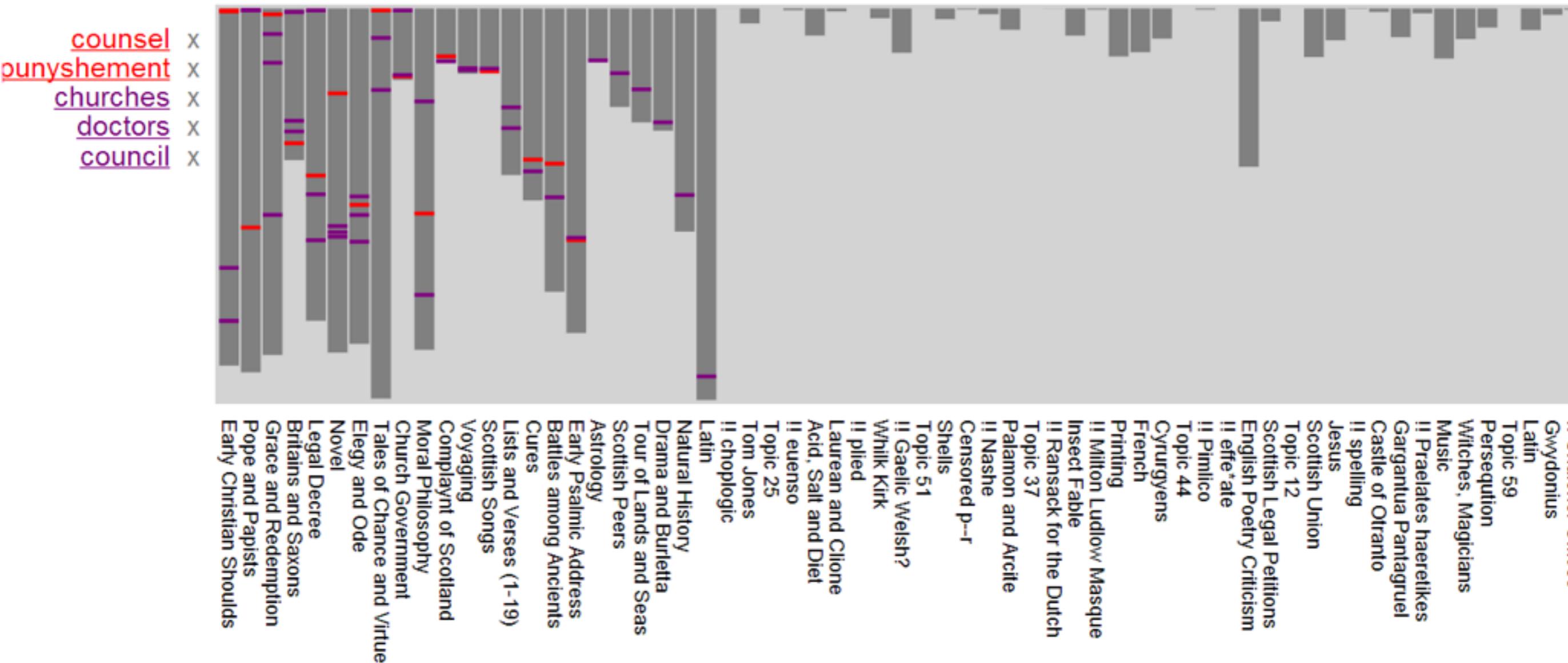
RankViewer

Serendip Word Rankings

Enter words separated by a space

purple ▾

Add



How are the factors for Serendipity implemented?

- Multiple access points: 3 Views, Ordered to user's liking
- Highlighting adjacencies: Through ordering and vis
- Flexible pathways: Jumping between views
- Curiosity and playfulness: Interaction & Discovery

2 use cases

- Vis Abstracts

1127 abstracts

From SciVis, InfoVis, VAST, BioVis, and PacificVis papers from 2007-2013
each 30 to 389 words

- Early Modern Literature

1080 digitized texts

From English literature published between 1530 to 1799
each few hundred words to few hundred pages

Formal evaluation still due

- To **confirm serendipitous discoveries** across multiple scales of data and abstraction
- Problem: How to evaluate serendipity?
- Long-term user studies needed

Discussion

+ strengths

- Suitable for any document and corpus size
- Three layers (whole corpus, single document, single words)
- Simple but effective visualizations
- Easily accessible (Online tool; *though topic modelling part still due*)

- weaknesses

- Not for quick exploration
- Sceptical about serendipitous discoveries

| | |
|---------------|--|
| System | Serendip |
| What | Data Document collection (Text, Metadata, Topics) |
| Why | Tasks Explore document collections Facilitate serendipitous discoveries |
| How | Encode Reorderable matrix, line graph, bar graph Reduce Aggregation, Ordering Manipulate Order, Color, Annotate, Details-on-Demand |
| Scale | Up to ~1,000 Documents (various length) |